

Науковий журнал
Радіоелектроніка
Інформатика
Управління



2003
1 (9)

Видається двічі на рік

Зареєстрований **29 січня 2003 року**
Державним комітетом інформаційної політики,
телебачення та радіомовлення України,
свідоцтво - серія **КВ № 6904**.

Засновник і видавник -
Запорізький національний технічний університет

Передплатний індекс - 22914

З М І С Т

| | |
|--|----|
| РАДІОЕЛЕКТРОНІКА | 3 |
| OPTIMIZING RADIATION PATTERN OF BROADSIDE ARRAY ANTENNA BY AMPLITUDE-POSITION PER- TURBATIONS USING A GENETIC ALGORITHM <i>Chao-Hsing Hsu, Kang K. Yen, Tadeusz Babij</i> | 3 |
| СТРУКТУРНАЯ УСТОЙЧИВОСТЬ ТОНКИХ ОСТРОВКОВЫХ ПЛЕНОК. РАСЧЕТЫ И КОМПЬЮТЕРНЫЙ АНАЛИЗ <i>В.И.Псарев, Л.А.Пархоменко, А.Ф.Куликов</i> | 16 |
| РАДИОФИЗИЧЕСКИЙ ПОДХОД К ИЗМЕРЕНИЮ СКОРОСТИ ГАЗОВЫДЕЛЕНИЯ В СВИНЦОВО- КИСЛОТНЫХ АККУМУЛЯТОРНЫХ БАТАРЕЯХ <i>В.А.Дзензерский, С.В.Плаксин, И.И.Соколовский</i> | 7 |
| СИСТЕМА МОДЕЛИРОВАНИЯ ФРАКТАЛОПОДОБ- НЫХ РАССЕЙВАЮЩИХ СТРУКТУР <i>С.Н.Романенко, Л.М.Карпуков, Р.Д.Пулов</i> | 21 |
| СОБСТВЕННЫЕ ФУНКЦИИ И ЧИСЛА ПРЯМОУГОЛЬНОГО ВОЛНОВОДА С ФРАКТАЛЬНЫМИ ШИРОКИМИ СТЕНКАМИ <i>А.А.Мисюра, В.М.Онуфриенко</i> | 12 |
| ЭЛЕКТРОСТИМУЛЯТОР ГЛАЗНЫХ МЫШЦ ЭКСПОНЕНЦИАЛЬНЫМИ МОДУЛИРОВАННЫМИ ИМПУЛЬСАМИ <i>Е.Я.Швец, Л.Л.Веревкин, О.Н.Поправка, А.П.Посунько</i> | 24 |
| ІНФОРМАТИКА | 27 |
| ON THE EDUCATIONAL COURSE OF LOCALYZING COMPUTING <i>Gregory G. Men'shikov</i> | 27 |
| HYBRID LEARNING ALGORITHM FOR NEURAL NET- WORKS <i>Jamil Ahmad</i> | 33 |
| DESIGN OF AN ADAPTIVE FUZZY LOGIC CONTROL- LER FOR NONLINEAR DYNAMIC SYSTEM <i>Ho Dac Loc</i> | 29 |
| THE INTELLIGENT CIRCUIT THAT IS OPERATED BY TRANSMISSION OF IMPULSE AS SYMBOL OF ACTIVITY <i>Karasawa S.</i> | 37 |

| | | | |
|--|-----|---|------------|
| SOLVING OF BOUNDARY VALUE PROBLEMS FOR MATHEMATICAL PHYSICS EQUATIONS IN CELLULAR NEURAL NETWORKS <i>М.А. Novotarskiy</i> | 42 | ОБЪЕКТНО-ОРИЕНТИРОВАННАЯ РЕАЛИЗАЦИЯ АЛГОРИТМА КЛОНАЛЬНОЙ СЕЛЕКЦИИ <i>В.И. Литвиненко, А.А. Фефелов, С.П. Горавский</i> | 81 |
| АППАРАТНО-ПРОГРАММНАЯ РЕАЛИЗАЦИЯ АЛГОРИТМА ВОССТАНОВЛЕНИЯ ИНТЕРПОЛЯЦИОННЫМИ СПЛАЙНАМИ <i>А.Ю. Балыкова, В.М. Жигачев, Б.В. Чувькин</i> | 47 | ИЕРАРХИЧЕСКАЯ СТРУКТУРА МОДЕЛИ АНАЛИЗА И ПРОГНОЗИРОВАНИЯ ПОДТОПЛЕНИЯ: АНАЛИЗ МОДЕЛЕЙ-ОСНОВАНИЙ <i>С.А. Митрохин</i> | 89 |
| СИНТЕЗ АВТОМАТА МИЛИ С ПРЕОБРАЗОВАНИЕМ МИКРООПЕРАЦИЙ В КОДЫ СОСТОЯНИЙ <i>А.А. Баркалов, Эль-Джейби Ахмад Камаль, А.А. Красичков</i> | 51 | АВТОМАТИЗИРОВАННАЯ СИСТЕМА АНАЛИЗА ФИНАНСОВО-ЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ ДЕЯТЕЛЬНОСТИ ПРЕДПРИЯТИЯ <i>А.В. Пархоменко, Н.А. Кузовенкова, О.В. Шинкарук</i> ... | 95 |
| КОМПЬЮТЕРНАЯ ПСИХОДИАГНОСТИЧЕСКАЯ СИСТЕМА ДЛЯ ОЦЕНКИ ПРОФПРИГОДНОСТИ ПЕРСОНАЛА РАЗДЕЛИТЕЛЬНЫХ ПРОИЗВОДСТВ <i>О.Г. Берестнева, В.А. Власов, А.А. Орлов</i> | 54 | ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ ПРИМЕНЕНИЯ ПАРАЛЛЕЛЬНЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ ДЛЯ МОДЕЛИРОВАНИЯ ДВИЖЕНИЯ КОСМИЧЕСКОГО АППАРАТА <i>Н.Д. Пиза</i> | 98 |
| ФОРМИРОВАНИЕ СТРУКТУРЫ ПОЛИНОМИАЛЬНОЙ НЕЙРОННОЙ СЕТИ С ПОМОЩЬЮ ГЕНЕТИЧЕСКОГО АЛГОРИТМА <i>Е.Ф. Васечкина, В.Д. Ярин</i> | 59 | КОНФИГУРАЦИЯ ОБЪЕКТА КАК ИНТЕРФЕЙС МЕЖДУ МАТЕМАТИЧЕСКИМ И ФИЗИЧЕСКИМ СОДЕРЖАНИЕМ МОДЕЛИ <i>В.П. Сизиков</i> | 104 |
| МОДЕЛИРОВАНИЕ ПРОЦЕССА ОБУЧЕНИЯ <i>А.И. Вершина, Б.Т. Солдатов</i> | 65 | ПРЕДВАРИТЕЛЬНАЯ НЕЙРООБРАБОТКА ИЗОБРАЖЕНИЙ В УСЛОВИЯХ ШУМОВ <i>О.В. Серпенинов, Е.А. Самойлин</i> | 111 |
| ЗАДАЧА КОММИВОЯЖЕРА И ГЕНЕТИЧЕСКИЕ АЛГОРИТМЫ <i>В.М. Гавриляко</i> | 72 | АЛГОРИТМЫ КЛАСТЕР-РЕГРЕССИОННОЙ АПРОКСИМАЦИИ ТА ЇХ НЕЙРОМЕРЕЖЕВИ ІНТЕРПРЕТАЦІЇ <i>С.О. Субботін</i> | 114 |
| РАЗРАБОТКА И ИССЛЕДОВАНИЕ ИЕРАРХИЧЕСКИХ АГЛОМЕРАТИВНЫХ АЛГОРИТМОВ РАЙОНИРОВАНИЯ <i>А.В. Заварзин, В.И. Месюра</i> | 76 | | |
| УПРАВЛІННЯ | | | 122 |
| АНАЛИТИЧЕСКИЙ МЕТОД НАХОЖДЕНИЯ ПЕРИОДИЧЕСКИХ НЕГАРМОНИЧЕСКИХ РЕШЕНИЙ В НЕКОТОРЫХ ЗАДАЧАХ ТЕОРИИ РАДИОТЕХНИЧЕСКИХ И ЭЛЕКТРОННЫХ ЦЕПЕЙ <i>М.Н. Горбачев</i> | 122 | СИНТЕЗ И АНАЛИЗ АНАЛОГО-ЦИФРОВЫХ АДАПТИВНЫХ НАБЛЮДАТЕЛЕЙ МНОГОФАЗНЫХ НЕОПРЕДЕЛЕННЫХ СИГНАЛОВ <i>Е.Е. Потапенко</i> | 139 |
| СИНТЕЗ ОПТИМАЛЬНЫХ ПО БЫСТРОДЕЙСТВИЮ ЦИФРОВЫХ РЕГУЛЯТОРОВ ДЛЯ ОБЪЕКТОВ УПРАВЛЕНИЯ С ЧИСТЫМ ЗАПАЗДЫВАНИЕМ <i>В.И. Гостев, В.В. Крайнев, С.Н. Скуртов</i> | 126 | СУММАРНЫЕ ФУНКЦИОНАЛЫ КАЧЕСТВА ИТЕРАЦИОННЫХ ДВУХКАНАЛЬНЫХ СИСТЕМ УПРАВЛЕНИЯ С ЭТАЛОННОЙ НАСТРОЙКОЙ И ПРИ СЛУЧАЙНЫХ ВОЗДЕЙСТВИЯХ <i>А.А. Худяев</i> | 142 |
| МИКРОПРОЦЕССОРНАЯ СИСТЕМА УПРАВЛЕНИЯ КОМПЛЕКСОМ ПО ПЕРЕРАБОТКЕ ОТХОДОВ ПОЛИМЕРНЫХ МАТЕРИАЛОВ <i>В.В. Дядичев</i> | 131 | МОДИФИКАЦИЯ МЕТОДА КОНЕЧНОГО СОСТОЯНИЯ ДЛЯ СИНТЕЗА УПРАВЛЕНИЯ РЕАЛЬНОГО ВРЕМЕНИ В МНОГОМЕРНЫХ НЕЛИНЕЙНЫХ ТЕРМИНАЛЬНЫХ СИСТЕМАХ С ДИФФЕРЕНЦИРУЕМЫМИ НЕЛИНЕЙНОСТЯМИ <i>Е.А. Шушлятин, А.Е. Шушлятина</i> | 149 |
| РАЗРАБОТКА ЦИФРОВЫХ АЛГОРИТМОВ ДЛЯ СИСТЕМЫ С ЗАПАЗДЫВАНИЕМ ПО УПРАВЛЕНИЮ <i>Е.Л. Еремин, Л.В. Ильина</i> | 134 | | |
| АВТОРЫ НОМЕРА | | | 154 |

І. РАДІОЕЛЕКТРОНІКА

УДК 621.326

OPTIMIZING RADIATION PATTERN OF BROADSIDE ARRAY ANTENNA BY AMPLITUDE-POSITION PERTURBATIONS USING A GENETIC ALGORITHM

Chao-Hsing Hsu, Kang K. Yen, Tadeusz Babij

В статье описана реализация антенны с амплитудно-позиционным возбуждением. Генетический алгоритм используется для оптимизации паттерна излучения антенны.

This paper presents the implementation of the broadside antenna by using adaptive array antenna. Even though this kind of broadside antenna is little bit costly compared with the conventional ones, its performance can be improved because it is possible to null the interfering signals and maximize the main lobe towards the array normal at the same time. Thus the optimal radiation pattern can be obtained. In many applications, it is required to have the maximum radiation of an array directed toward the array normal. An optimal radiation pattern design for a broadside array antenna is not only to derive the maximum power radiation at the array normal direction but also to suppress interference by placing a null at the direction of the interfering sources. The Signal Interference Ratio (SIN) can be maximized. Genetic algorithms are used for the search of the optimal radiation pattern by amplitude-position perturbations. Details on the structure of the system, radiation pattern formulation, the application of the genetic algorithm and simulation results are given.

1 INTRODUCTION

Conventional broadside antennas can form a main lobe always towards the array normal. Their structures are simple, but they are not able to compress the interferences. As the development of the modern communications, higher and higher communication quality is required by users for different kinds of antennas. In order to improve the performance of the broadside antenna, implementing the function of the broadside antennas with adaptive antenna array is a possible approach.

A broadside array antenna must keep maximum radiation toward array normal. In the previous papers [1-6], adaptive pattern nulling technique has been studied in the optimal radiation pattern search. Adaptive pattern nulling technique means minimizing the power of the interfering signal coming from any direction by putting a null in its direction. In this paper, the optimal radiation pattern of broadside array will not only be able to suppress interferences by placing nulls but also be able to derive the maximum power radiation in the array normal direction. A deduced formula of the radiation pattern of broadside array antennas is suitable for the

optimal solution search. The formula can always keep maximum radiation of a broadside array antenna toward array normal. At the same time, it can cancel all the interferences from the interfering sources. A perturbation method consists of small perturbations with the element parameters. The technique features are amplitude and position perturbations. As known, compared with gradient-based search methods, genetic algorithms can derive a global search by mutation technique and avoid being trapped in local optima. A search procedure based on the genetic algorithm is used to obtain the required perturbations for the designed optimal radiation patterns.

This paper first presents the structure of the adaptive antenna, which will be used as a broadside antenna. The idea is trying to simplify the structure as much as possible so as not to increase the cost much. Second, the formulation of the radiation pattern is deduced based on the characteristics of the broadside antenna. Third, the application of the genetic algorithms to the optimal radiation pattern design of the broadside antenna is explained. Last, the simulation results are given.

2 DEDUCING RADIATION PATTERN FORMULA FOR SEARCHING OPTIMAL SOLUTION

For an adaptive linear array of $2N$ equispaced sensor elements as Figure 1, an interfering signal with wavelength λ impinges on any two adjacent element n and element $n + 1$ by a distance d from the direction θ with respect to array normal as shown in Figure 2. To reach any two adjacent elements, there is a time delay τ as follows [7-8]:

$$\tau = \frac{d \sin \theta}{v}, \quad (1)$$

v is the propagation speed of radio wave. τ corresponds to a phase shift of ψ .

$$\psi = \frac{2\pi}{\lambda} d \sin \theta = kd \sin \theta. \quad (2)$$

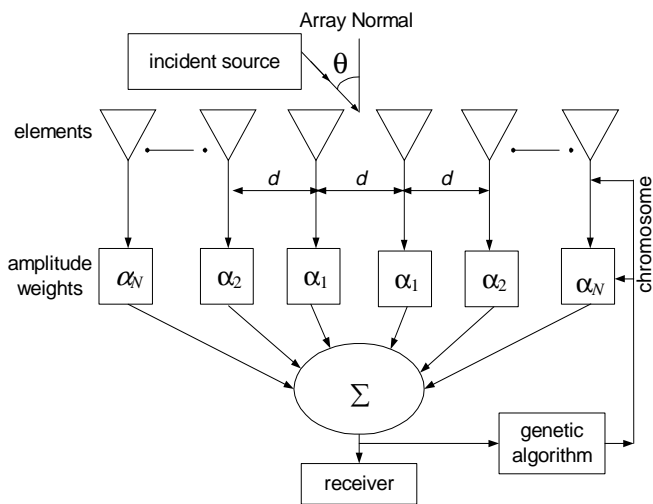


Figure 1 - Diagram of an adaptive linear array designed by amplitude-only perturbations using a genetic algorithm

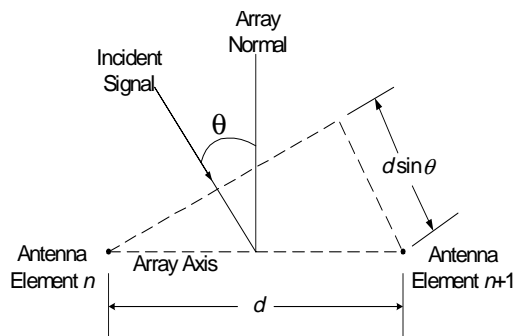


Figure 2 - The incident signal reaching any two adjacent elements

The array factor of adaptive linear array for far field using phase perturbations is given by

$$AF(\theta) = \sum_{n=1}^{2N} \alpha_n e^{j(n-1)\psi}, \quad (3)$$

where α_n - amplitude weight at element n .

If the reference point is at the physical center of the array, the array factor becomes

$$AF(\theta) = 2 \sum_{n=1}^{2N} \alpha_n e^{j(n-N-0.5)\psi}, \quad (4)$$

where $2N$ - number of elements, θ - incident angle of interfering signal or desired signal.

If amplitude weights are in even symmetry, the equation (4) can be simplified to

$$AF(\theta) = 2 \sum_{n=1}^N \alpha_n \cos[(n-0.5)\psi]. \quad (5)$$

In order to increase the performance of adaptive array, the antenna position perturbations in even symmetry are added. The array factor becomes as

$$AF(\theta) = 2 \sum_{n=1}^N \alpha_n \cos[(n-0.5)k(d+(\Delta d_n/(n-0.5)))\sin\theta], \quad (6)$$

where Δd_n is the position weight at element n .

The equation (6) can always keep maximum radiation toward array normal however the weights change. In addition, it is suitable for optimal solution searching.

3 GENETIC ALGORITHM APPROACH FOR OPTIMIZING RADIATION PATTERN

With fitness function given by the square of $AF(\theta)$ in equation (6), a genetic algorithm is used to adjust the position weights based on the power of the array in the interfering direction [9-10]. The goals are to minimize the total output power of the interfering signal. During the process of genetic iteration, the weight vector kept for the next step iteration should make the output power of the interfering signal to be decreased monotonically. Owing to a broadside array, the main lobe is always toward 0° with respect to array normal using equation (6), the radiation pattern could not only make the main lobe is always toward 0° or 180° but also minimize the total output power of the interfering signal at same time. Obviously, this technique can maximize the signal to interference ratio (SIR) for an broadside array antenna. The flow chart is shown in Figure 3. Steps of genetic algorithm are as follows:

Step 1. Initialization: A set of chromosomes is randomly generated. A chromosome (weight vector) is composed of genes (weights). For this problem, the genes are α_n and Δd_n . So, The initial step is generating a collection of random matrix vectors $[\alpha_n; \Delta d_n]$, $n=1, 2, \dots, N$, $m=1, 2, \dots, P$. The number of genes is $2N$. P is the number of chromosomes or population size. Define a vector and variable to which the gradually optimized chromosome and its fitness are saved separately. Their initial values are the first chromosome of the generated chromosome set and its fitness.

Step 2. Evaluation: For every chromosome, two objective functions are calculated for evaluating its fitness. One is the output power of the desired signal; the other is the output power of the interfering signal. Check every chromosome's fitness step by step. Compared with the present best fitness, if one chromosome can not only give the higher output power of the desired signal but also the lower output power of the interfering signal, renew the value of the defined vector and variable with this chromosome and its fitness. Otherwise, keep their values unchanged.

Step 3. Selection: A random number generator is used to generate random numbers whose values are between 0 and 1. If the value of random number is smaller than p_s , this chromosome survives; otherwise, it does not survive. The best fitness of the population always survives.

Step 4. Crossover: Pairs of parents are selected from these survivors. Single point crossover is selected to produce the next generation. The weight string from the beginning of chromosome to the crossover point is copied from one parent and the rest is copied from the second parent.

Step 5. Mutation: The basic mutation operator randomly generates a number as the crossover position and then changes the value of this gene randomly.

Step 6. Termination: Steps 2-5 are repeated till the predefined number of generations has been reached.

The best set of weights (amplitude and position weights) can be generated after Termination.

The flow chart of genetic algorithms is as follows:

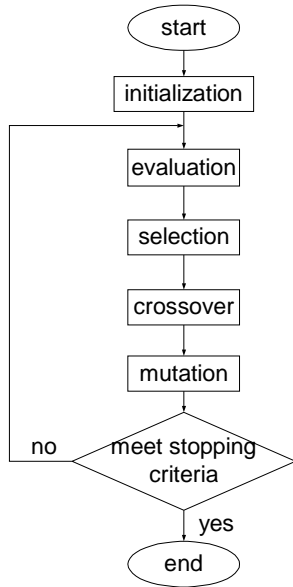


Figure 3 - Flow chart of genetic algorithm

4 SIMULATION RESULTS

For this problem, the necessary variables of genetic algorithm are defined as follows: population size p equals to 20; the selected survival rate p_s equals to 0,5; the probability of crossover is 0,5; the proportion of mutated genes is 0,5.

In this problem, assumed a linear antenna array is composed of 20 isotropic elements. So, N equals 10.

$1 \leq n \leq N$. The distance d of two adjacent elements is half of λ . The optimal radiation pattern technique features are by position perturbations. The fitness function is given by the square of $AF(\theta)$ in equation (6). Both amplitude weights and position weights are in even symmetry. The value of α_n is set between 0,1 and 1. The value of Δd is set between -0,24 and 0,24.

Case and Result: With respect to array normal, the interfering signal directions are from 40° and -20° respectively, and main lobe is toward 0° . The genetic algorithm is going to stop after 10000 iterations. The results are as showed in Table 1 and Figure 5. The optimal radiation pattern is derived. The SIN=125dB so that the interfering noise can be completely ignored.

Table 1 The weight for the optimal radiation pattern

| Main lobe toward 0° Nulling direction in from 40° and -20° | |
|--|-------------------------|
| $\alpha_1 = 0,572$ | $\Delta d_1 = 0,158$ |
| $\alpha_2 = 0,965$ | $\Delta d_2 = 0,077$ |
| $\alpha_3 = 0,506$ | $\Delta d_3 = -0,161$ |
| $\alpha_4 = 0,949$ | $\Delta d_4 = -0,005$ |
| $\alpha_5 = 0,761$ | $\Delta d_5 = -0,078$ |
| $\alpha_6 = 0,986$ | $\Delta d_6 = -0,168$ |
| $\alpha_7 = 0,678$ | $\Delta d_7 = -0,240$ |
| $\alpha_8 = 0,922$ | $\Delta d_8 = -0,197$ |
| $\alpha_9 = 0,830$ | $\Delta d_9 = -0,024$ |
| $\alpha_{10} = 0,166$ | $\Delta d_{10} = 0,132$ |

The relative power pattern is getting convergent as Figure 4. It can be seen that SIR reaches to 100dB only in the 45th generation. So, the efficiency of genetic algorithm approach is very high and the interference can be completely ignored in real time signal processing.

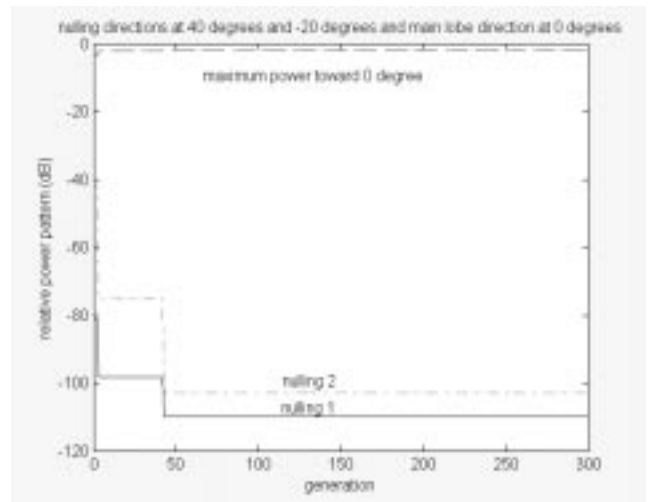


Figure 4 - Relative power pattern getting convergent using iteration

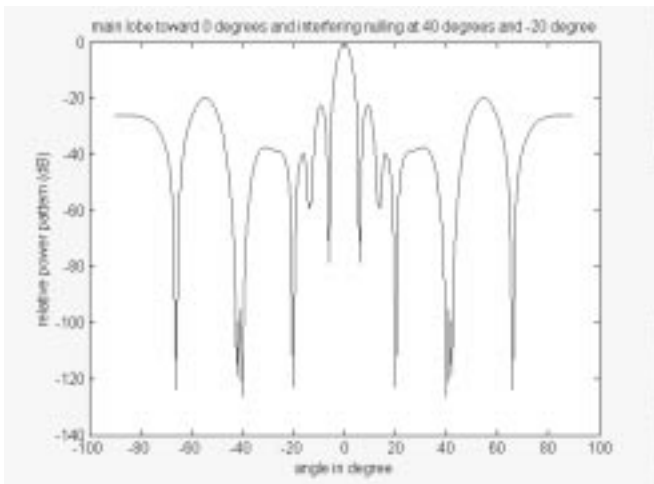


Figure 5 - Adaptive radiation pattern of broadside array antenna for main lobe toward 0° and interfering source at 40° and -20°

The interfering signal directions are from 40° and -20° (i.e. 340°) respectively, and the maximum radiation toward 0° . The radiation patterns are mapped from 90° to 270° as showed in Figure 6 in dB scale and in Figure 7 in decimal scale.

5 CONCLUSIONS

The optimal radiation pattern design of broadside array antenna based on amplitude-position perturbations using a genetic algorithm is proposed and achieved. For a broadside array, the desired signal direction is always at 0° with respect to array normal. In the paper, first, the output power formula based on amplitude and position for broadside array antenna composed by adaptive antenna array is deduced. In order to be able to adopt genetic algorithm to search the optimal radiation pattern, the formula is reformed through assuming amplitude and position weights are in even symmetry. Genetic algorithms are applied to find the optimal radiation pattern of the proposed adaptive antenna. As the optimal radiation pattern means that the power of desired signal should be highest and the power of interfering signals should be lowest as much as possible at the same time. In this paper, the optimal radiation pattern of broadside array was obtained. The convergence curves of genetic algorithm iteration show that it is effective for this problem. The approach of genetic algorithm in an adaptive linear array places nulls in the directions of interference and forms maximum main lobe in the direction of array normal to the far-field radiation pattern. The adaptive radiation pattern of broadside array antenna has been derived in this paper.

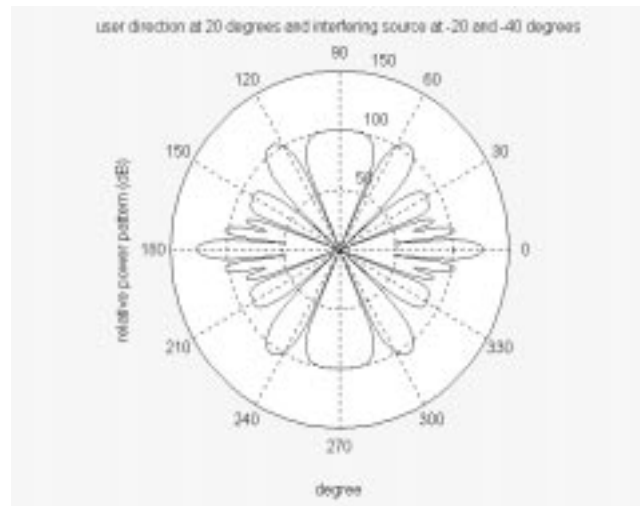


Figure 6 - The whole radiation pattern of adaptive broadside array antenna in dB scale

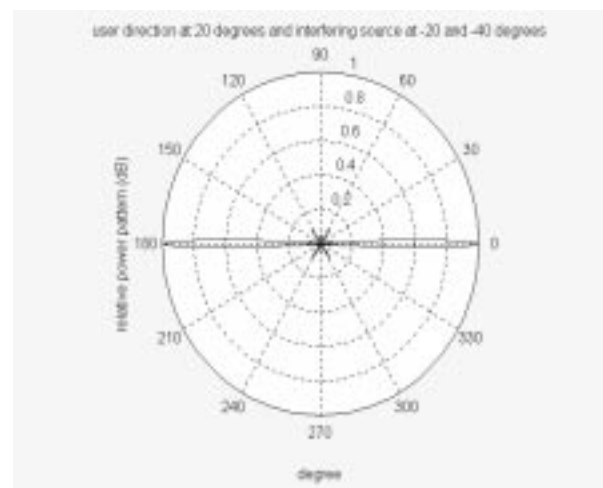


Figure 7 - The whole radiation pattern of adaptive broadside array antenna in decimal scale

REFERENCES

1. Alan J. Fenn, "Theoretical Near Field Cluster and Interference Cancellation for an Adaptive Phased Array Antenna," Proceedings of International Symposium Digest, Antennas and Propagation, pp. 47-49, 1987.
2. M. H. Er, "Synthesis of Antenna Array Pattern with Broad Null Constraints," Journal of Electrical and Electronics Engineering, Australia-IE Aust. & IREE Aust., vol. 10, no. 2, pp. 136-145, June 1990.
3. C. H. Hsu and T. M. Babij, "Pattern Nulling of Adaptive Antenna by Phase and Amplitude Perturbations Using Genetic Algorithm," Knowledge-Based Intelligent Information Engineering Systems & Allied Technologies, KES'2001, Part 2, pp.1047-1051, 2001.
4. Wen-Pin Liao and Fu-Lai Chu, "Array pattern Nulling by Phase and Position Perturbations with the Use of the Genetic Algorithm" Microwave and Optical Technology Letters, Vol. 15, No. 4, pp. 251-256, 1997.
5. You Chung Chung and Randy L. Haupt, "Amplitude and Phase Adaptive Nulling with A Genetic Algorithm," Proceedings of 15th Annual Review of Progress in Applied Computational Electromagnetics, Monterey, vol. 1, pp. 359-364, CA Mar. 15-20, 1999.
6. A. Tennant, M. M. Dawoud and A. P. Anderson, "Array Pattern Nulling by Element Position Perturbations Using a Genetic Algorithm," Electronics Letters, Vol. 30, No. 3, pp. 174-176, 3rd February 1994.

7. Monzingo, Miller, "Introduction to Adaptive Arrays" weley, New York, 1968.
8. Chao-Hsing Hsu and Kang Yen, "Optimizing Adaptive Linear Array Pattern by Phase-Position Perturbations" The 2nd International Conference on Information, pp. 56-59, July 24-28, 2002.
9. J. S. Pan, F. R. McInnes and M.A. Jack, "Codebook Design Using Genetic Algorithms" IEE Electronics Letters, Vol. 31, No. 17, pp.1418-1419, 1995.
10. D. E. Goldberg, "Genetic Algorithms in Search Optimization and Machine Earning", Addison-Wesley publishing Company, 1989.

УДК 621.355.1.035.32

РАДИОФИЗИЧЕСКИЙ ПОДХОД К ИЗМЕРЕНИЮ СКОРОСТИ ГАЗОВЫДЕЛЕНИЯ В СВИНЦОВО-КИСЛОТНЫХ АККУМУЛЯТОРНЫХ БАТАРЕЯХ

В.А.Дзензерский, С.В.Плаксин, И.И.Соколовский

Запропонований радіофізичний метод контролю газовиділення в свинцево-кислотних акумуляторах, заснований на зміні розподілу електричного поля в хвилеводній електродинамічній системі, що містить стабілізований напівпровідниковий ганновський генератор. Інформація про швидкість газовиділення дозволяє оцінити міру саморозряду і ступінь зарядженості акумуляторної батареї, оптимізувати режим заряду акумуляторної батареї.

Предложен радиофизический метод контроля газовой выделення в свинцево-кислотных аккумуляторах, основанный на изменении распределения электрического поля в волноводной электродинамической системе, содержащей стабилизированный полупроводниковый ганновский генератор. Информация о скорости газовой выделення позволяет оценить меру саморазряда и степень заряженности аккумуляторной батареи, оптимизировать режим заряда аккумуляторной батареи.

The radiophysical method of control of gassing in the lead-acid accumulators, based on the change of distribution of the electric field in the waveguide electrodynamic system containing the stabilized semiconductor Gann-generator, is offered. Information about the gassing rate allows to estimate the measure of self-discharge and degree of charge of accumulator battery, to optimize the mode of charge of accumulator battery.

ВВЕДЕНИЕ

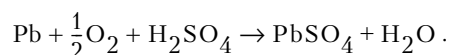
Одной из причин снижения ресурсных характеристик свинцево-кислотных аккумуляторных батарей (стартерных, стационарных, тяговых) является чрезмерный разряд и неконтролируемое время нахождения их в разряженном состоянии. Чрезмерный разряд при этом может быть связан как с условиями эксплуатации батарей, так и с саморазрядом. Саморазряд батарей проявляется при хранении их на складе или при длительных перерывах в их эксплуатации. Под саморазрядом подразумевают потерю емкости аккумуляторной батареи с разомкнутой цепью, и происходит он может за счет токов утечки по поверхностям батареи, смоченной электролитом, между выводами токоведущих мостиков, за счет токов утечки системы электропитания транспортного средства, а также за счет определенных физико-химических механизмов в аккумуляторе, включающих анодную реакцию ионизации металла и катодный восстановительный процесс. С точки

зрения термодинамики причина саморазряда (и коррозии металла токоотвода) обусловлена термодинамической неустойчивостью металлического свинца и диоксида свинца при работе их в качестве электродов аккумулятора. PbO₂-электрод располагается в зоне термодинамической неустойчивости воды диаграммы Пурбе ($E^0 = +1,629$ В) и взаимодействует с водой с выделением кислорода. Pb-электрод также расположен в зоне термодинамической неустойчивости воды ($E^0 = -0,356$ В) и должен взаимодействовать с водой с выделением водорода. И в этом смысле саморазряд свинцево-кислотных аккумуляторов и сопряженное с этим газовой выделення является неизбежным. Контроль скорости газовой выделення является поэтому важной процедурой, так как позволяет прогнозировать меру саморазряда и степень заряженности аккумуляторной батареи без контрольных измерений.

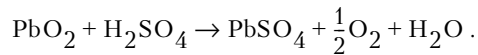
Кроме того, в реальных условиях у всех металлов, являющихся добавками или встречающихся в качестве примесей в сырье аккумуляторных электродов, значение водородного перенапряжения ниже, чем на чистом свинце. Снижение энергетического барьера процесса $2H^+ + 2e \rightarrow H_2$ повышает скорость реакций $Pb - 2e + SO_4^{2-} \rightarrow PbSO_4$, то есть скорость выделення водорода прямо связана с коррозией. Поэтому разработка надежного метода контроля темпа газовой выделення будет способствовать выявлению добавок, повышающих коррозию, и облегчит поиск иных добавок, снижающих коррозию.

1 МОТИВАЦИЯ РАЗРАБОТКИ ИЗМЕРИТЕЛЯ СКОРОСТИ ГАЗОВЫДЕЛЕНИЯ

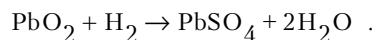
Имеется еще один мотив для проведения исследований по обеспечению надежных измерений скорости газовой выделення. Как указывалось выше, саморазряд и коррозия отрицательного электрода связаны с взаимодействием растворенного в электролите кислорода со свинцом:



Этот процесс по мере образования слоя сульфата свинца замедляется и его роль в негерметизированных аккумуляторах не столь существенна. Саморазряд положительных электродов обусловлен самопроизвольным восстановлением диоксида свинца в соответствии с реакцией:



Этот процесс, как и процесс саморастворения свинца $\text{Pb} + \text{H}_2\text{SO}_4 \rightarrow \text{PbSO}_4 + \text{H}_2$, является суммарной реакцией двух сопряженных процессов: катодного восстановления диоксида свинца: $\text{PbO}_2 + 2e + \text{H}_2\text{SO}_4 + 2\text{H}^+ \rightarrow \text{PbSO}_4 + 2\text{H}_2\text{O}$ и анодного окисления воды: $\text{H}_2\text{O} - 2e \rightarrow 1/2\text{O}_2 + 2\text{H}^+$. Причем, скорость процесса самопроизвольного восстановления диоксида свинца зависит от наличия примесей, снижающих потенциал выделения кислорода, от концентрации H_2SO_4 , от температуры и других факторов. Коррозия решетки свинцово-сурьмянистого положительного электрода, как и саморазряд, обусловлена окислением сурьмы, действием короткозамкнутых микрогальваноэлементов $\text{Pb}/\text{H}_2\text{SO}_4/\text{PbO}_2$, взаимодействием PbO_2 как сильного окислителя с веществами, способными восстанавливаться, например, по реакции:



Необходимость надежных и быстрых измерений скорости газовой выделению существует и в технологических операциях по ускоренному заряду аккумуляторных батарей как на стадии их серийного производства, так и при эксплуатации (подзаряде).

Как известно, зарядный ток приводит к возникновению химической реакции, изменяющей химический состав материала пластины и электролита. Скорость этой реакции зависит от величины зарядного тока - чем выше ток, тем быстрее протекает реакция. Безотносительно к условиям проведения зарядного процесса (режим заданного тока, режим заданного напряжения, импульсный, в том числе знакопеременный режим и пр.), в конечном счете, именно заряд, связанный с этим током, запасается в аккумуляторе для последующего использования. В процессе заряда аккумуляторная батарея накапливает все больший заряд и, в конце концов, наступает насыщение, то есть химическая реакция стабилизируется (или уравнивается) и дальнейшее накопление заряда прекращается. При этом большинство сульфатных ионов, которые были поглощены из раствора серной кислоты свинцовыми пластинами во время цикла разряда аккумуляторной батареи, возвращается из пластин в раствор. При этом пластины приобретают металлические свойства и начинают вести себя как электроды, помещенные в водный раствор, то есть прекрасную среду для электролиза. Зарядный ток в дальнейшем при продолжении заряда начинает разлагать воду в электролите на водород

и кислород ("кипение" батареи) с интенсивным газовой выделением. Статистика показывает, что газовой выделение начинается, когда в батарее запасено около 70-88% полного заряда. Если батарею заряжать и дальше с прежней скоростью, газовой выделение может привести к повреждению батареи. Необходимо или полностью прекратить подачу тока в аккумуляторную батарею, или изменить режим заряда ("дозарядка"). Существует, в частности, точка зрения, что не следует доводить батарею до полного заряда [1], так как в реальных условиях, например, при использовании батарей в качестве стартерных в транспортных средствах, даже небольшие отклонения в динамическом процессе заряд-разряд могут привести к избыточному газовой выделению. Это опасно, так как приводит к преждевременному износу аккумуляторной батареи и к нерациональному расходованию зарядного тока на электролиз.

2 ОПТИМАЛЬНОЕ ЗНАЧЕНИЕ ПОДДЕРЖИВАЮЩЕГО ТОКА

Из анализа вышеприведенных процессов в [2] предложен подход, направленный на снижение саморазряда, коррозии и газовой выделению. Суть этого подхода заключается в следующем. Устойчивое существование PbO_2 и Pb было бы возможным, если бы потенциал PbO_2 - электрода превышал значение +1,629В, а потенциал Pb - электрода был более -0,356В. Это можно реализовать, если на аккумулятор подать напряжение такой величины, чтобы протекающий ток был направлен навстречу току саморазряда и по абсолютной величине был равен току саморазряда. По существу это будет ток, поддерживающий оптимальный режим аккумулятора: этим будет обеспечена катодная защита отрицательного электрода и анодная защита положительного электрода, снижено до минимума газовой выделение. Важным моментом при этом является оценка оптимального значения указанного поддерживающего тока. Очевидно, что это оптимальное значение соответствует минимальному (по существу - нулевому) газовой выделению. Превышение этого оптимального значения поддерживающего тока приводит вновь к росту газовой выделению из-за разложения воды. Таким образом, измерения скорости газовой выделению могут дать достаточно надежные значения тока поддержания для конкретного типа аккумуляторной батареи, а динамический контроль указанной скорости газовой выделению - обеспечить необходимые вариации значений поддерживающего тока с учетом меры износа аккумуляторной батареи, температуры окружающей среды и потенциального ресурса.

3 ОБОСНОВАНИЕ ВЫБОРА ДАТЧИКА КОНТРОЛЯ ТЕМПА ГАЗОВОЙ ВЫДЕЛЕНИЯ

Можно значительно уменьшить разрушительные последствия выделения газа, снизив зарядный ток при появлении признаков газовой выделению. Для этого также необходим высокочувствительный датчик скорости газовой выделению, который позволял бы регулировать подачу зарядного тока в аккумуляторную батарею.

Попытки использования датчиков расхода газа для обеспечения оптимального режима заряда аккумуляторов ранее предпринимались [1, 3-5], где в качестве указанных датчиков использовалась комбинация реометр-датчик давления. Использование таких датчиков расхода газа не позволяет оперативно информировать регулятор зарядного устройства о необходимых изменениях режима заряда, а процедура выработки информационного сигнала для управления режимом заряда достаточно сложна. Необходим датчик, который непосредственно преобразовывал бы изменения скорости газовой выделении в изменения электрического сигнала, был бы безынерционным, а эксплуатация его, как и весь процесс заряда, была бы безопасной и обеспечивала бы возможность полной автоматизации процесса заряда.

Таковыми возможностями обладают датчики, построенные на радиофизических принципах, в которых информационный сигнал может быть выработан за счет изменения параметров электродинамической системы при внесении в последнюю импедансной среды, объем и скорость поступления которой в электродинамическую систему пропорциональны скорости газовой выделении в процессе заряда аккумулятора, аналогично разработанному ранее способу измерения гидрофильности сепараторов для химических источников тока [6].

4 УСТРОЙСТВО ДЛЯ ОПРЕДЕЛЕНИЯ РАСХОДА ГАЗА

Суть разработанного способа определения расхода газа и устройства для его осуществления в технологиях заряда и подзаряда аккумуляторов, преимущественно свинцовых, никель-кадмиевых и никель-железных, видна из рис. 1, где 1 - аккумулятор с токовыми клеммами 2 и газоотводящим каналом 3. К газоотводящему каналу подключена газоотводящая трубка 4, свободный конец которой входит в коническую часть бюретки 5, заполненной насыщенной кислородом жидкой импедансной средой (в простейшем случае водой).

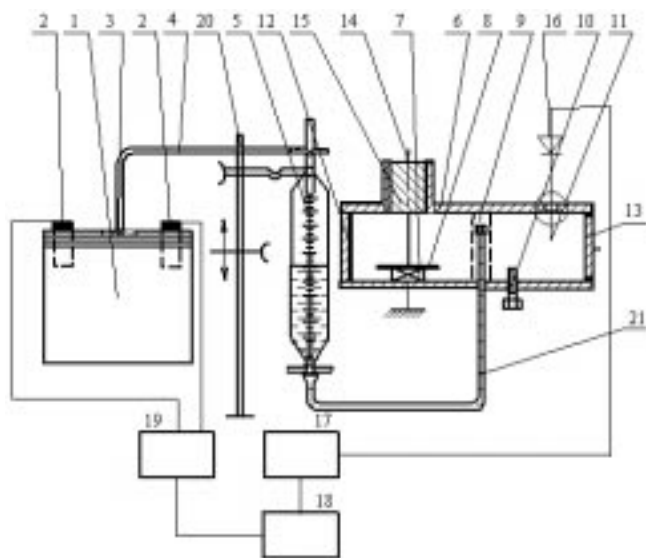


Рисунок 1 - Функциональная схема устройства для контроля заряда аккумулятора

Устройство содержит также сверхвысокочастотный (СВЧ) источник электромагнитного излучения с емкостным индикатором СВЧ поля, состоящий из отрезка закороченного с двух сторон прямоугольного волновода 6 с размещенными в нем полупроводниковым генераторным диодом 7 (преимущественно диодом Ганна), дисковым резонатором 8, индуктивной диафрагмой 9, настроечным штырем 10, емкостным зондом 11 с элементами гальванической связи (шайбовые резисторы) и элементами экранировки зонда от внешних полей (поглощающие втулки из прессованного порошкового железа) - на рисунке обведено кружком. Настроечный штырь 10 снабжен резьбой и изготовлен из поглощающего СВЧ излучение материала, преимущественно из магнетодиэлектрической массы. Первый короткозамыкатель, примыкающий к дисковому резонатору 8, снабжен СВЧ поглощающей вставкой 12 также из магнетодиэлектрической массы, а второй короткозамыкатель 13 выполнен подвижным с возможностью фиксации его в выбранном положении. Полупроводниковый диод запитывается от источника постоянного стабилизированного тока (на чертеже не показан) через ввод 14 и фильтр нижних частот 15, предотвращающий попадание СВЧ мощности в блок питания. С емкостным зондом 11 связан детекторный диод 16, подключенный к устройству сравнения 17 и через контроллер 18 к зарядному устройству 19. К зарядному устройству через клеммы 2 подключен аккумулятор 1. Бюретка 5 с газоотводящей трубкой 4 закрепляется на штативе 20 с возможностью изменения высоты установки. К нижней части бюретки подключена трубка 21 с кольцевым торцевым отверстием, входящая в индуктивную диафрагму 9, представляющую собой две пластины 22, 23, расположенные на узких стенках прямоугольного волновода, рис.2.

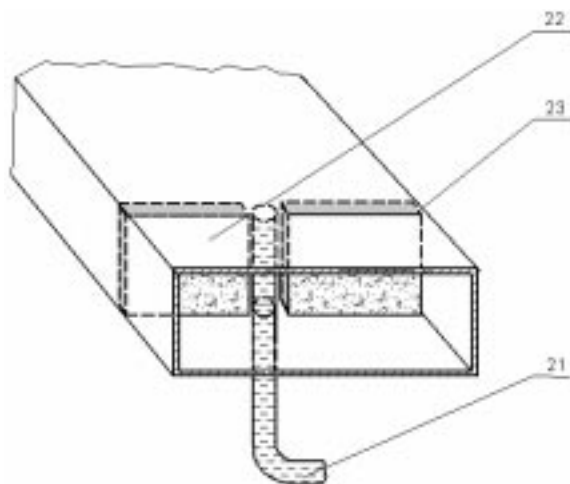


Рисунок 2 - Взаимное размещение индуктивной диафрагмы и газоотводящей трубки в сечении прямоугольного волновода

В зазоре между указанными пластинами размещена трубка 21. Емкостной зонд 11 связан с волноводом 6 в общем случае через щелевое отверстие в верхней широкой

стенке отрезка волновода с возможностью перемещения вдоль волновода и расположен на расстоянии $\lambda_B/4$, где λ_B - длина волны генерируемых СВЧ колебаний, от короткозамыкателя 13, то есть в максимуме ("пучности") напряженности поля. При связи зонда с волноводом через круглое отверстие, находящееся в фиксированном положении на плоскости широкой стенки волновода, необходимое расстояние между зондом и короткозамыкателем 13 устанавливается регулировкой положения подвижного короткозамыкателя. При практическом использовании указанные расстояния определяются по максимальным показаниям тока детектора 16 в положении, когда жидкая импедансная среда в трубке 21 не выше плоскости нижней широкой стенки отрезка прямоугольного волновода 6. В отсутствие заряда аккумулятора указанный уровень устанавливается соответствующим выбором высоты бюретки 5 на штативе 20. Устройство сравнения 17 при этом регулируется таким образом, чтобы сигнал на его выходе отсутствовал.

5 ПРОЦЕДУРА ИЗМЕРЕНИЙ

Контроллер 18, включающий микропроцессор и таймеры, отрегулирован таким образом, что в первоначальный момент от зарядного устройства, включающего регулятор тока и регулятор напряжения, зарядка аккумулятора начинается большим током без какого-либо ограничения по напряжению. При этом отношение длительности зарядного импульса τ_1 к длительности разрядного (деполяризующего) импульса τ_2 должно быть в пределах 80-120, отношение соответствующих амплитуд токов - в пределах 0,18-0,25, отношение паузы к длительности разрядного импульса - 4-5 (рис.3), причем амплитуда зарядного импульса соответствует значению, при котором аккумулятор должен быть заряжен за 2,5 часа.

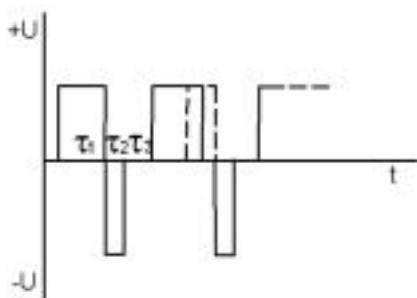


Рисунок 3 - Временное представление зарядного и разрядного импульсов, разделенных периодом стабилизации

Осуществляют зарядно-разрядный цикл до появления газовой выделения. При этом продуцируемый вследствие физико-химических реакций газ поступает через газоотводящую трубку 4 в бюретку 5 и выталкивает соответствующее количество жидкой импедансной среды в трубку 21, и жидкость оказывается в плоскости

диафрагмы 9, являющейся одновременно элементом связи собственно СВЧ генератора, образованного диодом Ганна 7, дисковым резонатором 8, отрезком волновода левее плоскости установки дискового резонатора с отрезком волновода, содержащим настроечный штырь 10, емкостной зонд 11 и короткозамыкатель 13, и частью СВЧ нагрузки (вторая часть СВЧ нагрузки представлена поглощающей вставкой 12). Так как до поступления импедансной среды в пространстве "индуктивная диафрагма - короткозамыкатель 13" образуется стоячая волна на частоте f_0 , что соответствует распределению напряженности поля, представленному сплошной линией на рис. 4, то при поступлении жидкой импедансной среды в колебательную систему из-за изменения частоты генерации вследствие эффекта "затягивания частоты" [7] до значения f_1 , пространственное расположение минимумов и максимумов поля изменится (минимум поля сместится на величину Δd , напряженность поля E изменится на величину ΔE , рис. 4), так что ток детектора 16, поступающий в устройство сравнения 17, изменяется (прерывистая кривая на рис. 4). При этом изменения тока детектора зависят от объема выделяющегося газа. Причем, в зависимости от применяемого типа устройства сравнения возникает необходимость изменения (например, уменьшения) коэффициента стоячей волны в пространстве "индуктивная диафрагма - короткозамыкатель 13". Это осуществляется за счет введения в полость отрезка волновода настроечного штыря 10, обладающего поглощающими свойствами в СВЧ-диапазоне.

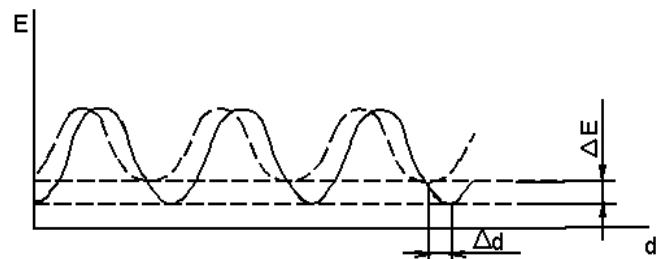


Рисунок 4 - Изменение напряженности электрического СВЧ поля на участке "индуктивная диафрагма - второй короткозамыкатель"

6 УПРАВЛЕНИЕ ИЗМЕНЕНИЕМ РЕЖИМА ЗАРЯДКИ

Ток детектора, изменившийся по величине вследствие перераспределения СВЧ-поля под зондом (см. рис. 4), поступает в устройство сравнения 17 (последнее может быть выполнено в виде электромагнитного реле, дифференциального усилителя или измерительного моста), с которого информационный сигнал поступает в контроллер 18, включающий таймеры и микропроцессор, который выдает команды зарядному устройству на изменение режима заряда с тем, чтобы либо исключить газодыделение, либо свести его к минимуму. Изменение режима заряда может состоять в изменении амплитуды зарядного тока или в изменении соотношения длительностей

зарядного и разрядного импульсов. Первый режим предполагает, что величина тока зарядного импульса составляет 10-12% от зарядного тока в предшествующий период зарядки. Во втором варианте соотношение длительностей зарядного и разрядного тока снижается до 10 и в таком режиме продолжается заряд до очередного газовыделения. Первый цикл газовыделения в свинцовых аккумуляторах происходит при заряженности аккумулятора до 80%, второй цикл - при 88-90%. После второго цикла газовыделения зарядный ток устанавливают величиной 4-5% от первоначального значения, разрядный импульс исключают, и заряд в таком режиме для свинцовых аккумуляторов осуществляют в течение времени, составляющего 0,4-0,45 интервала времени от начала зарядки до появления первого цикла газовыделения или 0,2-0,21 интервала времени до появления второго цикла газовыделения с момента начала заряда.

Выбор конструкции колебательной системы в виде дискового резонатора, установленного в отрезке прямоугольного волновода на расстоянии $\lambda_B/2$ от первого короткозамыкателя, и выбор элемента связи в виде индуктивной диафрагмы определяются тем, что в этом случае индуктивный реактанс диафрагмы сопрягается с емкостным реактансом генераторного диода Ганна, благодаря чему обеспечивается когерентное монохроматическое излучение с частотой, близкой к расчетному значению резонансной частоты дискового резонатора. При этом годограф точек стационарного режима на комплексной плоскости импедансов диода и нагрузки представляет собой окружность, совпадающую с окружностью постоянного коэффициента полезного действия генератора в значительном диапазоне изменений частоты, вызванных эффектом "затягивания частоты" без изменений генерируемой мощности. Это обеспечивает высокую надежность функционирования устройства в широком диапазоне интенсивностей газовыделения, так как изменения электрического поля под зондом, являющиеся информационными, вызваны только изменениями частоты генерации и мощности, поступающей в индикаторную часть отрезка волновода, а не изменениями генерируемой мощности диода Ганна.

ЗАКЛЮЧЕНИЕ

Разработанный метод контроля газовыделения адресован проблеме оптимального заряда и подзаряда негерметичных аккумуляторов и аккумуляторных батарей. Ускоренный метод заряда аккумуляторов важен в ряде применений: при зарядке аккумуляторов в солнечных энергетических установках, где необходим заряд в период максимальной освещенности фотопреобразователей, в серийном производстве, при заряде аккумуляторов в электромобилях, а также при хранении аккумуляторных батарей на складе. Использование знакопеременного зарядного импульсного тока известно [например, 8-15], однако в указанных источниках соотношения длительностей и амплитуд зарядного и разрядного импульсов различаются больше, чем на порядок и обусловлено это различными критериями в оценке требуемых параметров

зарядного тока или отсутствием каких-либо априорных критериев. Последнее вызвано тем, что в силу сложностей и неполного знания всех физико-химических процессов в аккумуляторах практически всех систем, технологии изготовления аккумуляторов на различных предприятиях неизбежно различаются и могут не воспроизводиться даже в пределах одного промышленного предприятия, например, при смене сырья или при замене технологического оборудования. Поэтому необходим такой подход в оценке оптимальной процедуры заряда и подзаряда аккумуляторных батарей, который был бы максимально адаптирован к конкретному аккумулятору или аккумуляторной батарее, достигалась бы эквализация их заряженности. Таким подходом и является использование информации о скорости газовыделения при заряде аккумуляторов. Это позволяет осуществить активное регулирование процесса заряда по реакции аккумулятора на зарядный ток, а благодаря использованию радиоволнового датчика получить максимальный объем информации о процессе - не только о наличии газовыделения, но и о динамике увеличения-снижения объема газовыделения, что позволяет гибко реагировать на изменение процесса зарядки. При этом достигается важнейшая цель в осуществлении заряда - аккумулятор доводится до состояния полного заряда за минимальное время при минимальном газовыделении, которое используется как параметр для достижения высокой скорости зарядки, так как удается практически мгновенно изменить режим заряда в самом начале процесса газовыделения. При этом в наибольшей степени могут быть реализованы преимущества импульсного знакопеременного заряда: в аккумуляторной батарее запасается максимальное количество заряда в кратчайшее время, при этом вероятность перезарядки и порчи аккумулятора исключается; периодическое изменение направления тока при заряде аккумуляторной батареи позволяет управлять восстановительными реакциями и структурными изменениями активной массы пластин, что дает возможность увеличить поверхность соприкосновения электролита с активной массой электродов, тем самым облегчая условия диффузии и выравнивания концентрации электролита в приэлектродном слое, что является принципиальным обстоятельством для повышения разрядной емкости аккумуляторов.

Разработанный метод контроля газовыделения апробирован при реализации ускоренной зарядки свинцово-кислотных стартерных аккумуляторных батарей серии 6СТ-55 и 6СТ-60 с токоотводами трех типов: А - оба токоотвода из свинцово-кальциевого сплава; Б - оба токоотвода из свинцово-сурьмянистого сплава; В - положительный токоотвод из свинцово-сурьмянистого сплава (содержание сурьмы 18%), отрицательный токоотвод из свинцово-кальциевого сплава. При этом, поскольку газовыделение прямо связано с расходом воды электролита (процесс диссоциации), то минимизация расхода воды является определяющим фактором в достижении оптимизации зарядки батареи. Так, при зарядке аккумуляторных батарей с токоотводами типа А удается снизить расход воды до уровня 1,7г/Ач (против 2,2г/Ач при стандартной зарядке), батарей с токоотводами типа Б до 3,8г/Ач (против 6,2г/Ач при стандартной зарядке), и батарей с

токотводами типа В до 2,45Г/Ач (против 3,5Г/Ач при стандартной зарядке). Соответственно снижается непроизводительный расход зарядного тока на диссоциацию воды.

Разработанный метод контроля газовой выделения в аккумуляторах при их заряде как инструмент в обеспечении оптимизации заряда может оказаться эффективнее и несомненно технологически проще по сравнению с методиками, которые используют для этих целей счетчики энергии, измерители напряжения на клеммах аккумуляторных батарей, измерители температуры и плотности электролита, измерители уровня электролита в аккумуляторе. Радиофизический метод контроля газовой выделения может быть особенно полезен при специальных исследованиях, когда аккумуляторные батареи необходимо размещать в камерах с повышенной или пониженной температурой, поскольку контроль газовой выделения может проводиться дистанционно.

ПЕРЕЧЕНЬ ССЫЛОК

1. Патент № 6,194,874, США, МКИ H02J 7/00. System and method for maintenance charging of battery cells / S. Kalogeropoulos, R.Andersson, J. Mercke etc. (СН); Telefonaktiebolaget LM Ericsson. № 397,001. Заявл. Сент. 15, 1999. Оpubл. февр. 27, 2001, НКИ 320/160. - 10 с.
2. Дзензерский В. А., Калиновский Е. А., Крысь В. Я., Палагин А. Ю., Привалов В. Н. Электрохимическая защита пластин свинцовых аккумуляторов от коррозии и саморазряда / Тез. доповідей 5 Міжвузівської конференції "Проблеми корозії та протикорозійного захисту конструкційних матеріалів". - Львів, 1999. - С.7-8.
3. Рашиев К.К. Коэффициент отдачи и использования зарядной емкости железоникелевого аккумулятора // Известия АН Латв. ССР, 1960, № 11. - С.73-81.
4. А.с. 41526 СССР, МКИ 21b/02. Устройство для зарядки железоникелевых аккумуляторов / К.К. Рашиев (СССР). - Бюл. изобретений и товарных знаков - 1961, № 19 от 29.06.60.
5. А.с. № 156207 СССР, МКИ H02P; 21C; 5/01. Устройство для автоматического управления подзарядом аккумулятора / К.К. Рашиев (СССР). - Бюл. изобретений и товарных знаков - 1963, № 15 от 24.02.62.
6. Дзензерский В. А., Плаксин С. В., Соколовский И. И. Радиофизический метод измерения гидрофильности сепараторов химических источников тока. - Д.: Изд-во Днепропетр. ун-та, 2002. - 20 с. (Препринт / Дніпропетр.нац.ун-т).
7. Слэтер Дж. Электроника сверхвысоких частот: Пер с англ. - М.: Сов. радио, 1965. - 336 с.
8. Zhang D., Popov V., Podrazhansky Y., Arora P., White R. Cobalt doped chlorium oxides as cathode materials for secondary lithium batteries // J. of Power Sources. - 1999. - Vol. 83. - P. 121-127.
9. Подражанский Ю. М., Шембель Е. М. Влияние импульсных режимов заряда на характеристики аккумуляторов // Вопросы химии и химической технологии. - 2000. - № 1. - С. 202-205.
10. Патент № 4,829,225 США, МКИ H02J 7/04. Rapid battery charger, discharger and conditioner / Podrashansky Y., Popp Ph. W. (США); Electronic Power Devices Corp. - № 790,461. Заявл. окт. 23, 1985. Оpubл. май 9, 1989, НКИ 320/14. - 9с.
11. Патент № 5,307,000 США, МКИ H02J 7/10. Method and apparatus for charging, thawing, and formatting a battery / Podrashansky Y., Popp Ph. W. (США); Electronic Power Technology, Inc. - № 824,113. Заявл. янв.22, 1992. Оpubл. июнь 26, 1994, НКИ 320/14. - 14 с.
12. Патент № 5,504,415 США, МКИ H02J 7/00. Method and apparatus for automatic equalization of series-connected batteries / Podrashansky Y., Podrashansky M., Golod M. (США); Electronic Power Technology, Inc. - № 162,581. Заявл. дек. 3, 1993. Оpubл. июнь 2, 1996, НКИ 320/18. - 10 с.
13. Патент № 5,694,023 США, МКИ H01M 10/44. Control and termination of a battery charging process / Podrashansky Y., Tsender B. (США); Advanced Charger Technology, Inc. - № 677,483. Заявл. июль 10, 1996. Оpubл. дек. 2, 1997, НКИ 320/21. - 14 с.
14. Патент № 5,889,385 США, МКИ H02J 7/0. Equalization of series-connected cells of a battery using controlled charging and discharging pulses / Podrashansky Y., Podrashansky M., Kusharskiy Y. (США); Advanced Charger Technology, Inc. - № 914,674. Заявл. сент.19, 1997. Оpubл. март 30, 1999, НКИ 320/130. - 17 с.
15. А.с. № 838828 СССР, МКИ H01 M 10/44. Способ заряда свинцового кислотного аккумулятора / В.А.Шуляев (СССР). - Бюл. изобретений и товарных знаков - 1981, № 22 от 17.10.79.

УДК 537.876.23+621.372.81.09

СОБСТВЕННЫЕ ФУНКЦИИ И ЧИСЛА ПРЯМОУГОЛЬНОГО ВОЛНОВОДА С ФРАКТАЛЬНЫМИ ШИРОКИМИ СТЕНКАМИ

А.А.Мисюра, В.М.Онуфриенко

Наведено результати дослідження прямокутного металевого хвилеводу з фрактальними широкими стінками. Для знаходження власних функцій та чисел електромагнітних хвиль, що розповсюджуються у розглянутій структурі, застосований апарат дробового інтегро-диференціювання. Визначено залежність значень власних хвильових чисел від величини скейлінгового показника, що характеризує ступінь фрактальності стінок хвилеводу. Показано збіжність результатів розрахунку з класичними, що відповідають ідеальній провідності металевих стінок хвилеводу.

Представлены результаты исследования прямоугольного металлического волновода с фрактальными широкими стенками. Для нахождения собственных функций и чисел электромагнитных волн, распространяющихся в такой структуре, применен аппарат дробного интегро-дифференцирования. Определена зависимость значений собственных

волновых чисел от величины скейлингового показателя, характеризующего степень фрактальности стенок в волноводе. Показано совпадение результатов расчета с классическими, соответствующими идеальной проводимости металлических стенок волновода.

The results of research of rectangular metallic waveguide with fractional wide walls are presented in this paper. The own functions and numbers of electromagnetic waves, which propagate in this structure, have been determined using the apparatus of fractional integro-differential calculus. The dependence of own wave numbers on values of scaling index, which is characterizes the degree of fractionation of the walls in waveguide, has been defined. The coincidences of calculation results with classic, which are corresponding to ideal conduction of metallic waveguides walls, have been showed.

ВВЕДЕНИЕ

Для направляющих систем с идеальными проводящими поверхностями алгоритм решения краевых задач весьма прост и изложен во многих источниках (см., например, [1-2]). Однако на практике эффективное исследование большинства СВЧ устройств с помощью традиционного математического аппарата затруднительно, что обусловлено наличием геометрических фрактальных сингулярностей на их поверхностях. В связи с этим весьма актуально применение специальных математических методов, среди которых наиболее эффективными являются приемы фрактального анализа, в основе которых задействован аппарат дробного интегро-дифференциального исчисления [3].

В фундаментальной междисциплинарной книге Б. Мандельброта [4] развивается понятия фракталов и анализируется огромное число примеров их природных и математических применений. Другие книги, посвященные фракталам, включают рассмотрение геометрических аспектов трактовки фрактальной природы линий и поверхностей [5]; прекрасно иллюстрированный обзор комплексной динамики приведен в [6]; физические применения - в [7]; особенности введения фрактальных контуров и поверхностей в граничные условия задач теории поля и электродинамики - в [8-10]; математическое обоснование применения интегро-дифференциалов для описания фрактальных структур - [11]. Число публикаций по данной тематике непрерывно растет, но в то же время нет работ по использованию теории фракталов при исследовании направляющих систем сверхвысококачественной техники.

1 ПОСТАНОВКА ЗАДАЧИ

В дальнейшем рассмотрении фракталом будем называть структуру, состоящую из частей, которые в каком-то смысле подобны целому. Это понятие базируется на перечне его геометрических свойств, а именно: фрактал имеет тонкую структуру (т.е. имеет детали малых размеров); является сверх нерегулярным, чтобы быть описанным в рамках традиционной геометрической терминологии в местном масштабе и глобально; определенная некоторым образом фрактальная размерность больше его топологической размерности; в большинстве случаев теории поля определяется достаточно простым, возможно рекурсивным, способом [3].

Понятия фрактальной размерности D и фрактальной меры H^α базируются на определении Хаусдорфа-Безиковича, центральное место в котором занимает понятие расстояния между точками в пространстве. Для определения размерности фрактального объекта необходимо покрыть его простыми компактами (прямолинейными отрезками, квадратами, прямоугольниками, кругами, эллипсами и т.д.) и измерить расстояние между ними. Так при измерении с помощью покрытия кубами, величина хаусдорфовой размерности описывается выражением

$$D_H = \lim_{r \rightarrow 0} \frac{\ln N(r)}{\ln(1/r)}, \quad (1)$$

где r - сторона куба покрытия;

$N(r)$ - количество кубов, в которые попадает хотя бы одна точка фрактального множества.

Величину D_H сравнивают с показателем α дробного интегро-дифференциала, который будет введен при решении следующей задачи. Рассмотрим прямоугольный волновод, у которого широкие стенки фрактальные, т.е. нелинейные, шероховатые.

Поставив целью исследовать E и H - волны в таком волноводе, мы должны решить первую и вторую краевые задачи Гельмгольца. Рассмотрим вначале алгоритм их решения для идеализированного прямоугольного волновода, а затем применительно к нашему случаю.

2 ОПРЕДЕЛЕНИЕ ИНТЕГРО-ДИФФЕРЕНЦИАЛЬНЫХ ХАРАКТЕРИСТИК ПОЛЯ

Для исследования волн, распространяющихся в идеализированном прямоугольном волноводе, необходимо решить первую (для E -волн) и вторую (для H -волн) краевые задачи Гельмгольца [1].

Однородное уравнение Гельмгольца относительно α -характеристик компоненты $u^\alpha(x, y, z)$ поля [8] в декартовых координатах имеет вид:

$$\frac{\partial^2 u^\alpha}{\partial x^2} + \frac{\partial^2 u^\alpha}{\partial y^2} + \frac{\partial^2 u^\alpha}{\partial z^2} + k^2 u^\alpha = 0, \quad (2)$$

где $k = \frac{2\pi}{\lambda}$ - волновое число.

Решение (2) будем искать в форме $u^\alpha(x, y, z) = X(x) \cdot Y(y) \cdot Z(z)$, используя метод разделения переменных, т.е. приводя уравнение с частными производными к обыкновенным дифференциальным уравнениям. В итоге получаем:

$$\begin{aligned} X &= A \cos k_x \cdot x + B \sin k_x \cdot x, \\ Y &= C \cos k_y \cdot y + D \sin k_y \cdot y, \\ Z &= E \cos k_z \cdot z + F \sin k_z \cdot z, \end{aligned} \quad (3)$$

где A, B, C, D, E, F - неизвестные амплитудные множители, определяемые с помощью граничных условий; k_x, k_y, k_z - некоторые константы, подчиненные равенству $k_x^2 + k_y^2 + k_z^2 = k^2$.

В двумерном случае, когда решение не зависит от z , с учетом граничных условий первой краевой задачи определяем

$$E_z^{mn\alpha}(x, y) = E_0 \sin \frac{m\pi x}{a} \cdot \sin \frac{n\pi y}{b}, \quad (4)$$

где E_0 - постоянная или амплитудный коэффициент; a - размер широкой стенки волновода; b - размер узкой стенки волновода; $m = 1, 2, \dots, n = 1, 2, \dots$

Это собственные функции первой краевой задачи, которым при $\alpha = 0$ соответствуют собственные числа $k_{mn}^2 = k_x^2 + k_y^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2$. Таким образом, каждая из функций $E_z^{mn\alpha}$ характеризует распределение в поперечном сечении волновода продольной компоненты вектора E^α той или иной свободной волны типа E_{mn}^α .

Для второй краевой задачи в прямоугольной области, используя граничные условия

$$\frac{\partial H_z^\alpha(0, y)}{\partial x} = \frac{\partial H_z^\alpha(a, y)}{\partial x} = 0 \quad \text{и} \quad \frac{\partial H_z^\alpha(x, 0)}{\partial y} = \frac{\partial H_z^\alpha(x, b)}{\partial y} = 0,$$

находим следующие α -характеристики собственных функций этой задачи

$$H_z^{mn\alpha}(x, y) = H_0 \cdot \cos \frac{m\pi x}{a} \cdot \cos \frac{n\pi y}{b}. \quad (5)$$

При $\alpha = 0$ им отвечают прежние собственные числа $k_{mn}^2 = k_x^2 + k_y^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2$. Таким образом, каждая из функций $H_z^{mn\alpha}$ характеризует распределение в поперечном сечении волновода продольной компоненты вектора H^α свободного поля типа H_{mn}^α .

Для воспроизведения из α -характеристик собственных функций в случае прямоугольного волновода с фрактальными свойствами поверхностей используем аппарат дробного интегро-дифференцирования. Следовательно, находим не сами собственные функции $\varphi(\alpha)$, а их интегралы дробного порядка α (при $\alpha > 0$) $\varphi(\alpha)$ или дробные производные (при $0 < \alpha < 1$) f^α :

$$\varphi^\alpha = {}_a I_x^\alpha(\varphi(\alpha)(x)) = \frac{1}{\Gamma(\alpha)} \int_a^x \frac{f(\alpha)(t)}{(x-t)^\alpha} dt \quad \text{при } \alpha > 0, \quad (6)$$

$$f^\alpha = {}_a D_x^\alpha(f(\alpha)(x)) = \frac{1}{\Gamma(1-\alpha)} \cdot \frac{d}{dx} \int_a^x \frac{f(\alpha)(t)}{(x-t)^\alpha} dt \quad \text{при } 0 < \alpha < 1. \quad (7)$$

Согласно [12] результаты интегро-дифференцирования (6) и (7) можно записать в виде рядов или гипергеометрических функций

$$I_{a+}^\alpha(\sin k_x x) = \frac{1}{2i\Gamma(\alpha+1)} \cdot x^\alpha [{}_1F_1(1; (\alpha+1); ik_x^{(\alpha)} x) - {}_1F_1(1; (\alpha+1); -ik_x^{(\alpha)} x)], \quad (8)$$

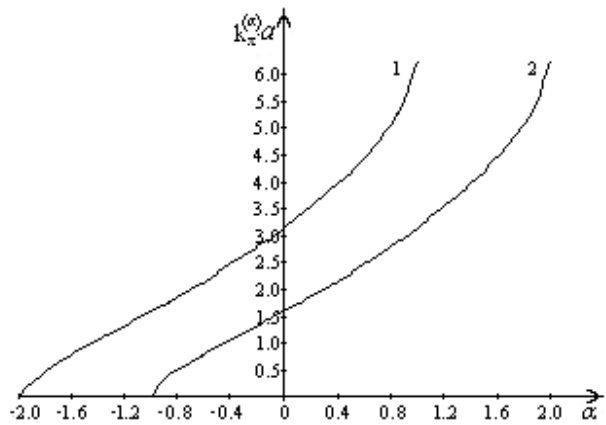
$$I_{a+}^\alpha(\cos k_x x) = \frac{1}{2\Gamma(\alpha+1)} \cdot x^\alpha [{}_1F_1(1; (\alpha+1); ik_x^{(\alpha)} x) + {}_1F_1(1; (\alpha+1); -ik_x^{(\alpha)} x)], \quad (9)$$

где ${}_1F_1(a; c; x) = \sum_{n=0}^{\infty} \frac{(a)_n}{(c)_n} \cdot \frac{x^n}{n!}$ - вырожденная гипергео-

метрическая функция Куммера: $(a)_n = a(a+1) \times \dots \times (a+n-1)$ - символ Похгаммера; $c \neq 0, -1, -2, \dots$ - параметр гипергеометрического ряда; x_n - переменная.

3 ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

Используя граничные условия для (8), (9), определим собственные числа $k_x^{(\alpha)}$, соответствующие вышеописанным собственным функциям $E(\alpha)$ и $H(\alpha)$ -волн, распространяющихся в такой структуре. Результаты численных расчетов собственных чисел и собственных функций электрических и магнитных волн в зависимости от различных значений скейлингового показателя α представлены на рис. 1-5.



1 - волна $E^{(\alpha)}$ -типа; 2 - волна $H^{(\alpha)}$ -типа

Рисунок 1 - Зависимость нормированного волнового числа $k_x^{(\alpha)}$ от дробного порядка α в случае широких фрактальных стенок

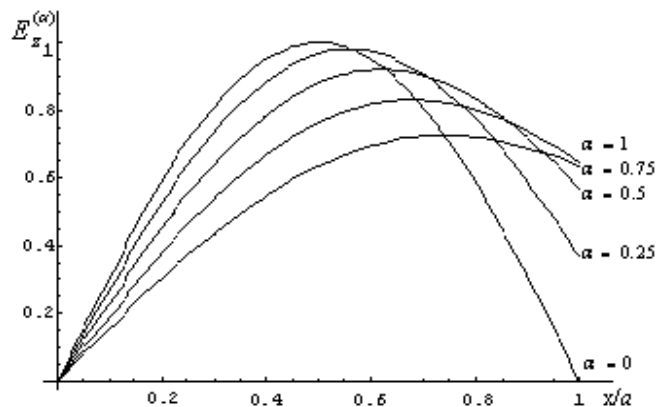


Рисунок 2 - Распределение компоненты $E_z^{(\alpha)}$ поля $E^{(\alpha)}$ -типа при различных положительных значениях дробного порядка α

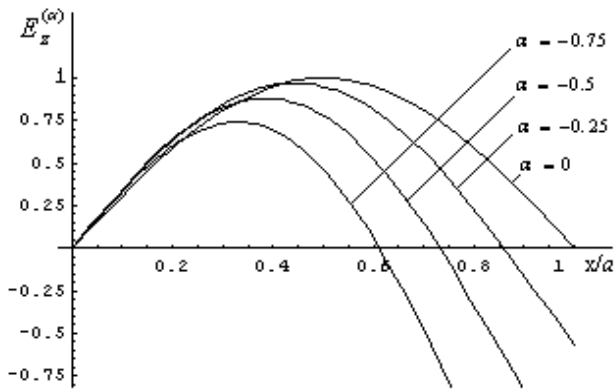


Рисунок 3 - Распределение компоненты $E_z^{(\alpha)}$ поля $E^{(\alpha)}$ -типа при различных отрицательных значениях дробного порядка α

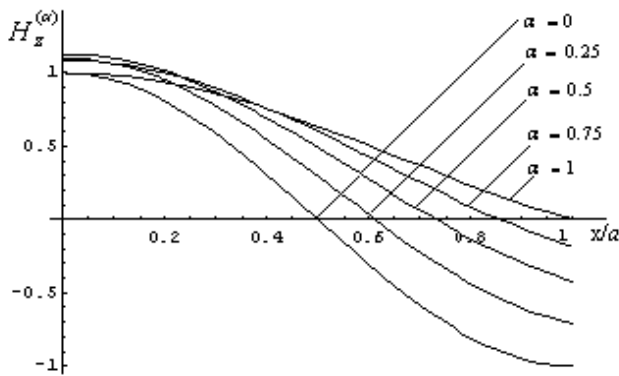


Рисунок 4 - Распределение компоненты $H_z^{(\alpha)}$ поля $H^{(\alpha)}$ -типа при различных положительных значениях дробного порядка α

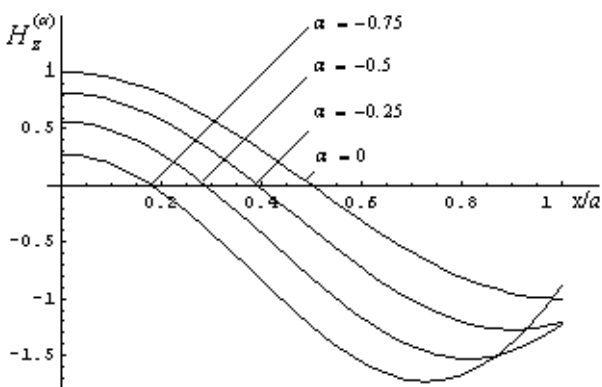


Рисунок 5 - Распределение компоненты $H_z^{(\alpha)}$ поля $H^{(\alpha)}$ -типа при различных отрицательных значениях дробного порядка α

Анализируя графики, представленные на рис.1, можно констатировать, что для волны $E^{(\alpha)}$ -типа (кривая 1) с увеличением порядка α от -2 до 1 и для волны $H^{(\alpha)}$ -типа

(кривая 2) с увеличением порядка α от -1 до 2 значение собственного волнового числа $k_x^{(\alpha)}$ возрастает от 0 до 2π , в случае $\alpha = 0$ значение волнового числа соответствует его значению для идеализированного волновода π и $\pi/2$ соответственно.

Так как волновое число является параметром, определяющим длину волны, распространяющейся в волноводе, следовательно, при его уменьшении ($\alpha < 0$) фактически увеличиваются размеры направляющей системы, и предоставляется возможность распространению высших типов волн. И, наоборот, при его увеличении ($\alpha > 0$) волноведущая система функционально сужается, что ограничивает распространение даже основного типа волны, как видно из рис. 2-5.

ВЫВОДЫ

Применением аппарата интегро-дифференциального исчисления определены собственные волновые числа и функции электромагнитных волн, распространяющихся в прямоугольном волноводе с широкими фрактальными стенками. Показана зависимость собственных волновых чисел от величины скейлингового показателя, характеризующего степень фрактальности стенок. Установлено, что благодаря такому строению поверхностей волновода возможно управление распространением тех или иных типов волн, что очень актуально при проектировании СВЧ фильтров. Предложенный математический аппарат позволяет описать волноведущую структуру произвольного сечения. Рассматриваемый подход позволяет определить такой важный для технических приложений параметр, как затухание и его зависимость от фрактальных свойств стенок волновода. Результаты настоящей работы могут использоваться для решения актуальных задач об управлении электромагнитным полем в волноведущих системах за счет варьирования скейлингового показателя, характеризующего степень фрактальности поверхностей.

ПЕРЕЧЕНЬ ССЫЛОК

1. Никольский В.В. Электродинамика и распространение радиоволн. Учеб. пособие для радиотехн. специальностей вузов. - М.: Наука, 1973. - 607с.
2. Нефедов Е.И., Сивов А.Н. Электродинамика периодических структур. - М.: Наука, 1977. - 208с.
3. Онуфриенко В.М. Диферентіальні α -форми у хаусдорфовій метриці на фрактальних множинах // Радіоелектроніка. Інформатика. Управління. - 2002. - №2 (8). - С. 31-35.
4. Mandelbrot B.B. The Fractal Geometry of Nature. Freeman.San Francisco, 1982. - 468p.
5. Falconer K.J. The Geometry of Fractal Sets. Cambr. Univ. Press, Cambridge, 1985. - 268p.
6. Peitgen H.-O. and Richter P.H. The Beauty of Fractals. Springer. Berlin, 1986. - 324p.
7. Федер Е. Фракталы: Пер. с англ. - М.: Мир, 1991. - 254с.
8. Onufriyenko V.M. On " α -features" of electrical waves above impedance plane// Proceedings 12 International. Conference on Microwaves & Radar. - V.1. - Krakov (Poland). - 1998. - PP. 212-215.
9. Onufriyenko V.M. Physical and Geometric Interpretation of Electromagnetic Field's α -Characteristics // Telecommunication and Radio Engineering. - V.53. - N 4-5, 1999. - PP. 136-139.
10. Onufriyenko V.M. The Differintegral Model for Describing Fractal Coupling between Waveguide Surfaces //Telecommunication and Radio Engineering, 2002. - V.57. - N 1. - PP. 30-36.

11. Onufriyenko V.M. Integro-Differential Charges and Currents Distribution on the Fractal Medium Topology // Conf. Proc. International Conference of Mathematical Methods in Electromagnetic Theory (MMET-2002). - V.2. - Kiev, Ukraine, 2002. -

PP. 382-384.
12. Самко С.Г., Килбас А.А., Маричев О.И. Интегралы и производные дробного порядка и некоторые их приложения. Минск: Наука и техника, 1987. - 688с.

УДК 548.53:548:536.4

СТРУКТУРНАЯ УСТОЙЧИВОСТЬ ТОНКИХ ОСТРОВКОВЫХ ПЛЕНОК. РАСЧЕТЫ И КОМПЬЮТЕРНЫЙ АНАЛИЗ

В.И.Псарев, Л.А.Пархоменко, А.Ф.Куликов

Наведено опис методу дисперсійно-комп'ютерного аналізу тонких островкових плівок з ілюстрацією його застосування. За його допомогою може бути встановлений можливий кореляційний зв'язок між ознаками змін експериментальних гістограм і внутрішньо-системними процесами, що викликають трансформацію розподілів островків за розмірами в процесі структурного огрублення плівок.

Приведено описание дисперсионно-компьютерного анализа тонких островковых пленок с иллюстрацией применяемого метода. С его помощью может быть установлена возможная корреляционная связь между признаками изменения со временем экспериментальных гистограмм и внутрисистемными процессами, вызывающими трансформацию распределений островков по размерам в процессе структурного огрубления пленок.

A description is given of the dispersion-computer analysis carried out on thin island films and the illustration of the technique used. A possible correlative relation among the time change signs of the experimental histograms and the intrasystem processes, which result in transformation of microislands size distribution in the process of their coarsening in thin films, can be established with the help of the technique.

Современный этап развития микроэлектроники характеризуется широким применением разного типа пленочных интегральных микросхем с обеспечением микроминиатюризации радиоэлектронной аппаратуры и совместного изготовления сразу многих элементов (сопротивления, конденсаторы, индуктивности, токоподводящие дорожки и т. п.). Качество и надежность работоспособного состояния микросхем, в процессе их функционирования, в особенности, под влиянием воздействия каких-либо внешних факторов, существенно зависят от структурной стабильности пленочного материала, достоверная информация о которой должна быть заранее получена методами физико-химического анализа. В научной литературе этому вопросу, фактически, не уделялось должного внимания. Рассмотрим в этой связи металлические пленки островкового типа, подверженные огрублению из-за оствальдовской коагуляции микроостровков, осложненной рядом сопутствующих процессов.

Формирование островковой пленки на подложке характеризуется образованием и ростом островков, а свободная поверхность между ними покрывается слоем адатомов (адсорбированных атомов) с разной плотностью в зависимости от ее микрорельефа. Степень выживаемости

островков (например, золота, платины, меди и др.) в такой дисперсной системе существенно зависит от их размера, огранки и структурного состояния. Переход пленки в более стабильное состояние происходит путем укрупнения островков и сопровождается значительным уменьшением свободной поверхностной энергии.

Перераспределение вещества островков при нагревании пленки определяется величиной их критического радиуса r_k . Микрочастицы-островки докритического размера $r < r_k$ постепенно расплываются по поверхности подложки, а расти будут только те из них, размер которых больше критического $r > r_k$. Такой процесс сопровождается уменьшением их общего числа, степени пересыщения адатами поверхности подложки, а значит, и ростом среднего r_s , критического r_k , модального r_m , наибольшего r_g размеров и размаха распределения $(0, r_g)$. Аналитически это условие выражается уравнением движения, а именно зависимостью скорости роста - растворения микрочастиц от их размера.

В случае дисперсной системы островков, имеющих форму сферического сегмента, предложена математическая модель уравнения движения [1, 2]

$$\frac{dr}{dt} = \frac{aD_sKV}{D_s + Kr^2} \left[\frac{1}{r_k} - \frac{1}{r} \right], \quad (1)$$

где r - пространственный размер островков (r_k - их критический радиус); D_s - коэффициент поверхностной диффузии адатомов; V - удельный объем вещества пленки; K - скорость присоединения адатомов к поверхности островков; a - постоянная величина [1, 2]. Отдельные частные решения, полученные на основе уравнения (1), соответствуют установившейся стадии оствальдовской коагуляции островков.

В настоящем сообщении, в свете изложенных в работах [1,2] представлений, произведено обобщение решения задачи и предложены аналитические разработки, предназначенные для системного анализа микроструктуры островковых пленок и опытной проверки полученных решений. С их помощью может быть установлена возможная корреляционная связь между признаками изменения со временем экспериментальных гистограмм и

внутрисистемными процессами, вызывающими трансформацию распределения островков по размерам в процессе структурного огрубления пленки.

1 ОБЩИЕ ТЕОРЕТИЧЕСКИЕ ПОЛОЖЕНИЯ

Двумерная задача об эволюции поверхностной системы островков, имеющих форму сферического сегмента, сводится к решению в трехмерной постановке [3]. Тем самым осуществляется ее осимметрирование путем ввода эквивалентного поверхностной системе трехмерного распределения микроостровков по их эффективному пространственным размерам. Введем безразмерную переменную $u = \frac{r}{r_k}$ и преобразуем уравнение (1), с учетом зависимости скорости K от размера островка [3]. Получим два варианта уравнения движения. Первый соответствует диффузионному механизму переноса вещества в системе и имеет вид

$$v(u) = v_k(t) \left[\frac{dr_k}{dt} \right]^{-1} \frac{dr}{dt} = \frac{u-1}{\gamma_1(u)}, \quad (2)$$

где $v_k(t) = \frac{dr_k}{dt} \frac{r_k^2}{aD_s V}$ - приведенная скорость изменения

r_k со временем t ; $\gamma_1(u) = u^3 + \varepsilon \cdot u^{1-\alpha}$; $\varepsilon = \frac{D_s \exp(mt)}{K_0 r_k^2}$.

Второй определяется величиной K - скоростью проникновения адатомов через межфазовую границу и соответствует реакционно-контролируемому механизму, описываемому уравнением

$$v(u) = v_k(t) \left[\frac{dr_k}{dt} \right]^{-1} \frac{dr}{dt} = \frac{u-1}{\gamma_2(u)}, \quad (3)$$

где $v_k(t) = \frac{dr_k r_k \exp(mt)}{dt a K_0 V}$; $\gamma_2(u) = \varepsilon_1 u^3 + u^{1-\alpha}$; $\varepsilon_1 = \varepsilon^{-1}$.

Выражение для производства функций распределения микрочастиц по их относительным размерам при заданном $v(u)$ имеет следующий вид [3, 4]

$$\varphi(u) = \frac{3v_k}{v_k u - v(u)} \exp \left[-3v_k \int \frac{du}{v_k u - v(u)} \right], \quad (4)$$

где $f(r, t) = \varphi(u) r_k^{-4}$ - функция плотности распределения микрочастиц по размерам в момент времени t ; $v_k = v_k(t)$ - приведенная скорость движения r_k в ансамбле размеров островков.

Постоянная интегрирования определяется из закона сохранения массы вещества островковой пленки, записанного в таком виде

$$M = \Delta + \frac{4}{3} \pi \alpha_1(\theta) \int_0^{r_g} \rho(r) f(r, t) r^3 dr, \quad (5)$$

где Δ - пересыщение адатомов на подложке; $\alpha_1(\theta) = 0,25(2 - 3 \cos(\theta) + \cos^3 \theta)$, θ - краевой угол ($r_s = r \sin \theta$, r_s - эффективный радиус островка на поверхности подложки); $\rho(r)$ - плотность вещества пленки.

При заданном $v(u)$ выражение (4) производит множество унимодальных функций распределения при условии [4]

$$v_k = v_k(t) = \frac{\partial}{\partial u} v(u) \Big|_{u_g} = \frac{v(u)}{u} \Big|_{u_g}, \quad (6)$$

где $u_g = \frac{r_g}{r_k}$ - верхняя граница относительных размеров островков. Их относительный модальный размер можно определить из соотношения [4]

$$4v_k(t) = \frac{\partial}{\partial u} v(u) \Big|_{u_m}. \quad (7)$$

Каждое из распределений множества (4) обладает собственными моментами относительно $u = 0$ и соотношением между ними. Последнее можно получить, решив уравнение [5]

$$dM_{nm}' + \left[\frac{4v_k - \frac{\partial}{\partial u} v(u)}{v_k u - v(u)} \right] M_{nm}' du = \left[\frac{n}{u} - \frac{m}{u_g - u} \right] M_{nm}' du, \quad (8)$$

где $M_{nm}' = u^n (u_g - u)^m \varphi(u)$ n и m - целочисленные или дробные положительные числа. В этой связи, при заданных из уравнений (2) и (3), с помощью (8) можно получить соотношение между смешанными моментами в таком виде

$$(n-3)v_k M_{nm} = nL_n + mL_m, \quad (9)$$

где $M_{nm} = \int_0^{u_g} u^n (u_g - u)^m \varphi(u) du$;

$$L_n = M_{n-1-\gamma(u),m} - M_{n-\gamma(u),m};$$

$$L_m = M_{n+1-\gamma(u),m-1} - v_k M_{n+1,m-1} - M_{n-\gamma(u),m-1}.$$

В терминах моментов записаны:

$$M_{n-\gamma(u),m} = \int_0^{u_g} M_{nm}' [u\gamma(u)]^{-1} du;$$

$$M_{n-\gamma(u),m} = \int_0^{u_g} M_{nm}' [\gamma(u)]^{-1} du;$$

$$M_{n+1-\gamma(u),m-1} = \int_0^{u_g} M_{nm}' u (u_g - u)^{-1} [\gamma(u)]^{-1} du;$$

$$M_{n+1,m-1} = \int_0^{u_g} M_{nm}' u(u_g - u)^{-1} du;$$

$$M_{n-\gamma(u),m-1} = \int_0^{u_g} M_{nm}' (u_g - u)^{-1} [\gamma(u)]^{-1} du.$$

Соотношение (9) обладает множеством признаков, которые могут быть положены в основу идентификации теоретических распределений с экспериментальными. Например, при $n = 3$ и $m \neq 0$ имеет место равенство $3|L_3| = m|L_m|$; при $n = 3$ и $m = 0$ справедливо $M_{2-\gamma(u)} = M_{3-\gamma(u)}$, где $\gamma(u) = \gamma_1(u)$ для уравнения движения (2) и $\gamma(u) = \gamma_2(u)$ - уравнения (3). Если экспериментальное распределение удовлетворяет соотношению (9) и вытекающим из него признакам, то оно совместимо с одним из теоретических распределений множества (4), подлежащим отысканию путем определения параметров α и ε .

2 ДИФФУЗИОННОЕ СТРУКТУРНОЕ ОГРУБЛЕНИЕ ПЛЕНКИ

Заменой из уравнения движения (2) значения в выражении (4) получим

$$\varphi(u) = \frac{3\nu_k \gamma_1(u)}{\nu_k u^4 + \nu_k \varepsilon u^{2-\alpha} - u + 1} \times \exp \left[-3\nu_k \int \frac{\gamma_1(u)}{\nu_k u^4 + \nu_k \varepsilon u^{2-\alpha} - u + 1} du \right], \quad (10)$$

где $\gamma_1(u) = u^3 + \varepsilon \cdot u^{1-\alpha}$.

Решение предусматривает определение u_g, ν_k, u_m при значениях $0 \leq \alpha \leq 1$ и $0 \leq \varepsilon \leq 1$, воспользовавшись соотношениями (6) и (7). Соответствующие уравнения имеют следующий вид [6]

$$3u_g^3 + \alpha - 4u_g^{2+\alpha} + (1-\alpha)\varepsilon u_g - \varepsilon(2-\alpha) = 0, \quad (11)$$

$$\nu_k = (u_g - 1)(\varepsilon + u_g^{2+\alpha})^{-1} u_g^{\alpha-2}, \quad (12)$$

$$u_m(\alpha\varepsilon + 3u_m^{1+\alpha}) + \varepsilon(1-\alpha) = 4\nu_k u_m^{2-\alpha}(\varepsilon + u_m^{2+\alpha}). \quad (13)$$

Для конкретных значений параметров α и ε , путем интегрирования (10), можно получить теоретическую функцию плотности распределения. Например, при $\alpha = 0$ и $0 \leq \varepsilon \leq 1$ аналитическое выражение для множества функций распределения имеет следующий вид

$$\varphi(u) = \frac{c\gamma(u)(u_g - u)^{-a} \exp \left[-\frac{b}{u_g - u} \right]}{(u^2 + 2u_g u + \varepsilon + 3u_g^2)^c \exp \left[\frac{d}{\Delta_1} \arctg \left[\frac{u + u_g}{\Delta_1} \right] \right]}, \quad (14)$$

где $a = \frac{5\varepsilon^2 + 39\varepsilon u_g^2 + 114u_g^4}{(\varepsilon + 6u_g^2)^2}$; $b = \frac{3u_g(\varepsilon + u_g^2)}{\varepsilon + 6u_g^2}$;

$c = 0,5(7-a)$; $d = \frac{u_g(6u_g^4 - 15\varepsilon u_g^2 - 6\varepsilon^2)}{(\varepsilon + 6u_g^2)^2}$;

$\Delta_1 = (\varepsilon + 2u_g^2)^{1/2}$.

При $\alpha = \varepsilon = 0$ имеем $u_g = \frac{4}{3}$; $a = \frac{19}{6}$; $b = \frac{2}{3}$; $c = \frac{23}{12}$; $d = \frac{2}{9}$; $\Delta_1 = \frac{4}{3}\sqrt{2}$. Это распределение Чакраверти [1], характеризуемое коэффициентом асимметрии $S_k = -1,311$ ($S_k = \mu_3/\sigma^3$, где μ_3 - центральный момент третьего порядка, σ - дисперсия распределения).

Отрицательная асимметрия характерна и для других распределений, соответствующих диффузионному механизму укрупнения островков. В этом случае максимумы кривых плотности распределения смещены в сторону более крупных размеров островков. Такая дисперсная система должна характеризоваться малой устойчивостью к сохранению высокой степени дисперсности островков. У пленок этого типа должна существовать тенденция к переходу от дискретного структурного строения (островкового) к сплошному строению. При сохранении дискретности пленка должна отличаться значительным контрастом в размерах островков.

3 РЕАКЦИОННОЕ СТРУКТУРНОЕ ОГРУБЛЕНИЕ ПЛЕНКИ

После подстановки (2) в выражение (4), получим

$$\varphi(u) = \frac{3\nu_k \gamma_2(u)}{\nu_k \varepsilon_1 u^4 + \nu_k u^{2-\alpha} - u + 1} \times \exp \left[-3\nu_k \int \frac{\gamma_2(u)}{\nu_k \varepsilon_1 u^4 + \nu_k u^{2-\alpha} - u + 1} du \right]. \quad (15)$$

При заданных α и ε_1 и выполнении условия (6) можно определить значения u_g и ν_k из уравнений

$$3\varepsilon_1 u_g^3 + \alpha - 4\varepsilon_1 u_g^{2+\alpha} + (1-\alpha)u_g - 2 + \alpha = 0, \quad (16)$$

$$\nu_k = (3u_g - 4)(2 + \alpha)^{-1} u_g^{\alpha-2}. \quad (17)$$

Каждой паре этих величин соответствует решение уравнения (15) - унимодальная функция плотности распределения островков по их относительным размерам. Распределения достигают максимума при значениях u_m , которые можно определить из уравнения

$$u_m(\alpha + 3\varepsilon_1 u_m^{1+\alpha} - 2\varepsilon_1 u_m^{2+\alpha}) + 1 - \alpha = 4\nu_k u_m^{2-\alpha} (1 + \varepsilon_1 u_m^{2+\alpha}). \quad (18)$$

Коэффициент асимметрии, производимых выражением

(15) распределений, принимает положительные, нулевые и отрицательные значения, характерные и для природной формы распределений островков по размерам.

Рассмотрим, например, решение при $\alpha=1/2$ и $\varepsilon_1=0$. Из уравнений (16) и (17) следует: $u_g=3$ и $v_k=0,3849$.

Произведем замену переменных: $x = u^{1/2}$, то $M_{nm} = x^{n+1}(x_g-x)^m\varphi(x)$ и $\gamma_2(x) = x$. Решая (15), получим

$$\varphi(x) = cx(x_g-x)^{-22/3} \left\{ x + \frac{1}{2}x^{\frac{1}{3}} \right\}^{-5/3} \exp\left\{ -\frac{4x_g}{x_g-x} \right\}, \quad (19)$$

где $x_g = u_g^{1/2} = 1,7321$.

В этом случае замена переменных означает замену признака, по которому осуществляется распределение островков. Следовательно, соотношение между моментами распределения (9) также преобразуется к виду

$$(6-n)v_k M_{nm} = n(M_{n-3,m} - M_{n-1,m}) - m(v_k M_{n+1,m-1} - M_{n,m-1} + M_{n-2,m-1}), \quad (20)$$

где $M_{nm} = \int_0^{x_g} x^{n+1}(x_g-x)^m\varphi(x)dx = \int_0^{x_g} M'_{nm}dx$; другие

моменты также зависят от x .

Заметим, в пределах каждого размерного интервала островки могут сочетаться по какому-либо другому признаку при $\alpha = 1/2$ иначе, чем при $\alpha = 2/4$ или же при любом другом $\alpha = \frac{n}{2n}$, где $n = 1, 2, 3, \dots$

Соответственно изменяется и аналитический вид функции плотности распределения. Например, при $\alpha = 3/6$ и $\varepsilon = 0$ имеем: $x_g = \sqrt[6]{u_g} = 1,2009$; $v_k = 0,3849$; $x = u^{1/6}$; $\gamma(x) = x^3$, тогда

$$\varphi(x) = cx^3(x_g^3-x^3)^{-22/3} \left\{ x^3 + \frac{1}{2}x^{\frac{3}{2}} \right\}^{-5/3} \exp\left\{ -\frac{4x_g^3}{x_g^3-x^3} \right\}.$$

Однако соотношение между моментами (20) сохраняется и для этой функции распределения.

Информацию о природе и особенностях протекания внутрисистемных процессов можно получить путем идентификации теоретических распределений - решений уравнения (4) с образом - экспериментальными гистограммами с последующим выявлением их сходства и различия. Методика подобного анализа описана в работе [3]. Воспользуемся ею при анализе экспериментальных распределений платиновых островков на $\gamma-Al_2O_3$ подложке по их эффективным радиусам [7].

Таблица 1 - Количественные характеристики гистограмм островковой пленки Pt на $\gamma-Al_2O_3$ - подложке после нагревания при $700^\circ C$ на воздухе [7]

| 1 час | | | 16 часов | | |
|--|--------------------------------------|---------------------------------------|---|--------------------------------------|---------------------------------------|
| $r \cdot 10^6, \text{ мм}$ | $n_s \cdot 10^{-7}, \text{ мм}^{-2}$ | $N_r \cdot 10^{-12}, \text{ мм}^{-3}$ | $r \cdot 10^6, \text{ мм}$ | $n_s \cdot 10^{-7}, \text{ мм}^{-2}$ | $N_r \cdot 10^{-12}, \text{ мм}^{-3}$ |
| 0 | 0 | 0 | 0 | 0 | 0 |
| 2,14 | 2,68 | 2,002 | 1,69 | 1,70 | 0,011 |
| 5,00 | 15,12 | 3,253 | 3,93 | 6,50 | 1,610 |
| 7,86 | 31,36 | 17,714 | 6,17 | 11,15 | 5,465 |
| 10,72 | 13,44 | 7,651 | 8,41 | 10,35 | 4,874 |
| 13,58 | 0,84 | 0,390 | 10,65 | 9,00 | 4,731 |
| 16,44 | 0,28 | 0,136 | 12,89 | 4,50 | 2,134 |
| 19,30 | 0 | 0 | 15,13 | 2,70 | 1,455 |
| | | | 17,37 | 0,45 | 0,243 |
| | | | 19,61 | 0 | 0 |
| $n_0 = 69,72 \cdot 10^7 \text{ мм}^{-2}$; $\bar{r}_s = 7,18 \cdot 10^{-6} \text{ мм}$; $N = 31,15 \cdot 10^{12} \text{ мм}^{-3}$; $\bar{r} = 8,0 \cdot 10^{-6} \text{ мм}$; $r_m = 6,78 \cdot 10^{-6} \text{ мм}$; $r_k = 9,26 \cdot 10^{-6} \text{ мм}$; $r_g = 22,3 \cdot 10^{-6} \text{ мм}$; $u_g = 2,408$; $\alpha = 0,50$; $\varepsilon = 1,02 \cdot 10^{-2}$; $v_k = 0,345$; $S_k = -0,457$; $ex = 1,115$; $M_{3-\gamma_2(u)} = 0,7957$; $M_{3-u\gamma_2(u)} = 0,8225$; $g = 26,25 \%$ | | | $n_0 = 46,35 \cdot 10^7 \text{ мм}^{-2}$; $\bar{r}_s = 8,34 \cdot 10^{-6} \text{ мм}$; $N = 20,51 \cdot 10^{12} \text{ мм}^{-3}$; $\bar{r} = 9,05 \cdot 10^{-6} \text{ мм}$; $r_m = 7,13 \cdot 10^{-6} \text{ мм}$; $r_k = 11,03 \cdot 10^{-6} \text{ мм}$; $r_g = 26,6 \cdot 10^{-6} \text{ мм}$; $u_g = 2,4124$; $\alpha = 0,60$; $\varepsilon = 1,36 \cdot 10^{-2}$; $v_k = 0,363$; $S_k = 0,422$; $ex = -0,452$; $M_{3-\gamma_2(u)} = 0,7412$; $M_{3-u\gamma_2(u)} = 0,7604$ $g = 18,69 \%$ | | |

электронов или же высвечиванием).

ВЫВОДЫ

В таблице 1 приведены данные частотной заселенности размерных интервалов островками на единице площади поверхности подложки, а также соответствующее эквивалентное число микрочастиц в единице объема (N_r). Переход от $n_s = f(r)$ к эквивалентному распределению $N_r = f(r)$ производился с помощью формулы (7), приведенной в работе [8]. В результате компьютерных расчетов наряду со средним \bar{r} , модальным r_m и наибольшим r_g размерами микрочастиц (в эквивалентном объеме) определены численные значения критического радиуса r_k островков.

В процессе нагревания пленки происходит укрупнение островков, уменьшается их число на единицу площади подложки (n_0) и увеличивается их средний радиус (\bar{r}_s). Аналогичные изменения имеют место и для характеристик системы эквивалентных микрочастиц. Характеризующая ее функция плотности распределения со временем изменяет свою асимметрию с отрицательной ($S_k < 0$) на положительную ($S_k > 0$) и форму кривой возле точки максимума (см. значение эксцесса (ex)). Все это связано с накоплением в системе большого количества островков докритического размера ($r < r_k$) в результате значительного смещения r_k к более крупным размерам и увеличения параметра α . Выдержка пленки в течение 16 часов при 700°C привела к уменьшению доли растущих островков (g). Этот факт зафиксирован также и в работе [7]. Он должен приводить к увеличению степени заселенности докритических размерных интервалов. Методика определения r_k и g и оценка достоверности их численных значений описана в работе [3].

Обозначения: n_s - число островков радиуса r на единице поверхности подложки; N_r - число эквивалентных микрочастиц в единице объема; n_0 - общее число островков на единице площади подложки; \bar{r}_s - их средний радиус; N - общее число эквивалентных микрочастиц в единице объема; \bar{r} - их средний радиус. О других обозначениях см. в тексте.

Рассмотренный механизм укрупнения микрочастиц не исключает возможности существования островковых систем со стабильной структурой ([9]). В этом случае функция распределения должна стремиться к равновесной, характеризующей невариантное устойчивое состояние. Такие пленки подводимую энергию способны переводить в излучение (путем холодной эмиссии

Предложен метод системного анализа эволюции островковых пленок за счет их структурного огрубления - оствальдовской коагуляции микрочастиц-островков, осложненной разного рода сопутствующими процессами. Для этой цели получены теоретические решения - функции плотности распределения островков по размерам с положительной, нулевой и отрицательной внешней асимметрией и соотношения между их моментами для разных механизмов структурного огрубления пленок. С их помощью может быть установлена возможная корреляционная связь между признаками экспериментальных распределений-гистограмм (со временем термической обработки пленки или же продолжительности ее работы, под влиянием действия внешних факторов: давления, вибрации, радиационного воздействия и др.) и внутривещных процессами, вызывающими трансформацию распределения островков по размерам.

Достоверность результатов такого системного анализа оценивается с помощью соотношения между моментами функции распределения островков по размерам. Проведение анализа и использование предлагаемых расчетов в практике пленочного материаловедения возможны при наличии ЭВМ и соответствующих программных разработок.

ПЕРЕЧЕНЬ ССЫЛОК

1. Chakraverty В.К. Grain size distribution in thin films. 1. Conservative system // Journ. Phys. Chem. Solids. - 1967. - V.28. - P. 2401 - 2412.
2. Венгреневич Р.Д., Псарев В.И. К теории коалесценции островков в тонких пленках // Изв. вузов. Физика. - 1973. - №7. - С.149 - 151.
3. Псарев В.И. Компьютерный анализ металлических островковых пленок // Металлы. - 1999. - № 6. - С. 105 - 110.
4. Псарев В.И. Морфогенезис распределений микрочастиц по размерам при огрублении дисперсных систем // Изв.вузов.Физика. - 1990. - №12. - С.53 - 58.
5. Псарев В.И. Проблема моментов распределений в статистической физике // Изв.вузов.Физика. - 1997. - №4. - С.92 - 97.
6. Псарев В.И., Пархоменко Л.А. Кинетика огрубления и стабилизация островковых тонких пленок // Изв. вузов. Физика. - 1996. - № 9. - С. 128. Деп. в ВИНТИ 20.06.96, № 2035 - В.96.
7. Wynblatt P., Gjostein N.A. A model study of catalyst particle coarsening // Scripta metallurgica. - 1973. - V. 7. - P. 969 - 975.
8. Псарев В.И., Куликов А.Ф., Пшенцов С.И. Влияние межфазной поверхностной энергии на процесс коагуляции микрочастиц при нагревании металлических сплавов // ПОВЕРХНОСТЬ. Физика, химия, механика. - 1985. - № 12. - С. 22 - 27.
9. Псарев В.И. Об устойчивом распределении микрочастиц по размерам в твердых дисперсных системах // Изв. вузов. Физика. - 1974. - № 7. - С.144 - 146.

СИСТЕМА МОДЕЛИРОВАНИЯ ФРАКТАЛОПОДОБНЫХ РАССЕЙВАЮЩИХ СТРУКТУР

С.Н.Романенко, Л.М.Карпуков, Р.Д.Пулов

В роботі наведено результати моделювання розсіюючих конструкцій, які створені на основі фракталоподібних структур. Розрахунки виконувались за допомогою автоматизованої системи, яка дозволяє інтерактивно змінювати геометрію структури, вивчати розподіл струмів і аналізувати діаграму спрямованості в різних системах координат. Проведено порівняння результатів моделювання з даними експериментальних вимірювань. Показано, що властивості фракталів можуть бути використані при мініатюризації антен.

В работе представлены результаты моделирования рассеивающих конструкций, построенных на основе фракталоподобных структур. Расчеты выполнялись с помощью автоматизированной системы, которая позволяет интерактивно изменять геометрию структуры, изучать распределение токов и анализировать диаграмму направленности в различных используемых системах координат. Проведено сравнение результатов моделирования с данными экспериментальных измерений. Показано, что свойства фракталов могут быть использованы для миниатюризации антенны.

In the work the results of modeling of scattering constructions based on the fractal-like structures are presented. The calculations were accomplished with the help of the CAD system, which allows interactive changing of structure geometry and analyzing of current distribution and antenna diagram in different coordinate systems. The results of modeling are compared with experimental measurements. It was shown that properties of fractals may be used for miniaturization of antennas.

ВВЕДЕНИЕ

Теория фракталов в последнее время находит широкое применение при описании различных физических явлений в природе, например, турбулентности, облаков и зон дождей, разрушения горных пород, высоты и формы морских волн, пробоя в диэлектрике, размерности ландшафтов и параметров окружающей среды, структур нефтегазовых месторождений, акустических явлений, свойств шаровой молнии и разряда молний в атмосфере, биологических и информационных процессов и т.д. [1-3]. Ряд работ (см., например, [4, 5]) посвящены изучению фрактальных свойств рассеивающих поверхностей в строгой электродинамической постановке.

Регулярные фрактальные модели обычно строят на основе различных математических алгоритмов с использованием, как правило, итеративных процедур генерации самоподобных форм при последовательном измельчении масштаба.

В качестве примера на рис. 1 показаны три последовательных этапа итеративного построения широко известных: а) фрактальной кривой Минковского, б) фрактальной кривой фон Кох и в) треугольника Серпинского.

На n -м этапе построения длина кривой Минковского вычисляется по формуле

$$L_n = \left[1 + \frac{2}{3}\right]^n L_{n-1}, \quad (1)$$

где $0 \leq w \leq 1$ - высота прямоугольного выступа на первом шаге построения.

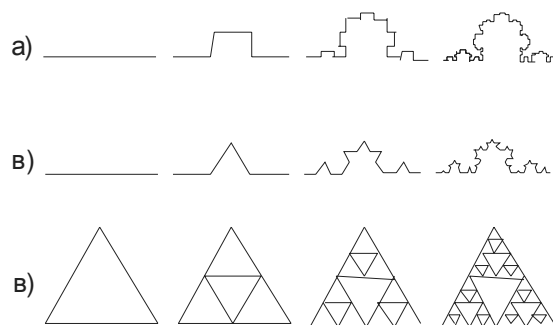


Рисунок 1 - Примеры построения фракталов

Для кривой фон Кох аналогичная формула имеет вид

$$L_n = \left[\frac{3}{4}\right]^n L_0, \quad (2)$$

где L_0 - длина первоначального (образующего) прямолинейного отрезка.

В случае треугольника Серпинского сумма длин периметров всех входящих в структуру на n -м шаге треугольников вычисляется по формуле

$$P_n = 2 \left[\frac{3}{2}\right]^{n+1}. \quad (3)$$

При этом длина периметра начального треугольника равна 1.

Из приведенных примеров видно, что при неограниченном увеличении числа шагов n длины периметров в формулах (1), (2) и (3) также неограниченно растут по степенному закону. При предельном переходе получаются самоподобные фрактальные кривые, которые обладают свойством локальной масштабной инвариантности или скейлингом. Характерной особенностью этих кривых является разрывность в каждой точке

и, как следствие, недифференцируемость. Кроме того, в отличие от гладких кривых, фрактальные кривые имеют нецелые значения размерностей. В частности, для приведенных на рис.1 кривых размерности D имеют следующие значения [1]: а) $-D = \ln(5)/\ln(3)$ при $w = 1$, б) $-D = \ln(4)/\ln(3)$, в) $-D = \ln(3)/\ln(2)$.

Приведенные рисунки показывают также одно из наиболее интересных геометрических свойств фракталов - свойство фрактальной линии плотно заполнять область малых размеров. При этом длина линии, как уже отмечалось, может быть очень большой (в пределе - бесконечно большой). Поэтому при разработке фракталоподобных антенн существенное влияние на размеры и характеристики излучения антенны будет оказывать степень упаковки фрактальной кривой и ее электрическая длина.

При практической реализации фракталов (в виде линий, поверхностей или объемов) невозможно создать бесконечно глубокую самоподобную структуру, какой является регулярный фрактал. Поэтому в процессе итеративного построения фрактала обычно ограничивают число шагов n самоподобного преобразования исходя из масштаба глубины анализа объекта и с учетом технологических и других физико-механических ограничений. Получающаяся в результате такого ограничения структура называется предфракталом, топологические свойства которого в рамках используемого масштаба мало отличаются от свойств истинного фрактала.

В данной работе представлены результаты компьютерного моделирования некоторых фракталоподобных рассеивающих структур, представляющих интерес для разработки малогабаритных антенн со специфическими характеристиками. Необходимо отметить, что высокая сложность геометрии таких структур предъявляет повышенные требования к эффективности математического аппарата алгоритмов и процедур автоматизированного моделирования.

В настоящей работе моделирование излучающих фракталоподобных структур производится на основе интегрального уравнения Поклингтона, решение которого осуществляется методом моментов с аппроксимацией токов кусочно-постоянными функциями. С этой целью разработана система автоматизированного моделирования, основанная на представлении излучающей поверхности решетчатой моделью. Такая модель отличается универсальностью и простотой описания геометрии элементов фракталоподобных антенн.

ОСНОВНАЯ ЧАСТЬ

Уравнение Поклингтона, используемое в системе для расчета распределения токов в сегментах решетчатой модели, имеет вид [7]:

$$\int_{-L/2}^{L/2} \left[\frac{\partial^2 G(z, z')}{\partial z^2} + k^2 G(z, z') \right] dz' = -j\omega \epsilon E_z^i(z), \quad (4)$$

где L - длина проводника, $G(z, z')$ - функция Грина свободного пространства, E_z^i - падающее или наведенное поле, z - точка наблюдения, z' - точка источника.

Расчет напряженности электрического поля от токов в проволочных сегментах решетчатой модели производится в цилиндрической системе координат по соотношениям:

$$E_z = \frac{1}{4\pi j k \sqrt{\epsilon}} \int_{-L/2}^{L/2} [(1 + jkr)(2r^2 - 3a^2) + k^2 a^2 r^2] \frac{e^{-jkr}}{r^3} dz', \quad (5)$$

$$E_\rho = -\frac{\rho I}{4\pi j k \sqrt{\epsilon}} \left. \frac{e^{-jkr}}{r^3} (1 + jkr) \right|_{z=-L/2}^{z=L/2}. \quad (6)$$

Здесь $r = \sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2}$, (x, y, z) - координаты точки наблюдения, (x', y', z') - координаты источника, $k = \frac{2\pi}{\lambda}$ - волновое число, I - амплитуда тока в сегменте, ρ - проекция r на плоскость xy .

В качестве источников возбуждения используются дельта источник стороннего напряжения, линейный диполь с заданным током и плоский кольцевой магнитный ток [8].

Интеграл (5) вычисляется с использованием квадратур Ромберга с автоматическим выбором шага интегрирования [9]. Это обеспечивает хорошую точность во всем диапазоне изменения расстояний.

Система имеет дружественный интерфейс, широкий набор сервисных функций, позволяет строить диаграммы направленности (ДН) в различных системах координат, вычислять входной импеданс антенны, а также рассчитывать распределение токов в сегментах решетчатой модели, которое может быть представлено в табличной или графической форме.

Тестирование системы осуществлялась путем расчета ДН известных антенн и сравнения с данными эксперимента. В качестве тестовой была проанализирована 15-ти элементная антенна, состоящая из вибратора, рефлектора и 13 направляющих элементов. Размеры элементов антенны: вибратор - $0,47\lambda$, рефлектор - $0,5\lambda$, направляющие элементы - $0,408\lambda$. Расстояние между вибратором и рефлектором - $0,25\lambda$, между направляющими элементами - $0,34\lambda$. Рабочая частота - 300 МГц.

На рис. 2 показана геометрия антенны и ее ДН в различных системах координат.

Результаты расчета диаграммы по азимуту и по углу места с графической точностью совпадают с данными работы [7].

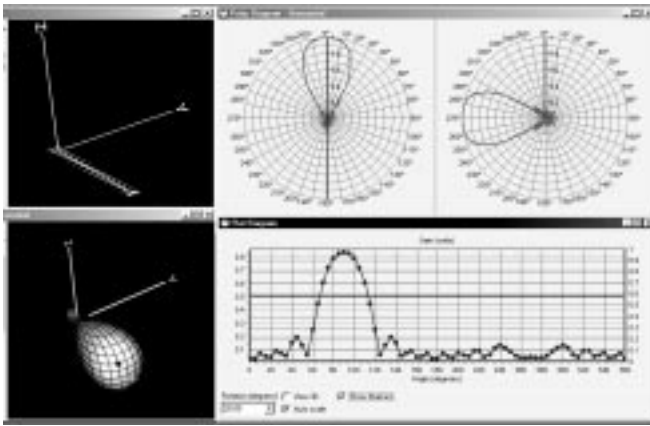


Рисунок 2 - ДН 15-ти элементной антенны в различных системах координат

С помощью системы моделирования проведены исследования рамочной антенны с длиной λ при использовании преобразования по кривой Минковского. На рис. 3 показана геометрия рамочных антенн, а на рис. 4 и 5 представлены результаты их анализа в виде ДН в различных системах координат.

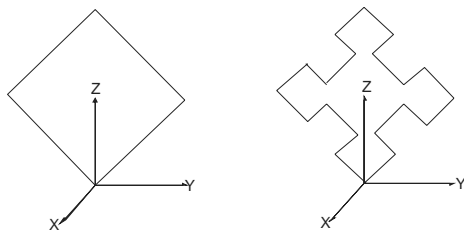


Рисунок 3 - Геометрия рамочных антенн

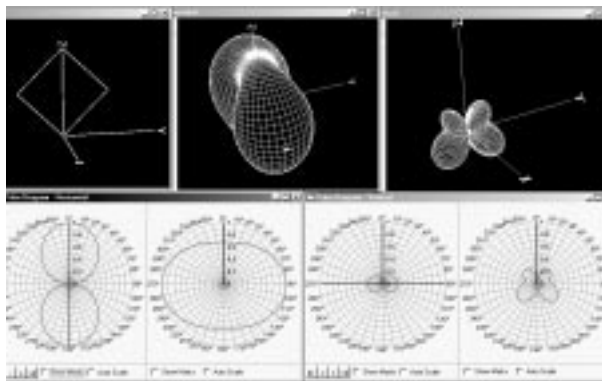


Рисунок 4 - ДН рамочной антенны для горизонтальной и вертикальной поляризации

Результаты моделирования совпадают с экспериментальными данными из [6].

На рис. 6 показана геометрия вибратора длиной $\lambda/2$ и треугольной рамочной антенны после третьего шага преобразования по кривой фон Кох. Результаты моделирования этих антенн представлены на рис. 7 и 8 соответственно.

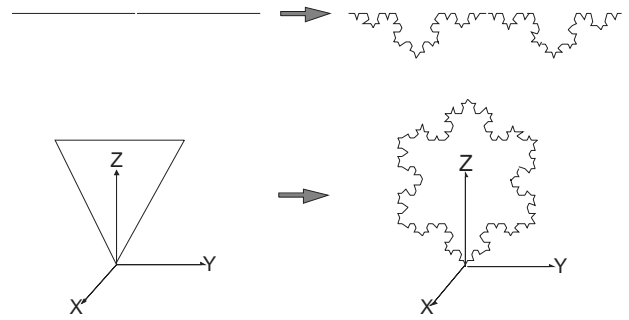


Рисунок 6 - Геометрия антенн

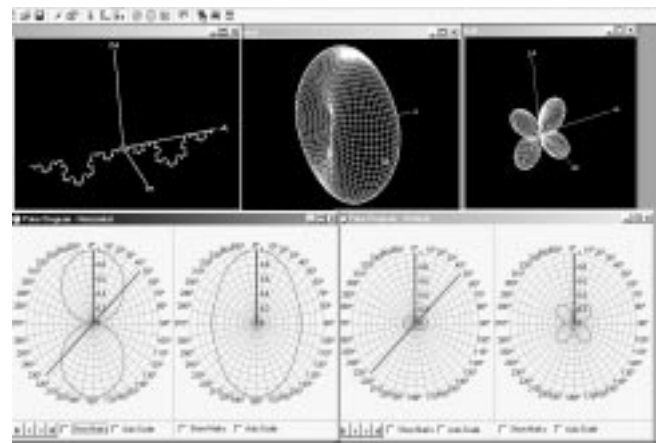


Рисунок 7 - ДН $\lambda/2$ вибратора на третьем шаге трансформации по кривой фон Кох

В верхней части рис. 7 приведены объемные ДН, а в нижней части - ДН в полярных координатах. Слева показаны диаграммы для горизонтальной поляризации, а справа - для вертикальной.

На рис. 8 представлены аналогичные результаты для рамочной антенны.

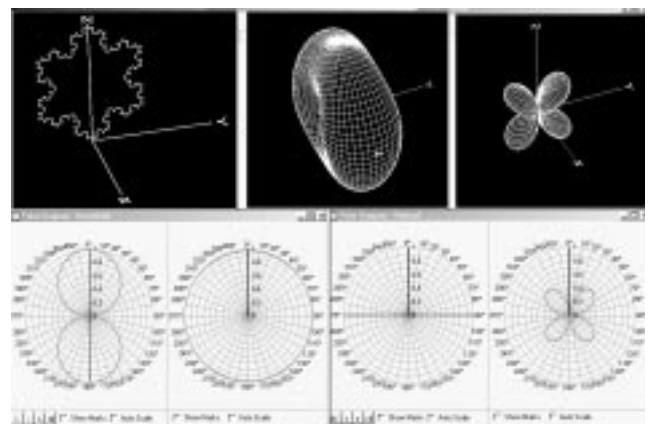


Рисунок 8 - Третий шаг трансформации рамочной антенны по кривой фон Кох

Результаты численного моделирования вибратора и рамочной антенны показывают, что с увеличением номера итерации преобразования фрактальной кривой уменьшается резонансная частота антенны при неизменной форме ДН.

Данные моделирования представленных структур сведены в таблицу, в которой λ соответствует резонансной длине волны антенны, а периметр равен полной длине фрактальной кривой на данном шаге трансформации.

Таблица 1

| Фрактальная размерность | Номер итерации | Отношение периметр/ λ вибратора | Отношение периметр/ λ рамки |
|-------------------------|----------------|---|-------------------------------------|
| 1,0000 | 0 | 0,450 | 1,364 |
| 1,2619 | 1 | 0,546 | 1,272 |
| 1,2619 | 2 | 0,437 | 0,712 |
| 1,2619 | 3 | 0,513 | 0,770 |

Из результатов моделирования также следует, что по мере возрастания степени изрезанности структуры уменьшаются амплитуды реактивных токов в линейных сегментах, составляющих фрактальную кривую.

ЗАКЛЮЧЕНИЕ

Разработана система автоматизированного моделирования фракталоподобных антенн, отличающаяся универсальностью, простотой освоения и удобством в эксплуатации. Система имеет дружественный интерфейс, широкий набор сервисных функций как для построения

геометрии антенны, так и для анализа результатов расчета.

Проведено исследование рамочной и вибраторной антенн на основе фрактальных кривых Минковского и фон Кох. Отмечено уменьшение резонансных частот антенн с увеличением степени трансформации кривой при сохранении формы ДН.

Дальнейшее изучение свойств фрактальных излучающих структур с помощью разработанной автоматизированной системы предполагает детальное исследование распределений токов и резонансных характеристик таких объектов.

ПЕРЕЧЕНЬ ССЫЛОК

1. Falconer K. J., The Geometry of Fractal Sets. Cambridge University Press, Cambridge, 1985. - 287 p.
2. Мандельброт Б. Фракталы в физике / Пер. с англ. под ред. Синая Я.Г. и Халатникова И.М. - М.: Мир, - 1988. - 544с.
3. Федер Е. Фракталы. - М.: Мир, 1991. - 478с.
4. Engheta N. On Fractional Calculus and Fractional Multipoles in Electromagnetism, IEEE Trans. on Antenna and Prop., vol. 44, No.4. - 1996. - pp. 554-566.
5. Онуфрієнко В.М. Взаємодія плоскої електромагнітної хвилі з металізованою фрактальною поверхнею // Радиофизика и электроника. - Харьков: Ин-т радиофизики и электроники НАН Украины. - Т4, №2. - 1999. - С.19-22.
6. Gianvittorio John P., Yahya Rahmat-Samii, Fractal Antennas: A Novel Antenna Miniaturization Technique and Applications, IEEE Antennas and Propagat., vol. 44, No.1, February, 2002. - pp. 20-36.
7. Computer techniques for electromagnetics. Edited by R. Mitra. Oxford, New York, Toronto, Sydney, Braunschweig: Pergamon press, 1973. - 488 p.
8. Tsai L., A numerical solution for the near and far fields of an annular ring of magnetic current, IEEE Trans. Antennas Propagat., vol. AP-20, May, 1972. - pp. 569-576.
9. Форсайт Дж., Малькольм М., Моулер К., Машинные методы математических вычислений. - М.: Мир, 1980. - 280 с.

УДК 681.32:007

ЭЛЕКТРОСТИМУЛЯТОР ГЛАЗНЫХ МЫШЦ ЭКСПОНЕНЦИАЛЬНЫМИ МОДУЛИРОВАННЫМИ ИМПУЛЬСАМИ

Е.Я.Швец, А.Л.Веревкин, О.Н.Поправка, А.П.Посунько

Розглядаються питання розробки і використання мініатюрних імпульсних електростимуляторів очних м'язів для офтальмології. Розроблено прилад, що являє собою генератор імпульсів із несучою середньою частотою 5000 Гц, фронти яких змінюються по експоненціальному законі. Період повторення імпульсів регулюється в межах від 5 до 100 Гц. Прилад виготовлений за технологією гібридних мікробірок, що дозволило значно зменшити його габаритні розміри і зручність для використання в офтальмології.

Рассматриваются вопросы разработки и использования миниатюрных импульсных электростимуляторов глазных мышц для офтальмологии. Разработан прибор, который представляет собой генератор импульсов с несущей средней частотой 5000 Гц, фронты которых изменяются по экспоненциальному закону. Период повторения импульсов регу-

лируется в пределах от 5 до 100 Гц. Прибор изготовлен по технологии гибридных микроборок, что позволило значительно уменьшить его габаритные размеры и удобство для использования в офтальмологии.

The questions of development and use of tiny pulse electro stimulators of eye muscles for ophthalmology are considered. There is a device developed, which represents the generator of pulses with bearing average frequency 5000 Hz, the fronts of which change by exponential rule. The period of recurrence of pulses is adjusted in limits from 5 up to 100 Hz. The device is made using technology of hybrid micro assembly, that has allowed to reduce its overall dimensions and convenience to use in ophthalmology.

ВВЕДЕНИЕ

Косоглазие не только косметический дефект, но и тяжелый функциональный недостаток, выражающийся в отсутствии бинокулярного зрения и снижении центрального зрения косящего глаза. Неполноценность зрительных восприятий влияет также на физическое и умственное развитие ребенка, что ограничивает выбор профессии. Предупреждение и устранения косоглазия с одновременным восстановлением утраченных функций - важная задача научной и практической офтальмологии [1].

Для лечения косоглазия без оперативных вмешательств применяются специальные оптические устройства и электронные стимулирующие аппараты. Использование электрических импульсов для стимуляции работы глазодвигательных мышц позволяет восстановить ослабленные сократительные функции мышц и получить положительный терапевтический эффект.

Под влиянием раздражения импульсным током волна возбуждения распространяется по мышечным волокнам, происходит пассивное сокращение мышц, наблюдается постепенное восстановление ослабленной сократительной функции глазодвигательной системы. Одновременно появляется присущее однонаправленному току сосудорасширяющее действие и улучшение кровоснабжения мышечной ткани, активизация обменных окислительно-восстановительных процессов. Происходит восстановление трофики и чувствительности нервно-мышечных структур [2].

В офтальмологии для электростимуляции применяют импульсные токи различной формы: прямоугольной, синусоидальной, экспоненциальной и др. В зависимости от характера и степени поражения нервно-мышечного аппарата глаза определяется выбор того или иного вида импульсного тока. При выборе формы электрических импульсов учитывают адаптацию мышцы или нерва к импульсному раздражению. Чем ниже способность мышцы к аккомодации, тем более физиологичными для нее будут импульсы, которые медленно нарастают и сравнительно быстро спадают. В качестве таких импульсов применяются импульсы с экспоненциальными фронтами [3].

Порог раздражения такими импульсами значительно ниже, чем для прямоугольных импульсов, а их действие менее болезненно и не вызывает истощение мышцы. При электро-стимуляции импульсные посылки чередуются с паузами, так как длительное сокращение ведет к нарушению питания мышцы, ее истощению и усилению атрофии. Частота модуляций и длительность импульсов, как и вид тока, определяются функциональным состоянием нервно-мышечного аппарата. Чем глубже нарушение иннервации, тем меньше выбираемая частота модуляций и более длительные импульсы.

В настоящее время для электростимуляции применяются приборы типа "Ампли-пульс", "Тонус" и др. При достаточно широком наборе формируемых ими сигналов

модулированные импульсы с высокочастотной несущей и плавно изменяющимися фронтами отсутствуют. Кроме того, они имеют достаточно большие габариты, что не всегда удобно в офтальмологии.

ПОСТАНОВКА ЗАДАЧИ

Целью настоящей работы является разработка малогабаритного устройства для электростимуляции глазных мышц специальными импульсными токами.

МЕТОДЫ РЕШЕНИЯ

Для решения поставленной задачи был разработан прибор, структурная схема которого приведена на рис.1. Он представляет собой генератор импульсов с несущей средней частотой 5000 Гц, фронты которых изменяются по экспоненциальному закону. Период повторения импульсов регулируется в пределах от 5 до 100 Гц. Прибор изготовлен по технологии гибридных микросборок, что позволило значительно уменьшить габаритные размеры по сравнению с аналогами ("Амплипульс", "Тонус").

Генератор вырабатывает синусоидальное напряжение с частотой 5 кГц, стабильной амплитуды, которое поступает на один из входов смесителя. На второй вход смесителя подается напряжение экспоненциальной формы с генератора, частота которого изменяется в пределах от 5 до 100 Гц.

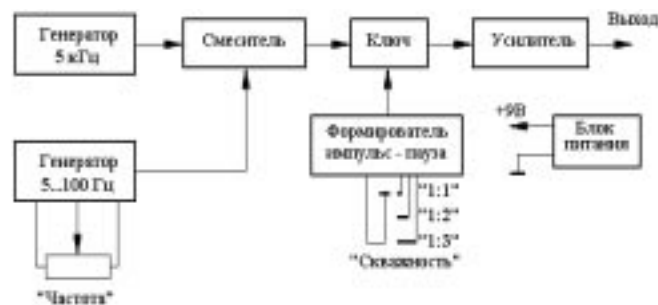


Рисунок 1 - Структурная схема электростимулятора

На выходе смесителя формируется высокочастотный сигнал 5 кГц с низкочастотной огибающей, которая повторяет экспоненциальную форму сигнала с генератора 5...100 Гц. Сформированный сигнал поступает на ключ. На управляющий вход ключа подключен формирователь скважности "импульс-пауза". На выходе ключа формируются пакеты импульсов со скважностью "1:1", "1:2", "1:3", которые поступают на вход усилителя и усиливаются до необходимого уровня (рис. 2).

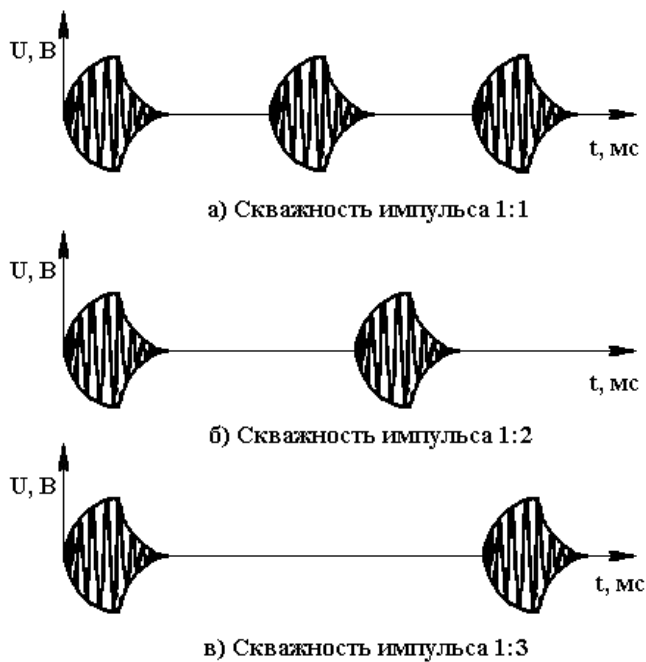


Рисунок 2 - Выходное напряжение, вырабатываемое электростимулятором

В отличие от низкочастотных токов, вызывающих сильное раздражение (ощущение жжения под электродом) и поэтому плохо переносимое больными, особенно с вегетативно-сосудистыми нарушениями, ток высокой частоты не встречает значительного сопротивления со стороны эпидермиса кожи, свободно проникает в глубину тканей, не вызывает заметного раздражения кожных рецепторов. Он хорошо переносится больными, оказывает возбуждающее действие на глубоко расположенные ткани.

Экспоненциальная форма импульса занимает промежуточное положение по эффективности между прямоугольной и синусоидальной формой. Токи, изменяющиеся по экспоненциальному закону, действуют более плавно, чем прямоугольные, но в то же время обладают большим стимулирующим действием, чем синусоидальные. В электростимуляторе "Амплипульс" синусоидальные низкочастотные импульсы модулируются синусоидальными токами более высокой частоты. Такая модуляция снижает раздражающий побочный эффект. Существуют приборы, генерирующие экспоненциальные токи без высокочастотной модуляции. Такие токи имеют хорошее миостимулирующее действие, но плохо переносятся детьми.

ВЫВОДЫ

Разработанный прибор позволяет применить импульсы с нарастающими и спадающими фронтами экспоненциальной формы с регулируемой частотой заполнения для миостимуляции. Применение технологии гибридных микросборок позволило существенно снизить габариты и вес прибора, обеспечив удобство при использовании в офтальмологии. Для расширения функциональных возможностей прибора необходимо ввести регулирование параметров модулирующего сигнала и выполнить дальнейшее уменьшение габаритов для серийного производства.

ПЕРЕЧЕНЬ ССЫЛОК

1. Сосин И.Н., Буявых А.Г. Физическая терапия глазных болезней. - Симферополь: Таврия, - 1998. - 248 с.
2. Рухлова С.А. Основы офтальмологии. - М.: Медицинская книга, Н. Новгород: Изд-во НГМА, - 2001. - 252 с.
3. Ремезов А.Н. Медицинская и биологическая физика. - М.: Высшая школа, 1999. - 616 с.

И.ИНФОРМАТИКА

УДК 579.67

ON THE EDUCATIONAL COURSE OF LOCALYZING COMPUTING

Gregory G. Men'shikov

Описываются основные математические идеи, лежащие в основе вычислений, осуществляющих интервальную локализацию искомого объекта.

The fundamental mathematical ideas are described which realize the localizing computing of the desired objects.

1 INTRODUCTION

A problem of the exactness is a very important one by itself since a success in its solution defines a power of the computer as the scientific tool. Further, only the well-verified (in this sense) computer allows to judge on the adequacy of the selected mathematical model to the technological problem to be solved.

In the tradition of scientific computation, it is not thought about consequences of the truncation errors. And remainder terms are simply neglected.

Meanwhile, it is well known, that numerous forms of errors are inherent to the traditional approach of the computation. They are quite uncontrolled. After the numerical experiment, the long and complicated investigation must be done to verify its results. Presently, this problem does not have a satisfactory solution.

It is enough to refer to the fine book of Babushka, Vitacek, and Prager [1], where non-complicated examples show an unfitness of the computer execution of algorithms.

It is seldom the case that the initial data are known with absolute exactness and the computation is quite integer-valued, the errors are absent. In such case, the computer is able to give the exact answer to the question about the object x "to what is value x equal?".

In the practice of science and engineering, more often the computer is not able to give an answer to this question since the exact value of the quantity to be found has no expression in machine numbers. For example, it is impossible to express an infinitely long binary fraction with machine numbers. Therefore the irrational numbers and the most part of the rational ones have no such expressions too.

In this typical situation the traditional computational methods give an answer to another question: "to what is a value x approximately equal?". But then they contribute the answer to one more question: "How much precise is a result which has been found by the computer?".

But it is very hard to get the answer. Thus, a quality of the traditional computing is doubtful.

A computer is not capable to address a thing about which it has no information.

I would like to talk about a new promising technology in computing experiments, namely localizing or enclosing computations. It finds the answer to the question, which is for-

mulated in the form "In what enclosure does x contain?". This enclosure may be named as localizer of x . The beginning of the localizing computation is connected with the name of the american mathematician Ramon E. Moore [2]. His main predecessors were:

- Vladimir M. Bradis [3] (USSR),
- Rosalind C. Young [4] (Great Britain),
- Mieczyslaw Warmus [5] (Poland),
- Teruo Sunaga [6] (Japan),
- Leonid V. Kantorovitch [7] (USSR).

The first russian book after Bradis is written by Yuri I. Shokin [8].

The program of reconstruction the educational course of the Computational Methods at the Faculty of the applied mathematics-control processes of the Saint Petersburg State University on localizing base was pioneered by recently died (2000) professor Vladimir I. Zubov. He was the chairman of the Methodical Commission of that faculty, the corresponding member of the Russian Academy of Sciences. I was set to do this in 1993. Now the reconstruction is close to the finishing [9].

In 1998 Vyacheslav B. Orlov jointed to this work. As part of reconstruction program we published 12 issues of lecture notes [9.1-9.12]. We prepared the subjects of the home checking works specific for the each student and developed their program support. We gained the greatest amount of experience.

Here we concern the theoretical foundation of localizing computing. Some fundamental mathematical ideas compose this foundation. And I would like to present their review.

2 IDEA OF AN INTERVAL ENCLOSING

The first one is the idea of localizing or enclosing itself. It represents the reality that some fuzziness, uncertainty, non-stability are inherent to numerical characteristics of real phenomena, processes, events, constructions.

Let us consider phenomena of the technical world.

Any composite technical device consist of its components. Therefore a mass production of technical devices is impossible without the corresponding mass production of the components. Each component must satisfy different standards of precision. Otherwise it will be impossible to gather a device from them. Thus, a modern manufacture is impossible without standartization.

The end of 1920-ties - beginning of 1930-ties, saw the organized mass production of the components of radio engineering devices and among them resistors. On the body of a resistor it is pointed out the nominal value of the resistance

(presented in Ohms Ω , kiloOhms ($k\Omega$) and so on) and the tolerance in percents. For instance the information ($10 k\Omega$, $\pm 10\%$) means that the true value of the resistance lies between 9 and 11 $k\Omega$.

This meant a rejection of exact number representation of the electrical characteristics and a transition to the enclosing representation, or the interval representation.

It may be a measure of the resistance by a voltmeter. An uncertainty will decrease, but will not be vanish due to an usual inaccuracy of a measuring. Moreover, a natural non-stability will remain (partly, due to temperature changes). And it will not bring success by reducing the enclosure for the resistance beyond a certain point.

Equally the capacity of a condenser and a characteristic of other electrical dipoles are represented now by their nominal values and a tolerances.

In mathematical computations presence of the errors may be explained with unboundness of a set of the mathematical objects and with restriction of a machine resources for their representataion. Here idea of localizing has sense, too. It was advanced by Bradis in 1927 in the form of method of bounds [3].

This idea in some form is as following: a closed interval serves as the enclosure for all sets which it contains.

3 USEFULLNESS OF THE ENCLOSING FOR THE MATHEMATICAL COMPUTATIONS

Let us show an usefulness of the interval enclosing by means of another clear example.

Example 1. Let it be necessary to compute points of the curve (some trajectory) for several given values of an abscissa. The traditional computation gives some points without any guarantee of nearness to the desired curve (see Fig. 1). It remains unknown, where this curve lies. But an enclosing computation gives the vertical intervals - enclosures for the ordinates of desired points (see Fig. 2). They play the role of the notches, through which the desired curve passes necessarily.

Hence a more narrow localizer is a more accurate one.



Figure 1



Figure 2

4 ACKNOWLEDGEMENT

I am very grateful to Dr. S. Neuschl (Slovakia) for the attention to this manuscript. I am very grateful also to Dr. L. Shakhmundes (Canada) and to my daughter Mrs. O. G. Men'shikova (Russia, S. Petersburg) for their efforts and helpful comments that facilitated in the improvement of this manuscript.

REFERENCES

1. Babuska, M. Prager, and E. Vitasek. Numerical Processes in Differential Equations. - SNTL, Praha, Czechoslovakia. 1966.
2. R.E. Moore. Interval Analysis. - Englewood Cliffs, N.J.: Prentice-Hall, 1966.
3. V.M. Bradis. The Experience of the Verification of Some Practical Rules of the Operations under Approximate Numbers. - Tver', 1927 (in Russian).
4. R.C. Young. The Algebra of many-valued quantities. // Math. Ann., 1931, v. 104, pp. 260 - 290.
5. M. Warmus. Rachunek przyblizony w zakresie pomiarow technicznych i wynikajacych z nich bledow. // Przegląd Elektrotechniczny. 1955, 31, N 10/11, 748 - 752.
6. T. Sunaga. Theory of an interval algebra and its application to numerical analysis. //RAAG Memoirs, 1958, v. 2, 547 - 564.
7. L.V. Kantorovich. On some new approaches to the computational methods and observations processing. //Siberian Mathematical Journal, Volume III, N 5, 1962, 701 - 709 (In Russian).
8. Yu.I. Shokin. Interval analysis. - Novosibirsk: Nauka. Siberian Department, 1981 (In Russian).
9. G.G. Men'shikov. Interval Analysis and Methods of Computing. Lecture Notes. Operative Poligraphy Department of Research Chemical Institute. Saint Petersburg State University. 1998 - 2001 (in Russian):
- Vol.1. Introduction to the Interval Organization of Computation. 2nd edition. 2000,
- Vol.2. Bilateral Solution of Elementary Problems. 2nd edition. 2000,
- Vol.3. Intervalization of Approximate Formulae. 2nd edition. 2000,
- Vol.4. An Introduction to Approximation of Functions. 2nd edition. 2000,
- Vol.5. Research the Function of One Variable. Differentiation of the Functions.1998,
- Vol.6. Localizing Integration of the Functions. 1998,
- Vol.7. The Approximation of the Functions and the Verification of the Approximates. 1998,
- Vol.8. The Iterative Processes and the Systems of the Number Equations. 1999,
- Vol.9. Elements of the Localizing Integration of the Differential Equations. 2nd edition. 2001,
- Vol.10. Localizing Integration of the Differential Equations. 2001,
- Vol.11. Analytical and Algebraical Support of Localizing Integration of the System of Differential Equations. 2001,
- Vol.12. Interval-Analytical Supplements. 2001.

DESIGN OF AN ADAPTIVE FUZZY LOGIC CONTROLLER FOR NONLINEAR DYNAMIC SYSTEM

Ho Dac Loc

В этой статье разработан адаптивный нечеткий контроллер для обобщенной высокопорядковой нелинейной непрерывной системы. Синтезированный контроллер обеспечивает: 1) глобальную стабильность циклически замкнутой системы; 2) асимптотическую сходимость к нулю ошибки слежения. Нечеткий контроллер синтезируется на основе коллекции правил "Если-то". Параметры членских функций характеризуют лингвистические выражения, измеряющиеся в соответствии с несколькими адаптивными законами с целью управления слежением за установкой по ориентировочной траектории.

In this paper, an adaptive fuzzy controller is designed for a general higher-order nonlinear continuous system. The synthesized controller ensures that 1) the close-loop system is globally stable and 2) the tracking error converges to zero asymptotically. The fuzzy controller is synthesized from a collection of IF-THEN rules. The parameters of the membership functions characterizing the linguistic terms change according to some adaptive law for the purpose of controlling a plant to track a reference trajectory.

1 INTRODUCTION

In recent years the adaptive control of nonlinear systems has attracted a lot of attention. Control methodology called feedback linearization has been proved sound and successful in some problems. The central concept of this approach is to transform the nonlinear system dynamic into an equivalent linear systems, so that conventional linear control techniques can be applied [1, 2]. A key assumption in these studies is that the system nonlinearities are known a priori, or can be linearizable parameterization. This presents a limitation of the theory because the real system, or more precisely the model of a system, may always contain with a nonlinear uncertain elements. Therefore, the design of a robust adaptive controller that deals with a nonlinear system with uncertainties is an important subject. So far, to deal with uncertain nonlinear systems, many adaptive control approaches have been proposed. Adaptive control approaches are applied to the systems with parameterized uncertainties, several results can be found in [3-6]. The above discussion makes apparent that adaptive control research has thus far been directed towards the systems with parametric uncertainties, for the totally unknown nonlinear systems cannot be discussed. The need to deal with uncreasingly complex systems, to accomplish in creasingly demanding design requirements and the need to obtain these requirements with less precise advanced knowledge of the plant and its environment, inspired many works that came mostly from the area of fuzzy logic control.

Recently, the integration of fuzzy logic techniques and conventional control approaches has been an active research

focus. It is claimed that this union will lead to new control algorithms that exploit the advantages of both paradigms. The property of uniformly approximating any nonlinear systems over compact input-space with fuzzy systems [7] lays the foundation for this integration and provides a bridge to convert a set of human heuristic rules into a mathematical description. Wang [8] suggested the utilization of conventional adaptive systems to the fuzzy system framework, and presented an in-depth and thorough analysis of adaptive fuzzy logic control based on the general error dynamics of adaptive system. The most important challenge to this approach is to solve the Lyapunov equation and find the positive definite matrix. Passino et. al. [9] developed a fuzzy model reference learning control by introducing a reference model for defining the desired process characteristics where a fuzzy inverse model has to be specified by the control engineer in advance. Also, many other excellent works on introducing the concept of fuzzy logic into the conventional control techniques and vice versa have been developed in the past years [10, 11-14], to name but a few.

However, most practical applications of fuzzy logic control have been limited to relatively simple problems, due to a lack of formal synthesis techniques which guaranties the very basic requirements of global stability and acceptable performances [15, 16]. The design of the globally stable fuzzy control systems was an open problem until recent efforts presented in [16]. Based on fuzzy logic systems which are capable of approximating, with arbitrary, any real continuous function on a compact set, a globally stable adaptive controller is first synthesized from a collect of IF-THEN rules. The fuzzy systems used to approximate an optimal controller, is adjusted by an adaptive law based on a Lyapunov function synthesis approach. However, this adaptive fuzzy control system is limited to the linearizable SISO nonlinear systems with control gain being constant. On the other hand, it cannot eliminate the effect of the modeling errors and of disturbances on the error output of the system. By introducing sliding mode control techniques, [15] designed a fuzzy controller for the same system as [16]. But the uses of the sliding mode, which is a discontinuous control, generally creates various problems, such as the chattering phenomena and possible excitation of high-frequency unmodeled dynamics.

In this paper, we will develop a robust adaptive control scheme for a class of unknown nonlinear systems. The basic architecture of our adaptive fuzzy controller is a standard fuzzy logic controller (FLC) [17] used in most fuzzy control systems, plus an adaptive law used for adjusting the parameters of FLC.

2 DESCRIPTION OF FUZZY LOGIC SYSTEMS

The basic configuration of the fuzzy logic system is shown in Fig.1. The fuzzy logic system performs a mapping from $U \subseteq R^n$ to R . Let $U = U_1 \times \dots \times U_n$ where $U_i \subset R$, $i = 1, 2, \dots, n$. The fuzzy rule base consists of a collection of fuzzy IF-THEN rules

$$R^l: \text{ IF } x_1 \text{ is } F_1^l \text{ AND } x_2 \text{ is } F_2^l \text{ AND } \dots \text{ AND } x_n \text{ is } F_n^l \\ \text{ THEN } y^l \text{ is } C^l \quad (l = 1, \dots, M), \quad (1)$$

where $\underline{x} = (x_1, \dots, x_n)^T \in U$, $y \in V \subset R$ are the input and output of fuzzy logic systems, respectively, F_i^l , C^l are the fuzzy sets defined on U_i and R , respectively. The fuzzy inference engine performs a mapping from fuzzy sets in U to fuzzy sets in R , based upon the fuzzy IF-THEN rules in the fuzzy rule base and the compositional rule of inference. The fuzzifier maps a crisp point $x = (x_1, \dots, x_n)^T$ into a fuzzy set in U . The defuzzifier maps a fuzzy set in V to a crisp point in V .

The fuzzy logic systems of Fig.1 comprise a very rich class of static systems mapping from $U \subset R^n$ to $V \subset R$, because many different choices are available within each block, and in addition, many combinations of these choices can result in a useful subclass of fuzzy logic systems. One subclass of fuzzy logic systems is used here as building blocks of our adaptive fuzzy controller, and is described by the following important result.

Lemma 1. The fuzzy logic systems with center-average defuzzifier, product inference and singleton fuzzifier are in the following form

$$u(\underline{x}) = \frac{\int_{l=1}^M \lambda^l \left(\prod_{i=1}^n \mu_{F_i^l}(x_i) \right)}{\int_{l=1}^M \left(\prod_{i=1}^n \mu_{F_i^l}(x_i) \right)}, \quad (2)$$

where λ^l is the point at which μ_{C^l} achieves its maximum value, and we assume that $\mu_{C^l}(\lambda^l) = 1$.

Eq.(2) can be written as

$$u(\underline{x}) = \underline{\lambda}^T \underline{\zeta}(\underline{x}), \quad (3)$$

where $\underline{\lambda} = (\lambda^1, \dots, \lambda^M)$ is a parameter vector, and $\underline{\zeta}(\underline{x}) = (\zeta_1(\underline{x}), \dots, \zeta_M(\underline{x}))^T$ is regressive vector, $\zeta_i(\underline{x})$ is defined fuzzy basic function (FBF) []

$$\zeta_l(\underline{x}) = \frac{\prod_{i=1}^n \mu_{F_i^l}(x_i)}{\int_{l=1}^M \left(\prod_{i=1}^n \mu_{F_i^l}(x_i) \right)}. \quad (4)$$

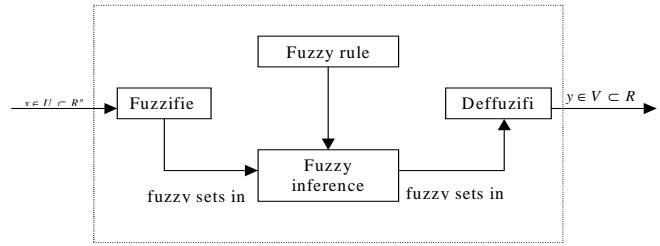


Figure 1 - Basic configuration of a fuzzy logic system

There are two main reasons for using the fuzzy logic system (3) as basic building blocks of adaptive fuzzy controller. First, it was proved that the fuzzy logic system in the form of (3) are universal approximates, i.e., for any given real continuous function g on the compact set U , there exists a fuzzy logic system in form of (9) such that it can uniformly approximate g over U to arbitrary accuracy. Therefore, the fuzzy logic system (3) are qualified as building blocks of adaptive fuzzy controllers for nonlinear systems. Second, the fuzzy logic system (3) are constructed from the fuzzy IF-THEN rules of (1), therefore, linguistic information from human experts can be directly incorporated into the controllers.

3 BASIC IDEAS OF CONSTRUCTING STABLE ADAPTIVE FUZZY CONTROLLERS

In this section, we first setup the control objectives, and then show how to develop an adaptive fuzzy logic controller to achieve these control objectives. Consider the n -th-order nonlinear systems of the form

$$\begin{aligned} \dot{x}_1 &= f_1(\underline{x}) \\ \dot{x}_2 &= f_2(\underline{x}) \\ &\dots\dots\dots \\ \dot{x}_n &= f_n(\underline{x}) + u \\ y &= h(\underline{x}), \end{aligned} \quad (5)$$

where $\underline{x} = (x_1, \dots, x_n)^T$ is the state vector, $f_i, i = \overline{1, n}$; $h(\cdot)$ are unknown functions, $u \in R$ and $y \in R$ are the input and output of system, respectively.

Control objectives: Determine a feedback control u based on fuzzy logic system (3) and an adaptive law for adjusting the parameter vector $\underline{\lambda}$ such that the following conditions are met:

a) $\lim_{t \rightarrow \infty} |y_m - y| = \lim_{t \rightarrow \infty} |y_m - h(\underline{x})| = 0$ where y_m is the output tracking signal;

b) the cost function $J = \int_0^\infty [\Psi(\underline{x})^2 + \dot{\Psi}(\underline{x})^2] dt$ is minimized, $\Psi(\underline{x})$ is a differentiable function of state variables and $\Psi(0) = 0$.

We now show the basic ideas of how to construct a direct adaptive fuzzy controller to achieve these control objectives.

To begin, let $\underline{e} = (e, \dot{e})^T = (e_1, e_2)^T$ and $\Psi(\underline{x}) = e_1 + \alpha e_2$. To ensure the control objective b), we have

$$\Psi(\underline{e}) + \dot{\Psi}(\underline{e}) = 0. \quad (6)$$

Using (5) we have:

$$\dot{\Psi} = \int_{i=1}^n \frac{\partial \Psi}{\partial x_i} \dot{x}_i = \int_{i=1}^{n-1} \frac{\partial \Psi}{\partial x_i} f_i(\underline{x}) + \frac{\partial \Psi}{\partial x_n} [f_n(\underline{x}) + \hat{u}], \quad (7)$$

where \hat{u} is the optimal control. Solving Eq.(6), using (7), we have

$$\hat{u} = -\left[\frac{\partial \Psi(\underline{e})}{\partial x_n}\right]^{-1} \left[\Psi(\underline{e}) + \int_{i=1}^{n-1} \frac{\partial \Psi(\underline{e})}{\partial x_i} f_i(\underline{x}) \right]. \quad (8)$$

If the functions $f_i(\cdot)$, $i = \overline{1, n}$ and $h(\cdot)$ are known, the control (8) ensures the control objective a). We can show it using the follow equation:

$$\begin{aligned} \Psi + \dot{\Psi} &= \Psi + \int_{i=1}^{n-1} \frac{\partial \Psi}{\partial x_i} f_i + \frac{\partial \Psi}{\partial x_n} (f_n + \hat{u}) = \\ &= \Psi + \int_{i=1}^{n-1} \frac{\partial \Psi}{\partial x_i} f_i + \frac{\partial \Psi}{\partial x_n} \left\{ f_n - \left[\frac{\partial \Psi}{\partial x_n}\right]^{-1} \left[\Psi + \int_{i=1}^{n-1} \frac{\partial \Psi}{\partial x_i} f_i \right] \right\} = 0. \quad (9) \end{aligned}$$

From (9) we can see that: $\lim_{t \rightarrow \infty} |\Psi| = 0$ or $\lim_{t \rightarrow \infty} |y_m - h(\underline{x})| = 0$ - the main control objective.

Since $f_i(\underline{x})$, $i = \overline{1, n}$ and $h(\underline{x})$ are unknown, the optimal control \hat{u} of (8) cannot be implemented. Our purpose is to design a fuzzy logic system to approximate this optimal control.

Eq. (5) can be written as:

$$\begin{aligned} \dot{x}_1 &= f_1(\underline{x}) \\ \dot{x}_2 &= f_2(\underline{x}) \\ &\dots\dots\dots \\ \dot{x}_n &= f_n(\underline{x}) + u + \hat{u} - \hat{u}. \quad (10) \end{aligned}$$

Substituting (8) and approximated control signal by fuzzy logic system $\hat{u} = \underline{\lambda}^* T \underline{\zeta}(\underline{e})$ into (10), we have:

$$\begin{aligned} \dot{x}_1 &= f_1(\underline{x}) \\ \dot{x}_2 &= f_2(\underline{x}) \\ &\dots\dots\dots \\ \dot{x}_n &= -\left[\frac{\partial \Psi}{\partial x_n}\right]^{-1} \left[\Psi + \int_{i=1}^{n-1} \frac{\partial \Psi}{\partial x_i} f_i \right] + (\underline{\lambda}^* T - \underline{\lambda}^* T) \underline{\zeta}(\underline{e}). \quad (11) \end{aligned}$$

Next, we develop an adaptive law to adjust the parameters of vector $\underline{\lambda}$, which provides that the close-loop system is generally stable. Define the Lyapunov function in the form

$$V = \frac{1}{2} \Psi(\underline{e})^2 + \frac{1}{2\gamma} \underline{\theta}^T \underline{\theta}, \quad (12)$$

where $\underline{\theta} = \underline{\lambda} - \underline{\lambda}^*$, γ is positive constant. Using (11) we have a derivative of Lyapunov function as:

$$\begin{aligned} \dot{V} &= \Psi \dot{\Psi} + \frac{1}{\gamma} \underline{\theta}^T \dot{\underline{\theta}} = -\Psi^2 + \frac{\partial \Psi}{\partial x_n} \Psi \underline{\theta}^T \underline{\zeta}(\underline{e}) + \frac{1}{\gamma} \underline{\theta}^T \dot{\underline{\lambda}} = \\ &= -\Psi^2 + \underline{\theta}^T \left[\frac{\partial \Psi}{\partial x_n} \Psi \underline{\zeta}(\underline{e}) + \frac{1}{\gamma} \dot{\underline{\lambda}} \right]. \quad (13) \end{aligned}$$

If we choose the adaptive law:

$$\dot{\underline{\lambda}} = -\gamma \frac{\partial \Psi}{\partial x_n} \Psi \underline{\zeta}(\underline{e}), \quad (14)$$

then (13) becomes $\dot{V} = -\Psi^2 \leq 0$, which guaranties that the close-loop adaptive system is generally stable.

The overall scheme of the developed adaptive fuzzy control system is shown in Fig.2. Now, we make some few remarks.

Remark 1. The developed, in this paper, adaptive fuzzy control system using the error vector as input signal for the controller is suitable in situations, where the state vector is not measured variables.

Remark 2. The simple adaptive law makes easy to implement the adaptive fuzzy control. The quality of the control system depends on form and parameters of function Ψ .

Remark 3. In using the adaptive law (14), the choice of the constant γ is important. At present, there has been no theoretical guidance about the choice of γ ; trial-and-error seems to be only practical option.

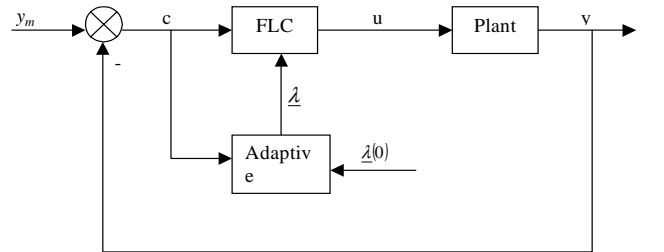


Figure 2 - The overall scheme of adaptive fuzzy control system

4 SIMULATION

In this section, we apply the direct adaptive fuzzy controller in the last section to control an unstable nonlinear dynamic systems.

Example1. In this example, we use our direct adaptive fuzzy controller to regulate the plant:

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -0,1x_2 - x_1^3 + 12 \cos(t) + u + d(t). \end{aligned} \quad (15)$$

If without control, the system is chaotic. Here the objective is to use the developed adaptive fuzzy logic controller to let system output track the desired sin input and square input, respectively. Moreover, a random disturbance with 0 mean and 1 variance is added to verify the performance of the control system. Fig.3 and Fig.4 are the simulation results, which indicate that the tracking performance is very good even in the presence of random disturbance.

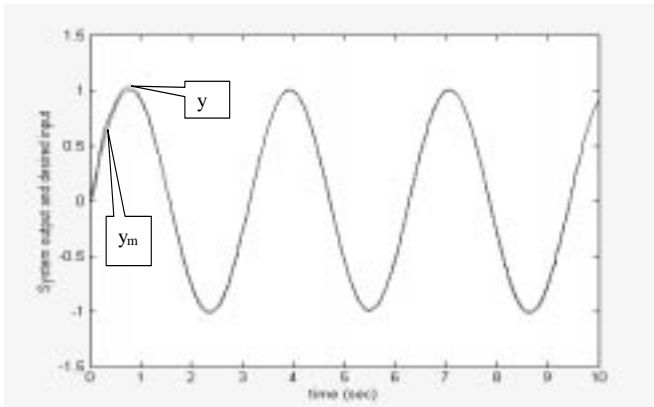


Figure 3 - Control results with AFLC ($\alpha = 1$; $\gamma = 500$)

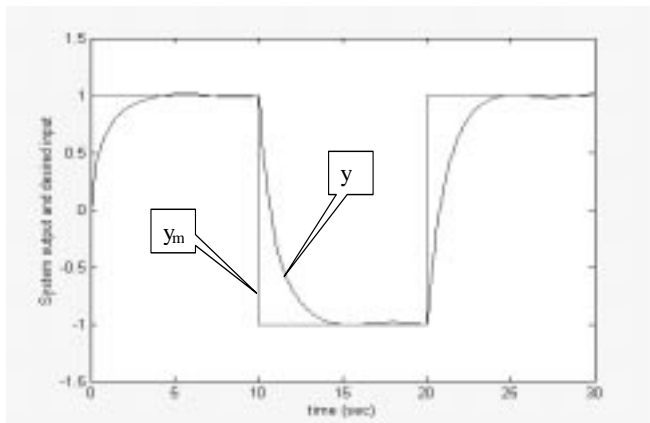


Figure 4 - Control results with AFLC ($\alpha = 1$; $\gamma = 500$)

Example 2. In this example, we apply the adaptive fuzzy controller to control the following nonlinear dynamic system:

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = -x_2 + \text{sat}(x_3), \\ \dot{x}_3 = -x_3 + u, \\ y = x_1. \end{cases} \quad (16)$$

The step respond of close-loop control system is shown in Fig.5, in which we can see the adaptive fuzzy controller could regulate the plant and the close-loop system is stable.

the adaptive fuzzy controller could regulate the plant and the close-loop system is stable.

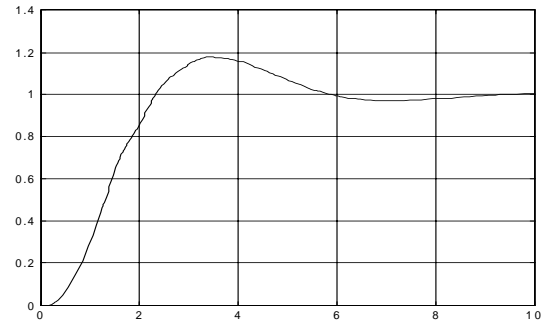


Figure 5 - The step respond of the control system

The step respond of the control system in situation, where γ has different values, is shown in Fig.6. From this we can see that the quality of the transient process depends on γ .

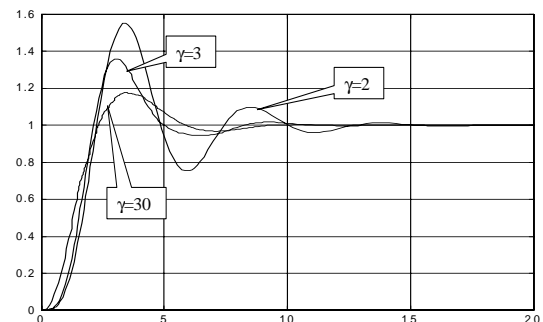


Figure 6 - The step respond

5 CONCLUSION

In this paper, we developed an adaptive fuzzy logic controller which: 1) does not require an accurate mathematical model of plant under control, 2) uses the error vector as controller's input, therefore it does not require all that components of state vector to be measurable, and 3) guarantees the global stability of close-loop system. The simulation results show that the adaptive fuzzy controller could successful control the unknown nonlinear dynamic system.

REFERENCES

1. A. Isodori. Nonlinear control systems, Springer, New York, 1989.
2. S. Sastry, A. Isodori. Adaptive control of linearizable systems, IEEE Trans. Automat. Control 34 (1989) 1123-1131.
3. I. Kanellakopoulos, P.V. Kokotovic, R. Maritio. An extended direct scheme for robust adaptive nonlinear control, Automat-

ica 27 (1991) 247-255.

4. I. Kanellakopoulos, P.V. Kokotovic, A.S. Morse. Systematic design of adaptive controllers for feedback linearization systems, IEEE Trans. Automat. Control 36 (1991) 1241-1253.
5. R. Marino, P. Tomci. Global adaptive output feedback control of nonlinear systems, part 1: nonlinear parameterization, IEEE Trans. Automat. Control 38 (1993) 17-32.
6. M.M. Polycarpou, P.A. Loannou. A robust adaptive nonlinear control design, Automatica 32 (1996) 423-427.
7. L.X. Wang. Adaptive fuzzy systems and control, Prentice-Hall, Englewood Cliffs, NJ, 1994.
8. L. X. Wang. Fuzzy systems are universal approximators, Proceedings of IEEE Conference on Fuzzy systems, San Diego, 1992, pp. 1163-1170.
9. R. Palm. Sliding mode fuzzy control, Proceedings of IEEE Conference on fuzzy systems, San Diego, 1982, pp. 519-526.
10. D. Driancov, R. Palm, (Eds.). Advances in fuzzy control, Heidelberg, 1998.
11. G.C. Hwang, S.L. Lin. A stability approach to fuzzy control design for nonlinear systems, Fuzzy Sets and Systems 48 (1992) 279-287.
12. T.A. Johansen. Fuzzy model based control: stability, robustness, and performance issues, IEEE Trans. Fuzzy Systems 2 (3) (1994) 221-234.
13. W.A. Kwong, K.M. Passino. Dynamically focused fuzzy learning control, IEEE Trans. SMS 26 (1996) 53-74.
14. J.R. Layne, K.M. Passino. Fuzzy model reference learning control for cargo ship steering, IEEE Control Systems Mag. 13 (6) (1993) 23-34.
15. C.Y. Sue, Y. Stepanenko. Adaptive control for a class of nonlinear systems with fuzzy logic, Trans. Fuzzy Systems 29 (1994) 285-294.
16. Wang Li-Xin. Stable adaptive fuzzy control of nonlinear systems, IEEE Trans. Fuzzy Systems 1 (1993) 146-155.
17. C.C. Lee. Fuzzy logic in control systems: Fuzzy logic controllers, parts I and II, IEEE Trans. Syst., Man, Cybern., 20 (2) (1990) 404-435.

УДК 004.93

HYBRID LEARNING ALGORITHM FOR NEURAL NETWORKS

Jamil Ahmad

Рассмотрен гибридный алгоритм для обучения нейросети, который адаптирует веса дважды на одной итерации. Этот новый алгоритм представляет собой комбинацию алгоритма обучения машины, предложенный Габором в 1960-х годах и метода наименьших квадратов. Этот гибридный алгоритм использует два разных уравнения, основанных на минимизации среднеквадратической ошибки для оптимизации весов. Алгоритм показал лучшие способности, чем метод наименьших квадратов и метод обратного распространения ошибки относительно задачи распространения образов.

This paper presents a hybrid algorithm for the learning mechanism of the neural network model which adjusts weights twice in any single iteration. This new algorithm is a combination of a machine learning algorithm proposed by Gabor in 1960s and LMS algorithm. This hybrid algorithm uses two different equations based on mean square errors to optimize weights. The algorithm showed better performance when compared with Least Mean Square (LMS) and Back Propagation (BP) learning algorithms using pattern recognition problem.

1 INTRODUCTION

In recent years, ANN models have made great leaps in solving complex problems such as prediction, classification, speech analysis, image analysis, and pattern recognition [1-5]. A number of sophisticated learning models have been developed to solve variety of problems. In spite the remarkable achievement by ANN model in some application areas, there is still space for improvement. Mostly, these models are suffered from problems of slow convergence and its structure definition. This paper presents a hybrid approach by combining two algorithms into a single learning model. This hybrid algorithm is mainly derived from the Gabor theory of Communication and Machine Learning, [6-7] which is modified by merging it with LMS based learning algorithm. Further information about the algorithm can be found in [8]. The algorithm is compared with standard BP [9] and LMS [10], with the pattern recognition problem. Various parameters such as initialization of weights, learning rate,

and learning curve are also investigated with the help of experimental study.

2 THE LEARNING PARADIGM

The general structure of the proposed hybrid learning algorithm is shown in Figure no. 1, which shows working mechanism of the algorithm. It can be noted that the algorithm adjusts weight in two stages. Both stages are carried out in each training run simultaneously, i.e., the adjustment of the weights takes place twice in single iteration. In the first stage, the algorithm calculates three errors associated with each weight and uses them to modify the associated weight (only one weight). Subsequently, the algorithm uses mean square errors to adjust all the weights which is considered as a second stage of the model. The flowchart for the proposed system is also shown in Figure no. 2.

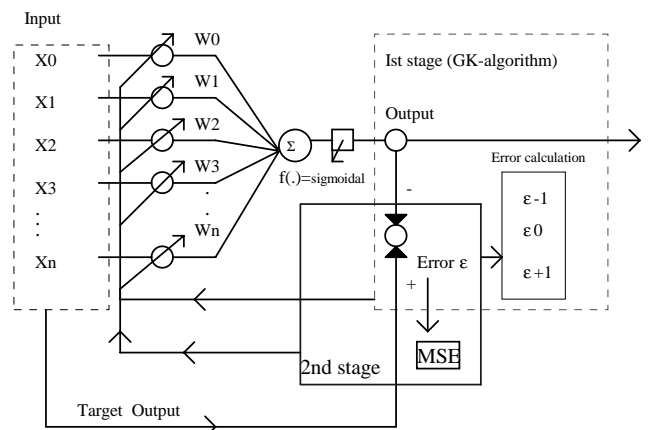


Figure 1 - General structure of the proposed Hybrid learning algorithm

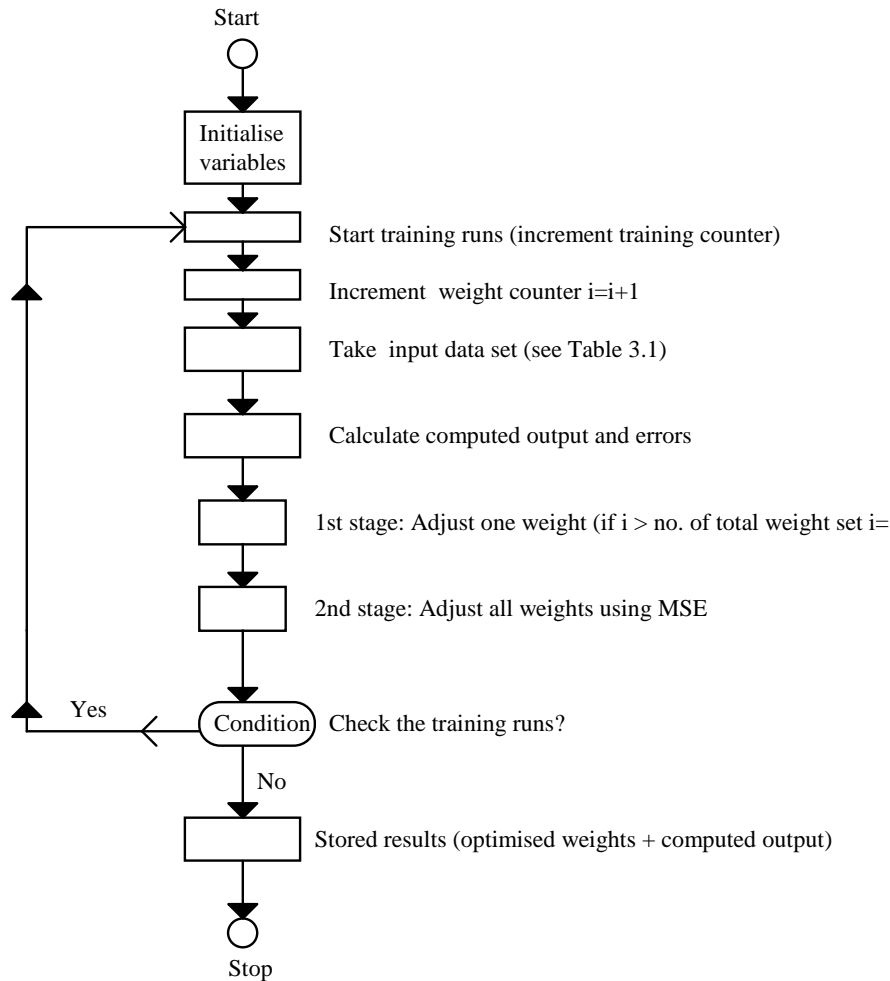


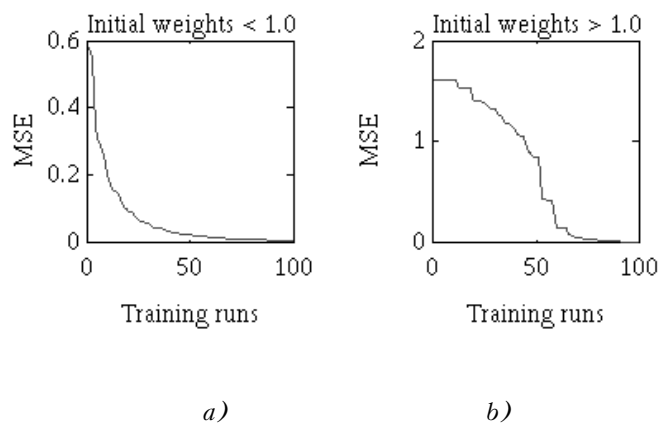
Figure 2 - Working flow of the proposed Hybrid learning algorithm

3 EXPERIMENTAL STUDY

The above two-stages model is analyzed through experimental study using a computer simulation package developed for this purpose. During the experiments a number of parameters such as, initial weights setting, learning constant, etc. are investigated thoroughly. The following paragraphs explain these parameters with more details and simulation results.

3.1 The Initial Weights values of the Algorithm

Most of the network models use random initial values for the weights to calculate their first approximated computed output. It is clear from the current literature that no specific rule exists to calculate the initial weight values; so in most cases these values are fixed after a number of experiments which are carried out with different random initial weight values. A number of experiments are carried out to find suitable initial weight boundaries for the proposed hybrid learning algorithm. Figures no. 3(a) and 3(b) show the learning curves for initial weights values, (< 1.0) and (>1.0), respectively. These graphs indicate that the initial weights' values less than 1.0 shows better performance.



Figures 3 - Initial weight values for the proposed Hybrid learning Algorithm. The weight values are randomly selected less than 1.0 and greater than 1.0.

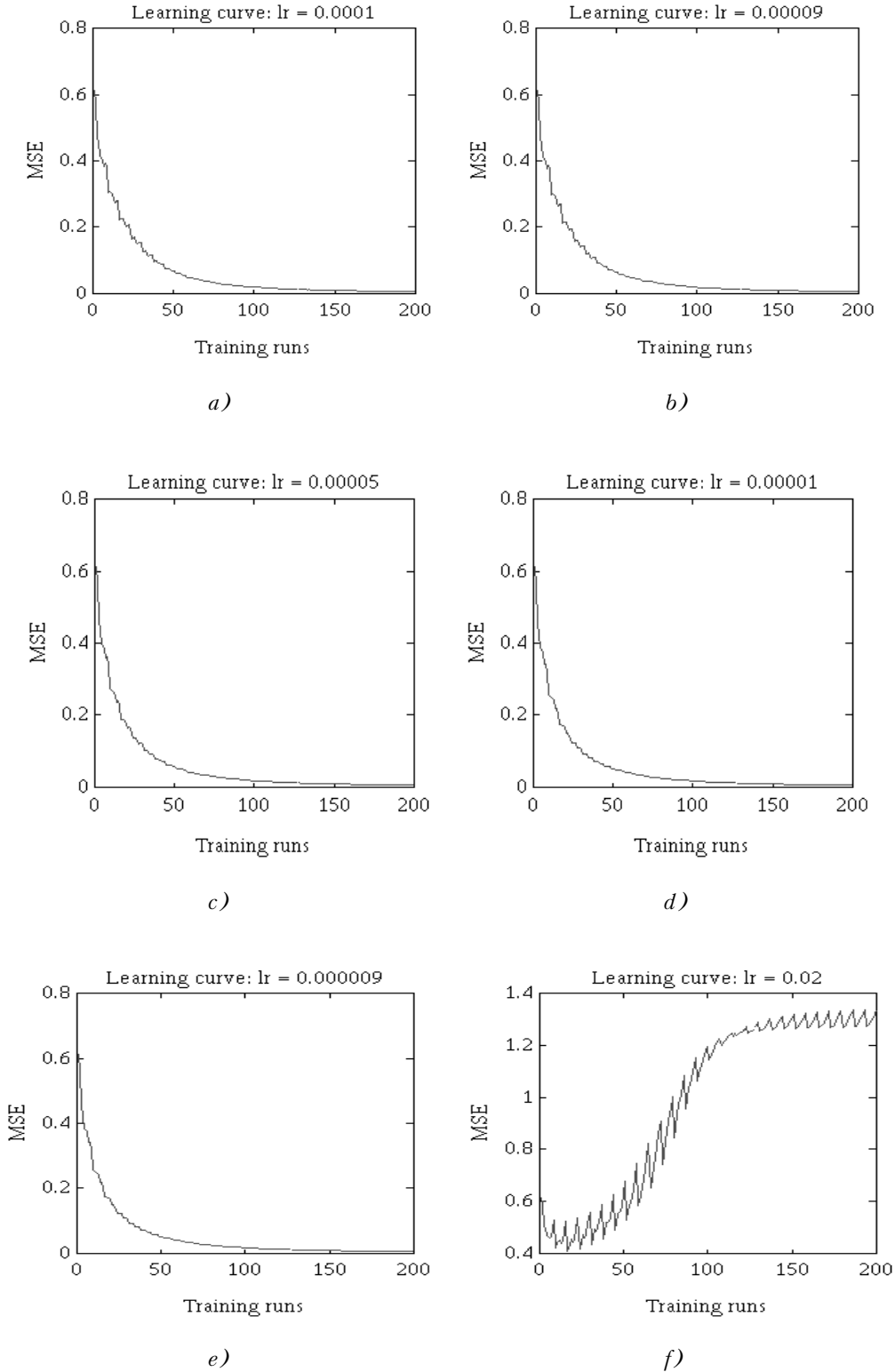
3.2 Learning rate of the Algorithm

The concept of learning rate is discussed intensively in the current literature, which controls stability and speed of

convergence. Generally, a small value is preferable, i.e., between 0.1 and 1.0 (in some cases even 2.0 is acceptable) [11]. High learning rates cause divergence from the required target output, as in the case of LMS learning algorithm, especially when the learning rate is greater than 2.

The learning rate of the hybrid learning algorithm should be very small, i.e. between 0.00005 - 0.0001. Several experi-

ments are carried out to set the learning rate for this algorithm and the results are graphically shown in Figures 4(a) to 4(f) for 0.0001, 0.00009, 0.00005, 0.00001, 0.000009, and 0.02, respectively. It is found that the proposed hybrid learning algorithm has a precise range of values for the learning rate. Further, learning rates of values 0.02 or more create problems during the convergence.



Figures 4 - Effect of learning constant with different values on the training performance of the proposed Hybrid learning algorithm

4 AN EXAMPLE OF PATTERN RECOGNITION

The proposed model is analyzed with the help of an example of pattern recognition. The input and target output are provided, as shown below in Figure no. 5. The learning curves for the proposed hybrid, LMS and BP learning algorithms are discussed in the following section.



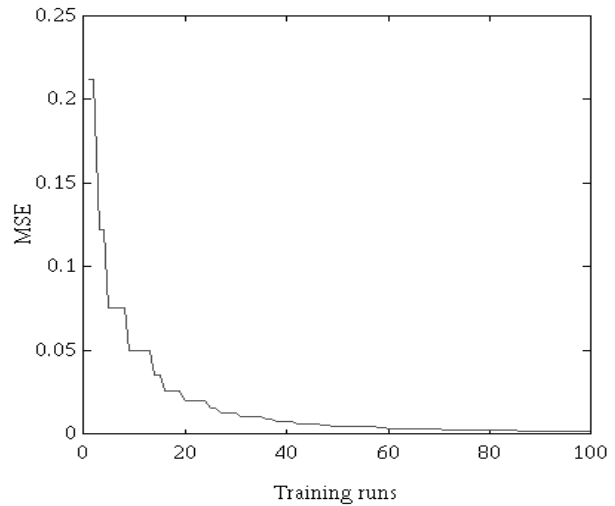
Figure 5 - Input and target patterns

4.1 The Learning Curves

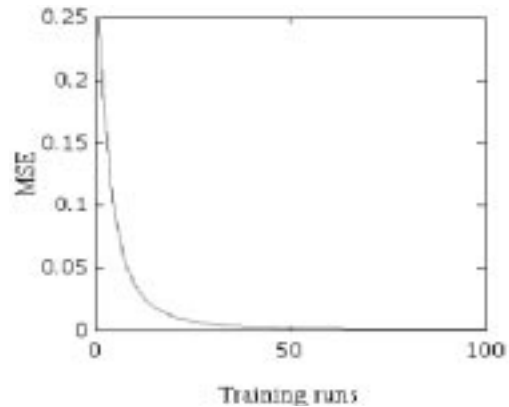
The criteria for the adjustment of the weight have been an object of the neural network literature for some time, but practically there is only one choice, the criteria of mean squares, as used by Widrow-Hoff, Gabor, Kolmogorov and Wiener for their models. The proposed model also uses MSE for its weights optimization. One can observe the speed, performance and adaptation of the neural network learning algorithms from the reduction of associated MSEs. The MSEs for the example pattern as shown above are presented for the hybrid, LMS and BP learning algorithms in Figure nos. 6(a), 6(b) and 6(c), respectively. These graphs indicate better performance for the hybrid learning algorithm than the other two algorithms. The comparison with the LMS and BP learning algorithms shows that the hybrid learning algorithm is more powerful and its optimization techniques are faster than the LMS and BP learning algorithm because it reaches quickly to the minimum level of the error, i.e., shown in Figure no. 6(a).

5 CONCLUSION

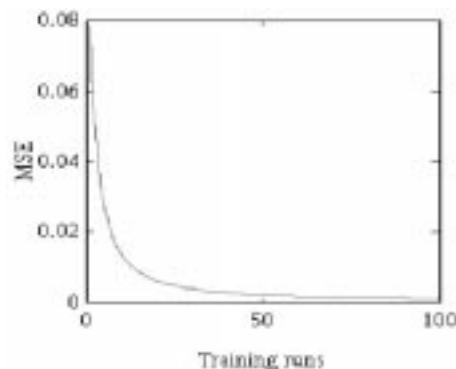
This paper presented a novel hybrid learning algorithm for neural networks. All important and necessary parameters of the algorithm are tested for the proposed model to show its capability for the implementation of complex problem. The proposed model is also implemented for the pattern recognition problem and compared with standard BP and LMS learning algorithms. The new proposed hybrid learning algorithm has shown better training performance, minimum error and better recognition rates. The overall minimum training time and smooth error reduction have found to be prominent features of the newly developed learning algorithm. Implementation of the proposed model for complex pattern recognition such as hand written digits and speech are amongst the major tasks for future study.



a) The proposed Hybrid Learning Algorithm



b) LMS Learning Algorithm



c) BP Learning Algorithm

Figure 6 - Learning Curves of Hybrid, LMS and BP Learning Algorithms

REFERENCES

1. S. Knerr, L. Personnaz, and G. Dreyfus Handwritten Digit Recognition by Neural Network with Single-Layer Training, IEEE Trans. On Neural Networks, 3, pp. 962-968, Nov. 1992.
2. A. K. Jain, J. Mao, J. and K. M. Mohiddin, Artificial Neural Network: A tutorial, IEEE Computers, pp. 31-44, March 1996.
3. W. A. Schmidt, and J. Davis, Pattern Recognition Properties of Various Feature Spaces for higher order Neural Networks, IEEE Trans. On PAMI, vol. 15, no. 8, pp. 795-801, Aug. 1993.
4. R. Lippman, An Introduction to Computing with Neural Network, IEEE ASSP Mag. Pp. 4-22, Apr. 1987.
5. A. Rajavlu, M. T. Musavi, and M. V. Shirvaikar, A Neural Network Approach to Character Recognition, Neural Networks, vol. 2, pp. 387-393, 1989.
6. D. Gabor, Communication Theory and Cybernetics, IRE Trans., CTI-4, pp. 19-31, 1954.
7. D. Gabor, W. P. Wilby, and R. Woodcock, Universal non-linear Filter Predictor Simulator which optimizes itself by a learning processes, Proc. IEE, vol. 108, Part B. no. 40, pp. 422-433, Jul. 1961.
8. J. Ahmad, "Novel Neural Network strategies based on the Gabor-Kolmogrov learning Algorithm for Pattern recognition and prediction", University of London, 1995.
9. D. E. Rumelhart, G. E. Hinton, and R. Williams, Learning internal representations by error propagation, in Parallel Distributed Processing (D. E. Rumelhart and J. L. McClelland, Eds.), Cambridge, MA:MIT Press, 1986, vol. 1, chapter 8.
10. B. Widrow and M. E. Hoff, "Adaptive Switching Circuits," IRE WESTCON CONV. Record, Part 4, pp. 96-104, 1960.
11. B. Widrow B. and M. A. Lehr, "30 Years of Adaptive Neural Networks Perceptron, MADALINE, and Backpropagation," IEEE, vol. 78, no. 9, pp. 1415-1442, Sep. 1990.

УДК 004.93

THE INTELLIGENT CIRCUIT THAT IS OPERATED BY TRANSMISSION OF IMPULSE AS SYMBOL OF ACTIVITY

Karasawa S.

Якщо вплив на датчиках представлений підмножиною імпульсів, а дії виконуючих пристроїв представлені іншою підмножиною імпульсів, тоді робоча пам'ять формується через дію одночасно шляхом поєднання точок, де існують імпульси. Інтелектуальна поведінка, що описана у логічній програмі, може бути досягнута за допомогою ланцюга, що передає імпульс як дію.

Если воздействия на датчиках представлены подмножеством импульсов, а действия исполнительных устройств представлены другим подмножеством импульсов, тогда рабочая память формируется через действие одновременно путем соединения точек, где существуют импульсы. Интеллектуальное поведение, описанное в логической программе, может быть достигнуто посредством цепи, которая передает импульс как действие.

If activities on sensors are represented by a subset of impulses and actions of actuators are represented by another subset of impulses, the working memory is formed through the activity concurrently by connecting the points where impulses exist. The intelligent behavior described in a flow chart can be achieved by means of the circuit that transfers an impulse as an activity.

1 INTRODUCTION

Neural engineering is an emerging discipline [1]. The greater part of artificial intelligence (AI) is researched by means of computational models. Those algorithms of the models are adapted to the digital computer. We appreciate the fruits of the software. The computer is omnipotent if the algorithm is perfect. Although the computing paradigms such as artificial neural networks [2] and connectionist models [3] are taking nature of a brain into account, the programs are difficult to form a working memory through an experience automatically.

The software met difficulties in the field of recognition if the object is things and affairs in the real world. An activity

is nonlinear and an activity is not continuous. The impulsive activity in the brain is not computational. The author as a researcher in the field of semiconductor electronics proposes the circuit where an impulse is transferred as an activity is transferred. The artificial intelligence in which the existence of impulse means occurrence of an action can be termed as hardware artificial intelligence.

Since a digital data is consist of signal on motionless state, the change of a state is represented by means of two states, i.e. one is obtained before the change, and the other is obtained after the change. There are two kinds of changes on a digital state, one is the positive impulse that causes an excitatory state, and the other is the negative impulse that inhibits an active state. A positive impulse that is the minimum unit of excitatory action is able to connect the lines at the transference, and the transmission of the positive impulse creates the working memory that makes possible to replay the same activity [4].

The action of connections makes possible to materialize the working memory. Although the definition of the working memory is difficult within the traditional concept of software, we can define the working memory as the function of a neuron, because the connections of a neuron are formed by means of activities, and the neuron decodes impulsive activities and it outputs an activity. The intelligence is a kind of activity and it is not restricted to the processing of information.

If we employ the segmentation of working memory in order to recognize things and affairs in the real world, the processing of recognition will become easy. If we divide a pattern into many pieces in order to compute, the processing of recognition becomes a jigsaw puzzle.

On the other hand, the intermittent operations of impulses make possible to carry out time-sharing of the operations. A circulating impulse in a loop is able to keep the activity, and the paralleling activities by means of transmit-

ting of impulses along plural flowcharts make possible to think while walking [5].

The activities by means of circuits driven by impulses are also possible to operate the information processing. The activity driven circuit that makes activity according to the transference of impulse is described by a flowchart on activities. The function of the flowchart is able to realize by means of impulse driven circuits. This new design concept is inspected through trial manufacturing and the results are reported in this paper.

2 PROBLEMS TO BE SOLVED BY MEANS OF THE CIRCUITS DRIVEN BY IMPULSES AS ACTIVITIES

We can memorize the first experience without lesson. But the computational learning model such as back propagation method [7] in neural network cannot form the circuits similar to the working memory that human possesses. It is difficult to process the signal by means of the central processing unit (CPU) that is under the construction.

The voice has an independent meaning only for the audience. The program and data are divided in a digital computer. Although the data are transferred, the machine cycle in a digital computer is not signified by the data. The meaning of signal is not given to wiring in the traditional technology. The digital signal that is independent from the media is difficult to form a working memory, for a working memory is a kind of circuit that is able to operate a function.

Another problem caused by the traditional concept of signal processing is difficulty in the processing for the recognition of things and affairs in the real world. The difficulty is caused by the structure of the signal. A digital computer deals with data adjusted to the signal processing, and there is the segmentation in a flow of signal processing. On the other hand, the things and affairs in the real world are possessed of the segmentation of itself. Since there are a great number of the combinations between two kinds of segmentations, the recognition of things and affairs in the real world becomes difficult. If we use the segmentation of the working memory for the recognition of things and affairs, the processing of recognition can be done at once.

The real activity is able to cover every information processing, but the information processing cannot cover every real activity. The serial processing of information is a part of the brain mechanisms. The traditional artificial intelligence cannot cover the function of consciousness in a brain well. The paralleling activities by means of transmitting of impulses along plural flowcharts make possible to emulate the communicating operations of the brain. The focused attention that is important for intelligent behavior will be emulated by means of the circuits designed by the concept of activity.

We can image the many new figures of a brain mechanism, if we remove the glasses of software. Although there are differences between human and machine, the interface that is designed by the concept of activity will contribute to decrease the gap between the activities in a brain and the traditional information processing.

3 THE METHOD TO ACQUIRE INTELLIGENT CIRCUITS THROUGH ACTIVITIES

3.1 The function of a neuron

There are a huge number of neurons in a brain. A number of impulses run through numbers of circuits. The activities can be memorized in the circuit if every signal is composed of the impulse. The subset of impulsive signals is packed by means of connections on a neuron at an instant. The neuron receives many impulses and it generates an impulse by means of a biochemical reaction.

Since resting potential in a neuron is -80mV sustained by many biochemical pumps and the narrow peak of positive impulse is $+40\text{mV}$ [6], the connections that the input terminals correspond to the points of existing positive impulses may form a decoder on the subset of impulses. A pattern of impulses is detected by the neuron formed through the same pattern of activities. This method of connections is able to form a working memory.

The subset of paralleling impulses or a series of intermittent impulses are transmitted by the working memory. The circuits for working memories are able to form automatically by making use of such activity driven system. The network of working memory is available for the cognitive device such as visual perception, or voice recognin.

3.2 Characteristics of the hardware artificial intelligence

The working memory is operated by means of minimum unit of activity where the process of making signal is omitted. The hardware artificial intelligence is based on the activity that depends on the body. The meaning of one impulse that comes from a sensor is able to understand as a reaction of the sensor under the situation. The activity of an actuator is able to understand from the effect that exerts on the surroundings. As for the activity of nerve cell, the meaning is contained in the wiring connections.

Since the nerve cell ignites in accordance with propagation of the impulse, the operation of a nerve cell is similar to the pin of the pin-drum of a music box. The activity of one pin produces one tone of sound. The whole activities of pins become music. The sophisticated reaction in the impulse driven system is achieved by means of linked activities. These activities are represented by a subset of impulses where many wiring are needed. In such case, although an activity in a cell has the meaning that depends on the cell, we cannot understand the meaning of an impulse well when we analyze the impulse that comes from only one nerve cell.

The concept that "an impulse is the representative of an activity in the circuit for intelligent activity" is different from the traditional concept of information processing system. The activity in the real world cannot separate from the substance and it cannot not free from the real time. The information that is translated from things and affairs in the real world makes possible to free from the real time.

The information of an impulse is only the timing that belongs to the real time. The device driven by activity is designed in order to make actions in the real time. In this design concept of impulse mode operation, the existence of

impulse in a circuit means the activity in the device. An impulse is the symbol of an activity.

Although the operation of working memory in an impulse system is similar to the digital decoder, the number of the input terminal of the working memory is adjusted with the object. Although there are intermittent transmissions in the impulse driven mode system, there is not the segmentation caused by the process of producing the signal of information.

The symbol itself is not computational, but it can be handled by means of circuits. These types of intelligent circuits will be useful for the electronic device that emulates the brain mechanism. Those circuits will be useful in the field of robotics or automaton.

3.3 Differences between digital circuits and activity driven circuits

We can replay an activity by a movie, but it is a copy on the visual image. The software is not able to function without hardware. But hardware is able to function without software. The impulse is the minimum signal on the activity, whereas the pulse is the minimum information.

The signal processing without comprehension such as a telecommunication stays within the world where activities improve expressions. The end user of a traditional signal processing is the human being. Fig.1 shows the architecture of a humanoid designed by the traditional concept of information processing.

The brain is not a mere signal processor, and it is an end user of the signal. The end user makes actions. The architecture of new machine is simple, but it operates every activity including information processing. Fig.2 shows the animal-robot in which the digital signal processing is omitted where only activities are transmitted along plural flowcharts.



Figure 1 - Block diagram of signal processing on a humanoid designed by traditional concept (Here, the data are transferred through media)

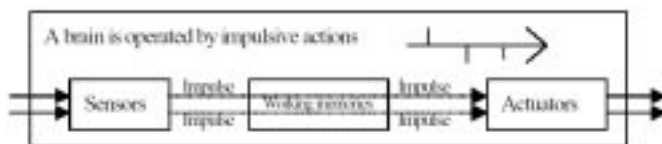


Figure 2 - The activity driven circuits for the model of brain mechanism where only impulses are transmitted as the symbol of activities (Brain of animal that is driven by transference of activities)

A combination of positive impulse and negative impulse is given by the derivative of a pulse with respect to time. But, the meaning of "impulse in the hardware artificial intelli-

gence" is different from that of "pulse in a traditional signal processing". In a traditional signal processing, the serial arrangement of pulses or the waveform of pulse carries information. But in the intelligence circuit driven by impulses as symbol of activities, the arrangement of series of impulses does not imply the meaning, and the waveform of an impulse does not imply the meaning. The frequent generation of impulses only indicates that the activity is breaking out frequently.

The electronic circuit driven by the activity can be manufactured with the method similar to digital circuits. That is, the activity is expressed with the presence of a certain amount of electric charge. The electric charge is transferred via the route selected by the logic circuit. The circuit in which the existing of electric charge means an operation of activity is realized by means of the semiconductor elements such as CCD (Charge Coupled Device) where the lines are connected by using floatation gate MOS FET (Metal Oxide Semiconductor Field Effect Transistor).

The implementation of working memory in the impulse driven system will be extremely easy compared with traditional digital technology of PLD (Programmable Logic Device) [8]. For example, if a pattern of activities is expressed with the pattern of the lighting spots, the pattern of lighting spots will be transferred via the route selected by the membrane in which a transparent point is formed by using the pattern of photochemical reactions on a membrane. The pattern of activities can be reproduced by using this path of lights.

4 THE HARDWARE ARTIFICIAL INTELLIGENCE THAT EMULATES BRAIN MECHANISM

4.1 The loop of delay elements for a short-term memory of activity

The function of a short-term memory for an activity is achieved by means of looped delay elements where circulation of an impulse represents a continuous activity. The excitatory state of this loop generates impulses continuously by the circulation. The plural paralleling impulses are able to excite one activity selected by the working memory that decodes the constituents.

There are activities of plural loops in the layer of constituents and that of one loop in the layered of output. The organization of activities yields activities of plural layers for the same agent. The meaning of each activity for the same agent is different. The organization of activities by means of working memories together with loops makes possible to form the frame of knowledge i.e. linkage of the activities.

The array of loops is able to operate the function of a register. The subset of impulses that should pay attention is transcribed on a register. The pattern of points activated on the register can be referred to the images those are inputted one after another successively. By using an activated image as a mark of alignment, a mapping of visual images can be carried out. In general, the consciousness that pays an attention to a certain image is realized through activities of the subset.

4.2 The organization of activities

The impulsive actions will be combined for a sophisticated activity. The subset of impulses is decoded by the working memory in which the intermittent activities are constituents. The linked working memories form a layered structure. A group of activities are decoded by means of multi-layered working memories.

The linguistic expression is transmitted in the form of subsets of impulses in a network of neurons. The faculty of language is acquired by means of the layered structure of working memories. The network of working memories is able to comprehend the meaning of speech voice and written sentence.

Since the organized structure makes use of existing circuits, the structure of extension economizes the circuit. There is compatibility in the element of component compared with the whole. The layered structure is able to represent a great many kinds things and affairs, because there are a great many kinds of combinations among many components. The things and affairs in the real world will be represented in the layered structure of working memories.

Since the output of working memory for a subset of serial activities is generated at the end of serial impulse, there are time differences among these impulses. The loop for a short-term memory is used in order to adjust the timing among plural activities. The working memory together with loops organizes sophisticated activities.

The linked activations of loops by means of a working memory make possible to provide impulses two directions. One direction is from the inputs side to the output side and the other direction is from the output side to the inputs side. The concurrent excitations of loops makes possible to form an afferent pathway and an efferent pathway. Here, the afferent pathway transfers a subset of components into the output, and the efferent pathway transfers a representative into the subset of components.

Positron Emission Tomography (PET) has enabled us to observe activities of a cerebrum [9]. The observations by PET show that the activities of cerebrum are held for a while. The working memory with loops that is driven by impulses will be able to emulate the brain activities.

4.3 The amplification of activities

The signal will be changed when the number of pulses changes in a traditional data processing where the arrangement of pulses is used as the signal. However, in the system driven by activities, the increase on the number of impulses is equal to the frequent activities. That is, to decrease the number of impulses is equal to the decrease of the frequency on activities.

Since the repetition of activity is carried out by an increase of impulses, the activity is emphasized by using of [OR] circuit for summing of impulse that was delayed. Since the repetition of an activity is diminished by a decrease of impulses, the activities are diminished by reduction of the number of impulses through AND circuit. Then, [OR] logic

circuit is able to use as an amplifier of activities. [AND] logic circuit is able to use as an attenuator of activities.

4.4 OR logic circuit for activities

Circuit of [OR] logic connection is needed where many circuits control an actuator. OR circuit outputs an impulse when an impulse inputs into one of the input terminals at least. However, OR circuit will ignite by a noise in case of many inputs. The automatic implementation of OR circuit makes such a problem.

Since the connections of AND circuit are equal to the points of existing impulses, the implementation of AND circuit can be carried out automatically. De Morgan's law can be stated as "AND of inverse of Boolean expressions is OR of the inverse of Boolean expressions". That is, OR logic on the complement of Boolean expression can be stated as "when even one pulse does not come the pulse is not output". The inhibitory (negative) impulse that suppresses excitatory (positive) impulses is used in the circuit driven by impulses.

Then, AND circuit for inverted inputs outputs an impulse when there is not an impulse. The output is used to suppress the positive impulse that is generated continuously. So, OR logic circuit is realized by using NOT, AND, and loop for generation of impulse. This circuit for OR logic connections will be able to emulate the activities on a cerebellum [10].

5 THE RESULTS; THE MANUFACTURING OF THE ACTIVITY DRIVEN CIRCUITS

5.1 The flowchart on a hardware artificial intelligence

Signal processing in a digital system is expressed by means of a flowchart. Each step of operation can be described by a proposition of "IF A=B, then Y=X". This logic is the fundamental statement for the software. The proposition can be carried out by a working memory. The flowchart on activity driven system expresses a flow of activity. Each step of activity can be described by a proposition of "IF present state of activities is A and the inputting activities is B, then the next state of activities is Y." The active state of this device is transferred to the other state assigned by the activities of conditions sent from the monitors of itself and outer world.

The format of flowchart on activity driven system is similar to that of digital system. But, the design of an activity driven system is based on the specification of activities.

Although a state in the activity driven system is represented by the state of circuit in which an impulse exists, the procedure to design the activity driven system is similar to the design of sequential system where one step of operation can be described by the change of state. The sequential system is known as "finite state machine". The tools and parts used in such digital system will be available in the activity driven system. The operations of an activity driven system is described by means of a state diagram and a timing diagram.

5.2 Manufacturing of the activity driven circuit that is made of digital IC

5.2.1 Digital circuits for the activity driven circuit

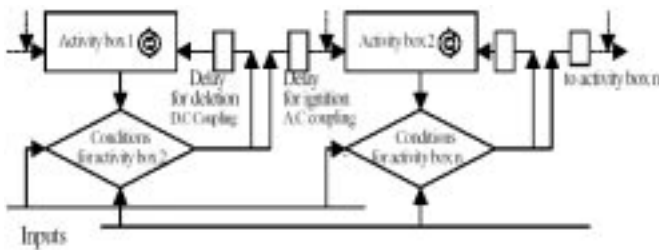


Figure 3 - The flowchart for transference of activities

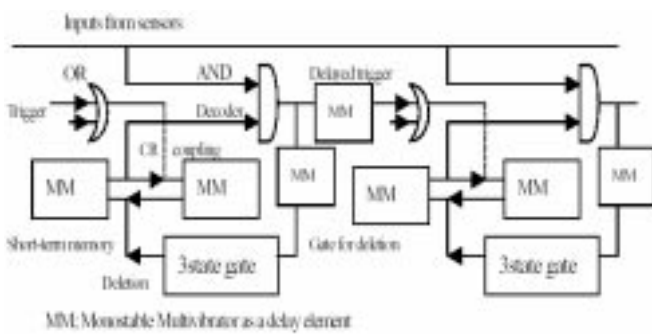


Figure 4 - The element of activity driven circuit that is composed of digital IC

Fig.3 illustrates the flowchart of transference of an activity. Fig.4 shows a circuit for the impulse driven mode. The plural outputs of monitors are connected by AND circuit in order to obtain the trigger signal that changes the present state. A loop made of a pair of Monostable Multivibrators (MM's) is used for a short-term memory. Two pieces of registers (1k ohm) are used as coupling elements of a loop, where the input into the input side of register is not forced to the output of another MM.

OR circuit connected from plural outputs to one next input is used in order to excite the next state. In order to delete an activity in a loop after the transition, the output of 3-state gate is connected to input port of the loop directly. When the control signal of deletion puts in the gate terminal of 3-state gate, the output is connected to the ground, and the rounding impulse is absorbed to the ground. When the control signal of deletion does not put in the gate terminal of 3-state gate, the output is unconnected to the ground.

5.2.2 Digital IC that is used for activity driven circuit

The activity driven circuit for a flowchart is comprised of delayed transferring element together with logic circuits as shown in Fig.4. The circuit is assembled by digital ICs listed as follows.

- a) Delay element (that is able to use as a timer for a controller) is realized by a MM; 74LS123 (14538B). The time period is given by the $T=0.7CR$ at 14538B;
- b) Decoder is composed of an [AND] circuit; 74LS08(4081);
- c) A loop for a short-term memory is made of a pair of MM;
- d) An impulse in a loop is deleted by means of a 3-state gate; 4LS244(74HC244);
- e) Common outputs are connected in a [OR] circuit; 74LS32(4071).

(Here, the registered number of IC is expressed as; TTL IC (c-MOS IC).

5.2.3 Trouble shooting of activity driven circuits

We manufactured several kinds of impulse driven control circuits [11], [12], [13], [14] Those circuits are the circuit for automatic vending machines, the circuit for automatic washing machine, the circuit for traffic controller, and the circuit for a sheet of short-term memories to replay the activity. Many Light Emitting Diodes (LED's) are used as monitors of the operations. These circuits functioned well.

We met some troubles. That is, the network of impulse driven circuits is delicate to the noise. Then, the chattering of switch changes the state. Moreover the trouble in activity driven circuit is transmitted by the impulse. The trouble shooting of activity driven circuit system is hard compared with the digital circuit.

5.3 The semiconductor device that is used for the activity driven circuit

The practical activity driven device must be manufactured as a semiconductor device. A quantity of electric charge is able to represent an activity. Since the function of a flow-chart is able to realize by means of a charge transfer device, the MOS technology will be useful in the field of impulse device. A dynamic Random Access Memory (RAM) consists of a capacitance where the charge is stored. The charge in the Charge-Coupled Device (CCD) can be moved along the surface of a semiconductor under the application of a proper sequence of voltage pulses, the charge packet can be transferred in a controlled manner. CCD is distinct from conventional device where current and voltage levels are generally used. However, CCD is able to realize digital logic circuits such as OR, AND, INVERT [15]. A charge-sensing amplifier nondestructively generates the same amount of another electric charge in one location and this information is able to transmit another location [16]. A memory element of a floating-gate avalanche-injection MOS memory (FAMOS) is fabricated as a dual-stacked poly-silicon structure in which the bottom gate is floating. The floating gate is isolated by thin oxide, and it provides the area for memorizing electrons. In EEPROM (Electrically Erasable Programmable Read-Only Memory), a circuit called a charge pump is used to generate the necessary programming voltage from the standard 5V, and the internal connection by means of the floating gate is provided to reverse the electron injection phenomenon [17]. Then, a pattern of activities can be copied to the pattern of

connections where the pre-injected electrons are removed. Since a positive impulse represents an activity, the function of a negative impulse is given by means of the circuit that absorbs the positive impulse.

6 SUMMARY

The characteristics on the activity driven circuits and the design tools are reported in this paper. We made several activity driven circuits by using digital IC's. The manufactured circuits functioned well. But, now, those circuits are not much superior than the digital circuits. In principle, the activity driven circuit can be designed to form through the experience. The function of flowchart that is designed by the concept of activity will be transferred into a charge transfer circuits. It is possible to fabricate activity driven circuits as a semiconductor integrated circuit. Such activity driven circuit will be superior to today's digital system.

The trial manufacturing of activity driven circuit without automatic implementation is the step to realize the activity driven system. The manufactured circuit operates the designed function. These experiences teach us the importance of timing on the operation.

As another fruit of this research, the concept of activity helps to manufacture the circuit model for the brain mechanism. The impulsive actions are combined for a sophisticated activity. The group of activities is decoded by means of a working memory. A loop is able to keep an activity. The consciousness is intermittently interlocked to the other activity. These linkages of the short-term memories make possible to explain the activities of a brain such as linguistic activity. The intermittent transmission of activity will provide the deeper understanding of the brain mechanism that has relations to the artificial intelligence.

Although the concept of activity driven circuit is based on the activity in the real world, the circuit is a model. The author hopes that this new concept of hardware artificial intelligence will contribute to developments of the technology of intelligence.

ACKNOWLEDGMENT

The author is grateful to them who manufactured the impulse driven circuits as graduation studies. Mr. Y. Uchiyumi made a control circuit for automatic vending machine [13], Mr. Y. Fujiwara made a control circuit for automatic

washing machine [14], and Mr. Y. Monma made a sheet of short-term memories. [15]. The author would like to thank Prof. Takeshi Kurobane and Prof. Kimio Shibayama at Tohoku Univ. for their advices and encouragements as author's former teachers.

REFERENCES

1. Proc. IEEE, Vol.89, No7, Edited by M. Akay, Special issue on Neural Engineering: Merging Engineering and Neuroscience pp. 991.
2. Schalkoff R.J., Artificial Neural Networks, The McGraw-Hill Co. Inc. 1997.
3. Levy J.P., Bairaktaris D, Bullinaria J.A., Cairns P., Connectionist Models of Memory and Language, UCL Press Limited, 1995.
4. Karasawa S., The strategy of impulse driven working memory for visual perception, Proc. of Inter. Conf. on Imaging Science, Systems, and Technology, June 24-27, Las Vegas, Vol.2, pp. 729-735, 2002.
5. Karasawa S., Impulse recurrent loops for short-term memory which merges with experience and long-term memory, Proc. of 3rd Int. Conf. on Cognitive and Neural systems, pp. 36, Boston Univ., May 26-29, 1999.
6. Nicholls J.G., Martin A.R., Wallace B.G., From Neuron to Brain, pp. 12, Sinauer Associates, Inc. Publishers, 1992.
7. Karasawa S., Decoder-for-neuron translation models for brain mechanism, Research report, Miyagi National College of Technology, pp. 19-25, 2000.
8. Carter J.W., Digital Designing with Programmable Logic Devices, Prentice-Hall, Inc, 1997.
9. Honjo I., Language Viewed from the Brain, Karger, ISBN 3-8055-6789-8, 1999.
10. Karasawa S., Oomori J., Impulse circuits for a distributed control inspired by the neuroanatomical structure of a cerebellum, Intelligent engineering systems through artificial neural networks, Vol.10, pp. 171-190, ASME press, New York, 2000.
11. Karasawa S., Monma Y., Fujiwara Y., and Uchiyumi Y., Semiconductor devices for intelligent circuit that is operated by means of transferring impulses, Academics lecture meeting, Northeast branch of Japan Society of Applied Physics, pp.26-27, Dec. 5, 2002.
12. Uchiyumi Y., Manufacturing of an impulse driven circuit for the control of vending machine, Graduation Thesis at Department of Electrical Engineering, Miyagi National College of Technology, March 4, 2003.
13. Fujiwara Y., Manufacturing of an impulse driven circuit for the control of automatic washing machine, Graduation Thesis at Department of Electrical Engineering, Miyagi National College of Technology, March 4, 2003.
14. Monma Y., Manufacturing of a sheet of impulse driven short-term memories, Graduation Thesis at Department of Electrical Engineering, Miyagi National College of Technology, March 4, 2003.
15. Zimmerman T.A., Allen R.A., and Jacobs R.W., Digital charge-coupled logic, IEEE Journal of solid-state circuits, Vol.SC-12, No. 5, pp. 473-485, Oct. 1977.
16. Handy R.J., Use of CCD in the Development of digital logic, IEEE Trans. ED, Vol.24, No.8, pp. 1049-1061, 1977.
17. Geiger R.L., Allen P.E., and Strader N.R., VLSI design techniques for analog and digital circuits, pp. 826, McGraw Hill Publishing Co. 1990.

УДК 681.32:007.52

SOLVING OF BOUNDARY VALUE PROBLEMS FOR MATHEMATICAL PHYSICS EQUATIONS IN CELLULAR NEURAL NETWORKS

M.A.Novotarskiy

Стаття присвячена вивченню застосування локально-асинхронних методів до вирішення крайових задач математичної фізики на кліткових нейронних мережах. Розглянуті архітектури традиційної аналогової та дискретної нейрон-

них мереж і архітектура багатопшарової кліткової нейронної мережі з багатоклітковим згладжуванням. Представлений алгоритм асинхронного функціонування для багатопшарової кліткової нейронної мережі.

Статья посвящена изучению применения локально-асинхронных методов для решения краевых задач математической физики на клеточных нейронных сетях. Рассмотрены архитектуры традиционной аналоговой и дискретной клеточных нейронных сетей, а также архитектура многослойной клеточной нейронной сети с мультисеточным сглаживанием. Представлен алгоритм асинхронного функционирования для многослойной клеточной нейронной сети.

This paper is devoted to studies of application of locally - asynchronous methods to the boundary value problems for mathematical physics equations solving in cellular neural networks. The following approaches are considered: traditional analog cellular network architecture, digital cellular network architecture, and multilayer cellular network with multigrid smoothing. The algorithm of asynchronous operation for the multilayer cellular neural network represented.

INTRODUCTION

The locally - asynchronous calculation concept is based on representation of a global task by a set of local tasks that are realized as independent local processes cooperating among themselves with the purpose of an interchanging by intermediate information. The interacting processes are implemented in cells of neural network [1]. We call them asynchronous, if no restrictions on interchange time exist. Localization is determined as the fixed amount of connections between neurons, which is much less than a possible amount of neurons in the neural network. Such organization of processes is characteristic for a living nature, therefore is rather actual to use it in an artificial neural network for realization of such concept.

1 MATHEMATICAL MODEL OF LOCALLY - ASYNCHRONOUS COMPUTATIONS

We will examine mathematical locally - asynchronous model of the form:

$$AU = F, \quad (1)$$

where A - the differential operator,

$U = U(x)|_{(x \in D \subset R^r)}$ with Dirichlet boundary conditions of the arbitrary geometrical form.

The boundary is considered as a point set with an amount sufficient for a faithful representation of boundary. Let us select M points $\bar{p}_1, \bar{p}_2, \dots, \bar{p}_M \in B$ for describing boundary B . Then the boundary conditions of Dirichlet look like:

$$U(\bar{p}_i) = b_i, \quad i = 1, 2, \dots, M. \quad (2)$$

We approximate the equation (1) on a homogeneous grid Ω with step h . Then $U_h = U(x_h)|_{(x_h \in D_h)}$. The resulting system of the linear algebraic equations will look like:

$$A_h U_h = F_h, \quad (3)$$

where A_h - difference operator; F_h - some grid function;

U_h - difference solution.

Arbitrary iterative one-step method for deriving an approximate solution of the equation (1) can be noted as:

$$B_n U_h(n+1) = C_n U_h(n) + \Phi_n, \quad n = 0, 1, 2, \dots, \quad (4)$$

where $B_n, C_n, \Phi_n: H \rightarrow H$.

Having made replacements $D = \frac{B_n - C_n}{\tau_{n+1}}, \Phi_n = \varphi \tau_{n+1}$

we shall receive two-layer iterative scheme

$$B_n \frac{U_h(n+1) - U_h(n)}{\tau_{n+1}} + D U_h(n) = \varphi, \quad n = 0, 1, 2, \dots \quad (5)$$

From here

$$U_h(n+1) = U_h(n) - \tau_{n+1} B_n^{-1} (D U_h(n) - \varphi). \quad (6)$$

Using (6) the iterative scheme is obtained

$$U_h(n+1) = U_h(n) - \rho [A_h (D U_h(n) - \varphi)]. \quad (7)$$

Locally - asynchronous method [3] defines the order of application of the iterative scheme (7) to nodes of grid area, on which the solution of a problem (1) exists. Such order is described with usage of a concept of a random sequence $\{J_n\}_{n=1}^{\infty}$ of nonempty subsets $\{1, 2, \dots, \eta\}$. Using this sequence, we create a sequence of iterations $\{U_h(n)\}_{n=1}^{\infty}$ according to a rule:

$$U_h^i(n) = \begin{cases} U_h^i(n-1), & i \notin J_n, \\ f^i[U_h^1(s_1(n)), U_h^2(s_2(n)), \dots, U_h^\eta(s_\eta(n))], & i \in J_n, \end{cases} \quad (8)$$

where U_h^i - is component of a vector U_h , and $\{s_i(n)\}_{n=1}^{\infty} = 1, 2, \dots, \eta$ - is a sequence of non-negative numbers, which meet conditions: $s_i \leq (n-1)p, s_i(n) \rightarrow \infty$.

The delay s_i allows using components of a vector of the previous iterations at an evaluation of a vector of the flowing iteration. This fact strictly justifies a possibility of usage of the asynchronous conditions for computations on parallel structures.

2 ANALOG CELLULAR NEURAL NETWORK ARCHITECTURE

The cellular neural networks [1] today are widely used for solving systems (3). Their structures are very similar to the structure of cellular automata, as they contain neurons, which are connected only to the nearest neighbors. An example of a two-dimensional cellular neural network is shown in fig. 1.

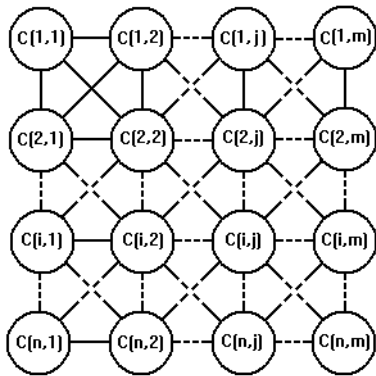


Figure 1 - A two-dimensional cellular neural network

It consists of $n \times m$ nodes arranged in n rows and m columns. The nodes of a regular graph are neurons with their neighbor connections. The r -neighborhood for the neuron $C(i, j)$ is determined in [1] as a subset of the cellular neural network that meets conditions:

$$N_r(i, j) = \left\{ C(k, l) \mid \max\{|k - i|, |l - j|\} \leq r, \begin{cases} 1 \leq k \leq n, 1 \leq l \leq m, \\ ((k, l) \in D_h) \end{cases} \right\}, \quad (9)$$

where r is a positive integer number.

For the two-dimensional cellular neural network shown in fig. 1, we have

$$N_1(i, j) = \{C(i-1, j), C(i, j), C(i+1, j), C(i-1, j-1), C(i, j-1), C(i+1, j-1), C(i-1, j+1), C(i, j+1), C(i+1, j+1)\}, r = 1. \quad (10)$$

The r -neighborhood $N_r(i, j)$ always includes a central neuron $C(i, j)$ that provides a symmetry property. If $C(i, j) \in N(k, l)$, then $C(k, l) \in N(i, j)$, for all $C(i, j)$ and $C(k, l)$ in a cellular neural network.

There are many equivalent circuits of neuron $C(i, j)$ now [1,2]. They depend on a type of a problem. One of them, useful for obtaining solution of system (3), was offered in [1]. It can be implemented with usage of operational amplifiers, as shown in fig. 2.

The voltages $U_{k,l}$, $U_{i,j}$ simulate output signals from neighbor neurons and feedback from the output signal. The node voltage $V_{i,j}$ is called the state of neuron $C(i, j)$. The circuit consists of linear capacitor C_v , linear resistor R_v , linear voltage controlled current source with the characteristic $I_v(i, j; k, l) = A(i, j; k, l)U_{i,j}$, $I_u = B(i, j; k, l)U_{k,l}$ and a subcircuit with piece-wise-linear function

$$U_{i,j} = 0.5(|V_{i,j} + 1| - |V_{i,j} - 1|). \quad (11)$$

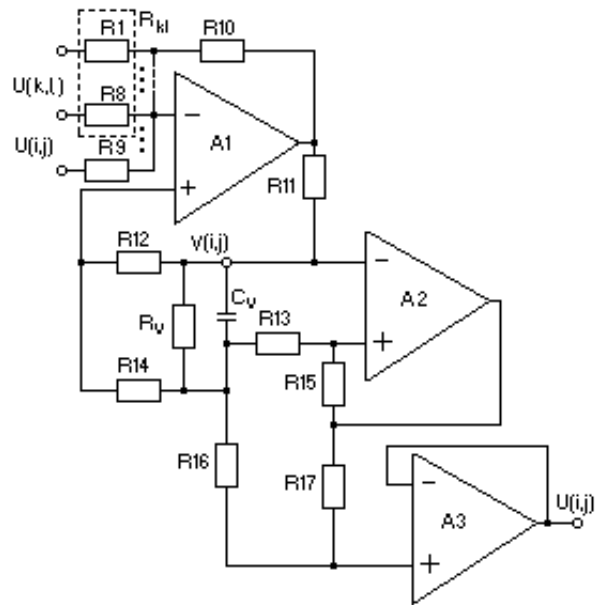


Figure 2 - Implementation of neuron $C(i, j)$ for the cellular neural network with $r = 1$

The internal state of the neuron is set by the equation:

$$C_v \frac{d}{dt} V_{i,j}(t) = \frac{V_{i,j}(t)}{R_v} + \sum_{\substack{k,l \\ C(k,l) \in N(i,j)}} A(i, j; k, l) U_{i,j}(t) + \sum_{\substack{k,l \\ C(k,l) \in N(i,j)}} B(i, j; k, l) U_{k,l}(t) + I. \quad (12)$$

Let's consider a two-dimensional heat equation:

$$\frac{\partial^2 u(x, y, t)}{\partial x^2} + \frac{\partial^2 u(x, y, t)}{\partial y^2} = \frac{1}{k} \frac{\partial u(x, y, t)}{\partial t}. \quad (13)$$

The heat equation (13) can be approximated by a set of equations:

$$u_{i,j}(t) \frac{8}{k} \frac{du_{i,j}(t)}{dt} = u_{i-1,j}(t) + u_{i+1,j}(t) + u_{i,j}(t) + u_{i,j-1}(t) + u_{i-1,j-1}(t) + u_{i,j-1}(t) + u_{i+1,j-1}(t) + u_{i-1,j+1}(t) + u_{i,j+1}(t) + u_{i+1,j+1}(t). \quad (14)$$

Using cellular neural networks for solving differential partial equations is grounded on physical analogy between the equation (12) and equation (14). The solving process is combined with learning process. It consists of searching a stationary state of the cellular neural network that is a solution of a set of equations (14).

Analog neural networks provide the high rate of solution but relative low accuracy. The alternate way is usage of digital neurons in cells of the cellular neural network. It allows applying the numerical methods to solution of a set of equations (3).

3 DIGITAL CELLULAR NETWORK ARCHITECTURE

The digital cellular neural network structure is similar to analog structure but it consists of two layers, because iterations are carried out in two steps. The two-dimensional network is shown in fig. 3.

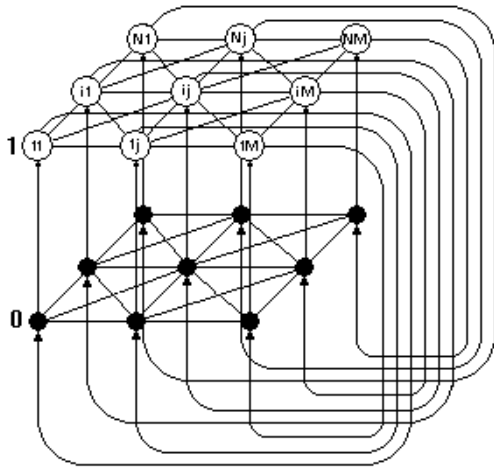


Figure 3 - A two-dimensional digital cellular neural network

Each layer of the neural network performs specific calculations; therefore neurons in layers have different structures. Let us consider two steps of iterations implemented by the two-layer cellular neural network.

Layer 0. In this layer neurons calculate the discrepancy for points of discrete space:

$$d_h(n) = F_h - A_h U_h(n). \quad (15)$$

For calculations, we use a neuron model, shown in fig.4. It contains a sigma-pi unit for product $A_h U_h(n)$ calculations and bias F_h subtraction. In both layers, we use the piecewise-linear activation function, which is typical of computing neural networks.

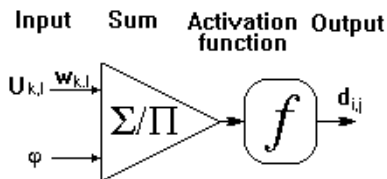


Figure 4 - A layer 0 neuron model

Layer 1. This layer forms new approximation of the solution vector by the iteration expression (7). A layer 1 neuron model is shown in fig.5. One can describe it as a mathematical processor that receives input signal in the form of r -dimensional vector $U(n)$ and produces scalar output $U_{i,j}(n+1)$.

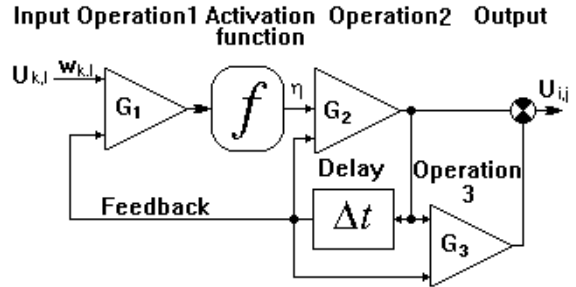


Figure 5 - A layer 1 neuron model

It realizes the mapping operation:

$$Ne: U(t) \rightarrow U_{i,j}(t+1), \quad (16)$$

where $U = \{U_{k,l}\}, \forall D(k,l) \in N_r(i,j), 1 \leq k \leq n, 1 \leq l \leq m$.

The neuron can be a digital controller or computer program. In both cases, it should contain operational units and f - unit. Δt - unit is used for simulation of computation time delay. Feedback connection provides execution of internal iterations according to locally - asynchronous algorithm.

In common case one can describe Ne -mapping as

$$\begin{aligned} U_{i,j}(t+1) &= U_{i,j}(t) + \eta f(WU(t) + C) = \\ &= U_{i,j}(t) + \eta f\left(\left[\begin{matrix} w_{k,l} U_{k,l}(t) + c_{k,l} \end{matrix} \right]\right), \end{aligned} \quad (17)$$

where $w_{k,l} \in W, c_{k,l} \in C, \forall D(k,l) \in N_r(k,l), 1 \leq k \leq n, 1 \leq l \leq m$.

Ne -mapping (17) is a physical analog of the iterative formula (7). Being grounded on this analogy, we shall set a way of definition of weight coefficients W . Generally W is a matrix interpretation of differential operator A_h . Therefore, the process of creation of a matrix of weight coefficients W is considered as a stage of synthesis of the cellular neural network by cloning a template.

One more function of this layer consists in detection of stopping moment at obtaining convergence of asynchronous algorithm. For this purpose, the neurons evaluate expression:

$$\frac{|U_h(n+1) - U_h(n)|}{|U_h(n)|} < \varepsilon. \quad (18)$$

If convergence has been obtained, the vector U_h is the problem solution; otherwise, we go to the layer 0 again.

4 MULTILAYER CELLULAR NETWORK WITH MULTIGRID SMOOTHING

The asynchronous way of operation of the digital cellular neural network can cause a mismatch in speed of operation of neurons. Such mismatch reduces convergence speed of the iterative method. Usage of multigrid smoothing [4, 5] allows achieving essential improvement of convergence speed. Some accelerating schemes have been proposed to modify standard multigrid method under different circumstances [6, 7, 8]. Here we present multilayer cellular neural network paradigm with multigrid smoothing.

Let $U_h(n)$ - some approximating of the equation (1) solution U_h , then the absolute error and discrepancy of the given iterative step are determined by expressions:

$$V_h(n) = U_h - U_h(n), d_h = F_h - A_h U_h(n). \quad (19)$$

Outgoing from (19) equations of a discrepancy looks like:

$$A_h V_h(n) = d_h(n). \quad (20)$$

According to [4] we receive solutions of the equation (20) on a subset of grids $\Omega^{(L)}, \Omega^{(L-1)}, \dots, \Omega^{(1)}, \Omega^{(0)}$, with steps

$$h^{(l)} = h^{(L)} 2^{L-l}, l < L, \quad (21)$$

where l - number of a flowing coarse grid.

Linear projective operator P carries out the transition from fine grid to rougher grid

$$d_{h^{(l-1)}}(n) = P^{(l)} d_{h^{(l)}}(n), \quad (22)$$

and the transition in an opposite direction is ensured with an interpolating operator

$$V_{h^{(l)}}(n) = I^{(l-1)} V_{h^{(l-1)}}(n). \quad (23)$$

Computation of the absolute error at the given grid level is performed with usage of the difference scheme

$$V_{h^{(l)}}(n+1) = V_{h^{(l)}}(n) - \rho [A_{h^{(l)}} V_{h^{(l)}}(n) - d_{h^{(l)}}(n)]. \quad (24)$$

The cycle of computations on coarse grids is completed by correction of the grid function:

$$U_h(n) = U_h(n) + V_{h^{(L)}}(n). \quad (25)$$

There is a set of multigrid algorithms (V -cycle, W -cycle etc.), which use in various sequences base operations of smoothing, restricting and interpolating. Each of variants of multigrid algorithm can appear effective for the separate problem.

Therefore, we speak about the multigrid technique, which defines some ideology of solution of boundary value problems.

The multilayer cellular neural network shown in fig. 6 performs computations to solve (3) and (20).

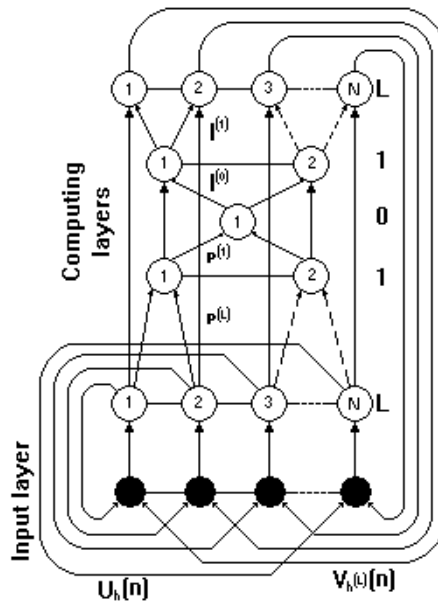


Figure 6 - A multilayer cellular neural network

It consists of the input layer and $2L + 1$ computing layers. The neurons of an input layer are intended for the grid function $U_h(n)$ correction according to expression (25). All computing layers are doubled except for the layer 0. These layers are intended for implementation of a multigrid smoothing algorithm. For simplicity, the one-dimensional variant of the neural network is represented in fig 6.

In the layer L . each neuron corresponds to the point of a fine grid with step $h = h^{(L)}$. According to expression (21) step $h^{(L-1)}$ is magnified twice in comparison with step $h^{(L)}$. Such relation of steps is accepted as a matter of convenience but this is not necessary.

The implementation of a multigrid locally - asynchronous method is determined by the following algorithm.

```

For  $1 \leq i \leq N$  do Choose initial value  $U_i(0)$ ;
1: For  $0 \leq l \leq L$  &  $(0 \leq j \leq \frac{Nh^{(L)}}{h^{(l)}})$  do
begin
Choose initial value  $V_j(0)$ 
 $d^{(l)} := F^{(l)} - A^{(l)} U^{(l)}$ ;
 $d^{(l-1)} := P^{(l)} d^{(l)}$ ;
 $V_j^{(l)}(n+1) := V_j^{(l)}(n) - \rho [A^{(l)} V^{(l)}(n) - d^{(l)}(n)]$ ;
 $V^{(l)} := I^{(l-1)} V^{(l-1)}$ ;
end;
For  $0 \leq i \leq N$  do
begin
 $U_i := U_i + V_i^{(L)}$ ;
 $U_i(n+1) := U_i(n) - \rho [A^{(L)} U^{(L)} - \phi^{(L)}]$ ;
if then  $\frac{U_i(n+1) - U_i(n)}{U_i(n)} < \varepsilon$  then exist else goto 1;
end;

```

This algorithm sets a sequence of data conversions that mainly are the characteristics of one neuron. We do not circumscribe an exact sequence of interneuron data exchange, as it cannot be defined because of asynchronous character of interaction.

The algorithm of asynchronous iterations (8) has allowed applying the numerical methods for PDE solving on cellular neural networks. We used multigrid smoothing to improve convergence speed. Our numerical results showed that remarkable accelerations rates were achieved by these techniques.

REFERENCES

1. Chua L. O., Yang L. Cellular Neural Networks Theory// IEEE Transactions on Circuits and Systems, 1988, vol. 35, No10,

- pp. 1257 - 1272.
2. Lijuan S., Panzhen W., Xinyu W. The Application of Cellular Neural Networks for Solving Partial Differential Equations// The Journal of China Universities of Posts and Telecommunications, 1997, vol. 4, No.1, pp.45-50.
3. Baudet G.M. The Design and Analysis of the Algorithms for Asynchronous Multiprocessors. Ph.D. Diss.-Pittsburg: Carnegie-Mellon Univ., 1978, PA, 182 p.
4. McCormic S., Rodrigue G.H. Multigrid Methods for Multiprocessor Computers. Techn. Rept. - Lawrence Livermore: Lab. Livermore CA, 1979, 29 p.
5. Wesseling P. An Introduction to Multigrid Methods. John Wiley & Sons, Chichester, 1992, 294 p.
6. Brandt A., Yavneh I. Accelerated multigrid convergence and high-Reynolds recirculating flows// SIAM Journal on Scientific Computing., 1993, vol. 14, pp. 607-626.
7. Reusken A. Steplength optimization and linear multigrid methods// Numer. Math, 1991, vol. 58, pp. 819-838.
8. Vaněk P. Fast multigrid solver// Appl. Of Math., 1995, vol.40, pp. 1-20.

УДК 681.2.06

АППАРАТНО-ПРОГРАММНАЯ РЕАЛИЗАЦИЯ АЛГОРИТМА ВОССТАНОВЛЕНИЯ ИНТЕРПОЛЯЦИОННЫМИ СПЛАЙНАМИ

А.Ю.Балыкова, В.М.Жигачев, Б.В.Чувькин

Розглянуто питання проектування інтерполюючих аналого-цифрових фільтрів, що реалізують алгоритми сплайн-інтерполяції. Приведено приклади розрахунку варіантів структур АЦФ і результати імітаційного моделювання.

Рассмотрены вопросы проектирования интерполирующих аналого-цифровых фильтров, реализующих алгоритмы сплайн-интерполяции. Приведены примеры расчета вариантов структур АЦФ и результаты имитационного моделирования.

The questions of designing of an interpolation analog-digital filters (ADF) realizing algorithms of spline-interpolation are considered. The examples of realization of variants of structures ADF and results of imitating modeling are given.

Использование сплайнов в вычислительной математике началось в 70-е годы. Они оказались эффективным инструментом в решении задач приближения функций, восстановления функций по неполной информации, сглаживания экспериментальных данных, цифровой фильтрации. В измерительной технике сплайны нашли широкое применение при решении задач синтеза весовых функций помехоустойчивых аналого-цифровых преобразователей (АЦП) и интерполирующих фильтров в первую очередь благодаря возможности их аппаратной реализации в аналого-цифровой форме с высокой точностью.

Рассмотрим наиболее важные свойства сплайнов.

Сплайны - это финитные функции, составленные из "кусочков" многочленов данной степени, которые состыкованы так, чтобы получившаяся функция была непрерывной и обладала несколькими непрерывными производными, т.е. была достаточно "гладкой".

Полиномиальный сплайн степени n для равномерных отсчетов (сплайн Шенберга) имеет вид [1]:

$$B_n(t) = \int_{k=0}^{n+1} (-1)^k C_{n+1}^k (t-k)_+^n, \quad (1)$$

где знак "+" обозначает одностороннее ограничение:

$$t_+^n = \begin{cases} t^n, & t > 0; \\ 0, & t \leq 0. \end{cases}$$

Если рассматривать полиномиальный сплайн (1) как импульсную характеристику (ИХ) некоторого аналого-цифрового фильтра (АЦФ), то можно найти его передаточную функцию через преобразование Лапласа:

$$H_n(p) = L(B_n(t)) = \left(\frac{1 - \exp(-ph)}{ph} \right)^n, \quad (2)$$

где L - оператор преобразования Лапласа, p - оператор дифференцирования, h - шаг дискретизации. Поскольку оператору аналоговой задержки $\exp(-ph)$ на шаг дискретизации h для цифрового входного сигнала соответствует оператор дискретной задержки z^{-1} , то передаточная функция (2) АЦФ можно представить в виде функции двух переменных, что соответствует аналоговой и цифровой представлению информации:

$$H_n(z, p) = (1 - z^{-1})^n \cdot (ph)^{-n}. \quad (3)$$

В соответствии с формулой (3) реализация алгоритма восстановления цифрового сигнала интерполяционным сплайном состоит из процедур формирования n -ой конеч-

ной разности и n -кратного интегрирования последовательности δ -импульсов.

Формально структуру АЦФ с передаточной функцией (3) можно представить в виде последовательного включения типовых динамических звеньев: конечной разности $(1 - z^{-1})$, формирователя δ -импульсов и интегрирования $1/(hp)$. Функция сплайна нулевого порядка ($n = 0$) реализует цифроаналоговый преобразователь (ЦАП), математическая модель которого определена в виде $(1 - z^{-1})/hp$. С учетом вышесказанного структура АЦФ с передаточной функцией (3) будет иметь вид, приведенный на рис. 1.

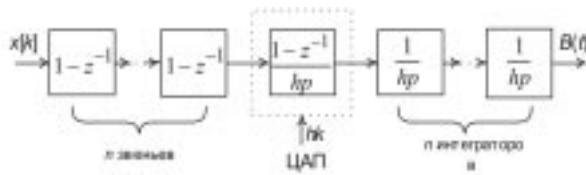


Рисунок 1

Полиномиальные сплайны Шенберга степени $n = 0$ и $n = 1$ являются интерполяционными (ступенчатая и линейная интерполяция), т.е. для них выполняется условие интерполяции:

$$B_n(hk) = \begin{cases} 1, & k = m; \\ 0, & k \neq m, \end{cases} \quad (4)$$

где m - номер узла интерполяции, $k = 0, 1, \dots, n$. В случае более высоких порядков условие (4) не выполняется и поэтому необходимо ввести дополнительный или цифровой корректирующие фильтры. Аналоговый корректирующий фильтр будет иметь передаточную функцию вида:

$$R_n(p) = a_0 + a_1p + \dots + a_kp^k + \dots + a_np^n. \quad (5)$$

Коэффициенты a_k находятся из условия интерполяции (4). Ниже в табл. 1 приведены значения коэффициентов a_k для интерполяционных сплайнов третьего и четвертого порядка [2].

Таблица 1 - Коэффициенты интерполяционных сплайнов

| N | a_0 | a_1 | a_2 | a_3 | m |
|---------|-------|-------|-------|-------|-----|
| $n = 2$ | 1 | 1/2 | 0 | - | 1 |
| | 1 | -1/2 | 0 | - | 2 |
| $n = 3$ | 1 | 1 | 1/3 | 0 | 1 |
| | 1 | 0 | -1/6 | 0 | 2 |
| | 1 | -1 | 1/3 | 0 | 3 |

Передаточную функцию корректирующего фильтра (5) можно реализовать путем введения в структуру АЦФ сумматора с весовыми коэффициентами a_k . Соответствующая структура аналоговой части АЦФ приведена на рисунке 2.

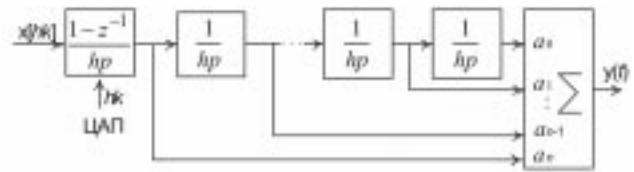


Рисунок 2

На рисунке 3 приведен пример кубического интерполяционного сплайна $y_3(t)$ ($n = 3, m = 2, a_1 = 0, a_2 = -1/6, a_3 = 0$). Для сравнения приведен сплайн Шенберга $B(t)$ этого же порядка.

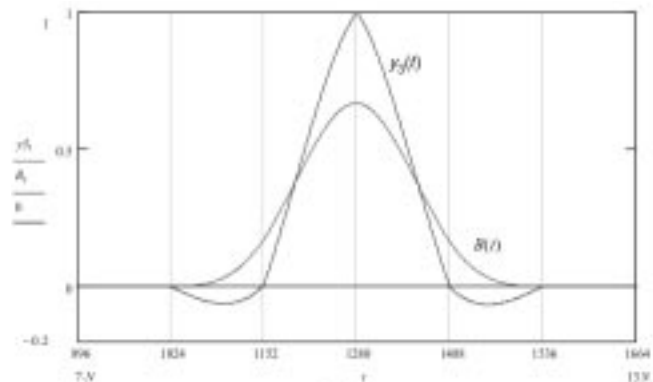


Рисунок 3

Для алгоритма цифровой коррекции передаточная функция ЦФ задается в виде:

$$R_s(z) = (a_0 + a_1(z + z^{-1}) + \dots + a_k(z^k + z^{-k}) \dots + a_s(z^s + z^{-s})) \cdot z^{-s}. \quad (6)$$

Коэффициенты находятся из условия интерполяции (4). Соответствующая структура цифровой части АЦФ приведена на рисунке 4.

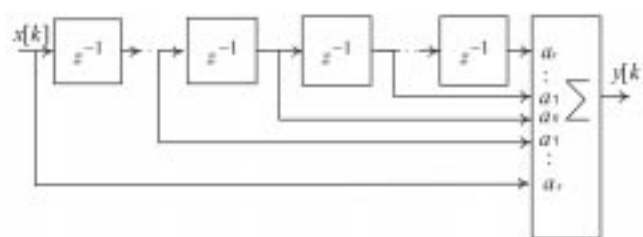


Рисунок 4

В качестве примера рассмотрим интерполяционный сплайн для $n = 3$ [3]

$$y_3(t) = B_3(t) - a_1(B_3(t+h) + B_3(t-h)) + a_2(B_3(t+2h) + B_3(t-2h)) - \dots - a_s(B_3(t+sh) + B_3(t-sh)), \quad (7)$$

где коэффициенты $a_k = (2 - \sqrt{3})^k$.

Значение s можно ограничить порядком 3-5, поскольку ряд коэффициентов a_k быстро сходится. На рисунке 5 приведен пример кубического интерполяционного сплайна $y_3(t)$ ($n = 3$) для варианта $s = 3$. Точность выполнения условия интерполяции в крайних точках интервала ИХ соответствует 10^{-3} . При необходимости точность интерполяции в узлах может быть повышена за счет увеличения s , но при этом увеличится длительность ИХ и потребуются увеличение операций умножений и суммирования при реализации алгоритма корректирующего фильтра. Для сравнения приведен сплайн Шенберга $B(t)$ этого же порядка (рис.5).

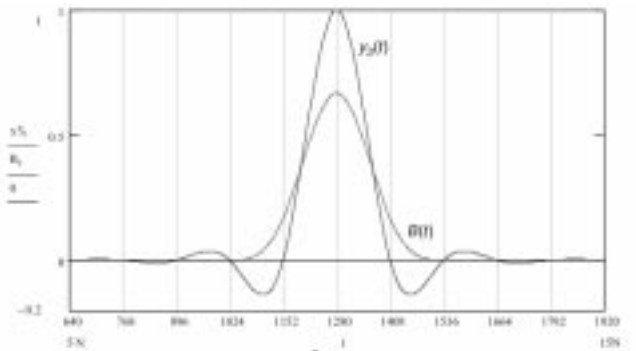


Рисунок 5

Рассмотренные выше сплайн-интерполирующие АЦФ имеют разомкнутую структуру, состоящую из звеньев не удовлетворяющих условию физической реализуемости (аналоговые интеграторы). Для обеспечения устойчивости АЦФ необходимо ввести цепи обратной связи (ОС), т.е. перейти к замкнутой структуре. Замкнутые структуры АЦФ могут отличаться вариантами ОС. Обобщенный вариант канонической структуры многопетлевой ОС приведен ниже. В этом случае в цепях ОС последовательно включены АЦП, нерекурсивные ЦФ с передаточными функцией $W_{n,k}(z)$ и рекурсивный ЦФ с передаточной функцией $1/V_{n,k}(z)$ в прямом канале преобразования.

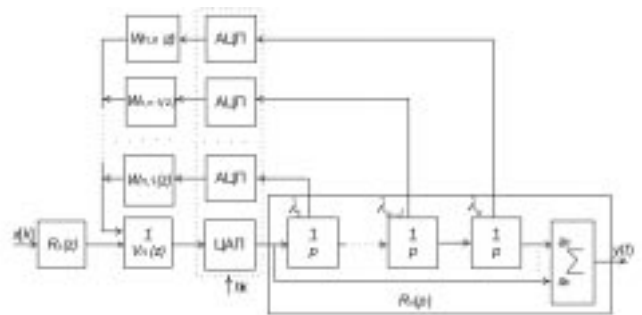


Рисунок 6

Коэффициенты передаточных функций $W_n(z)$ и $1/V_n(z)$ находятся из условия абсолютной устойчивости замкнутой структуры по критерию финитности ИХ АЦФ и приведены в таблице 2.

Таблица 2 - Коэффициенты передаточных функций

| $n = 1$ | | | $n = 2$ | | | | | | | | |
|---|-------------|-------------|-----------------------------|-------------|-------------|----------------------------|-------------|-------------|---------------|-------------|-------------|
| λ_1 | | | λ_1 | λ_2 | | λ_1 | λ_2 | | λ_1 | λ_2 | |
| -1 | | | 0 | -5/2 | | -3/2 | | | -11/6 | -2 | |
| $W_{1,1} = 1$ | | | $W_{2,1} = 0$ | | | $W_{2,1} = 1$ | | | $W_{3,1} = 1$ | | |
| $V_1 = 1$ | | | $W_{2,2} = 1 - 3/5z^{-1}$ | | | $W_{2,2} = 1$ | | | $W_{3,2} = 1$ | | |
| | | | $V_2 = 1 + 3/4z^{-1}$ | | | $V_2 = 1$ | | | $W_{3,3} = 1$ | | |
| | | | | | | | | | $V_3 = 1$ | | |
| $n = 3$ | | | | | | | | | | | |
| λ_1 | λ_2 | λ_3 | λ_1 | λ_2 | λ_3 | λ_1 | λ_2 | λ_3 | λ_1 | λ_2 | λ_3 |
| 0 | 0 | -35/36 | 0 | -23/6 | -1 | -1 | 0 | -3 | -11/6 | -2 | -1 |
| $W_{3,1} = 0$ | | | $W_{3,1} = 0$ | | | $W_{3,1} = 1 - 23/6z^{-1}$ | | | $W_{3,1} = 1$ | | |
| $W_{3,2} = 0$ | | | $W_{3,2} = 1 - 11/23z^{-1}$ | | | $W_{3,2} = 0$ | | | $W_{3,2} = 1$ | | |
| $W_{3,3} = 1 - 46/35z^{-1} + 17/35z^{-2}$ | | | $W_{3,3} = 1$ | | | $W_{3,3} = -2/3z^{-1}$ | | | $W_{3,3} = 1$ | | |
| $V_3 = 1 + 73/36z^{-1} + 73/36z^{-2}$ | | | $V_3 = 1 + 11/13z^{-1}$ | | | $V_3 = 1 + 7/2z^{-1}$ | | | $V_3 = 1$ | | |

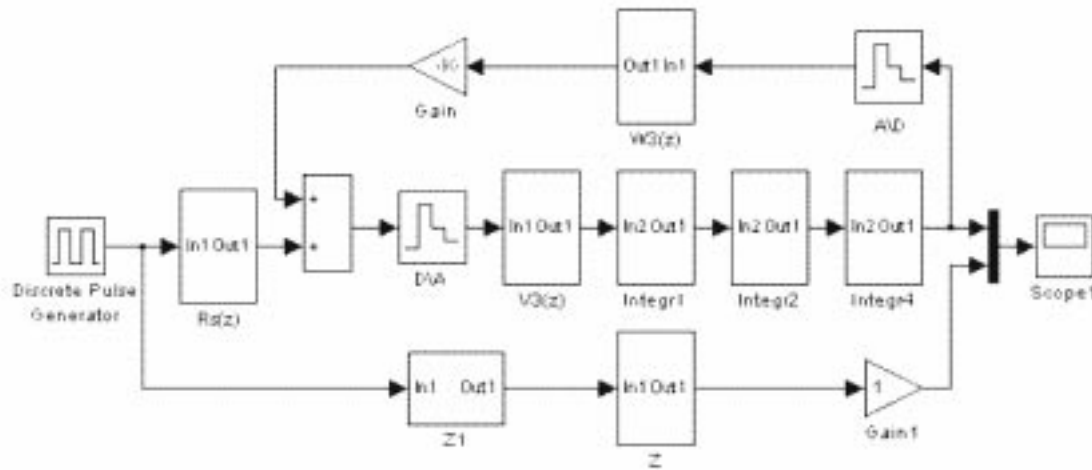


Рисунок 7

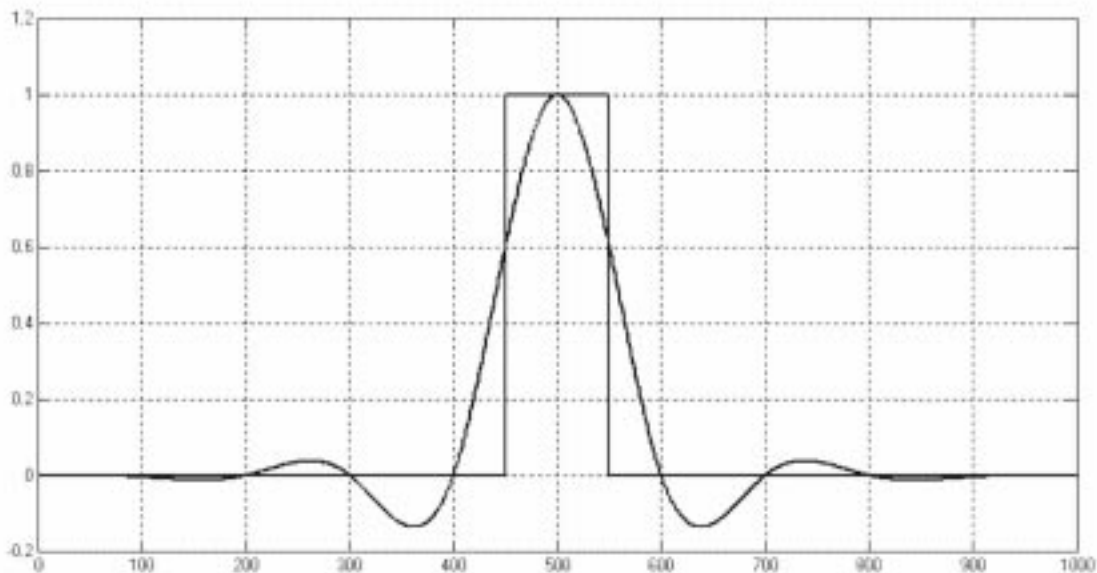


Рисунок 8

Для анализа погрешностей сплайн-интерполирующего АЦФ можно использовать метод имитационного моделирования в среде Simulink 4.0 пакета Matlab 6.0. Это позволяет получить качественные и количественные характеристики интерполирующего АЦФ, необходимые для решения задач проектирования и оценки методических и инструментальных погрешностей процедур восстановления цифровых сигналов. В качестве примера на рисунке 7 приведена имитационная модель сплайн-интерполирующего АЦФ на базе трех интеграторов ($n = 3$) с цифровым корректирующим фильтром ($s = 3$).

На рисунке 8 приведена ИХ, которая совпадает с ИХ, полученной аналитически.

Таким образом, использование АЦФ замкнутой структуры позволяют реализовывать в аналого-цифровой форме алгоритмы восстановления цифровых сигналов, в основе которых лежат методы сплайн-интерполяции.

ПЕРЕЧЕНЬ ССЫЛОК

1. Альберг Дж., Нильсон Э., Уолш Дж. Теория сплайнов и ее приложения. - М.: Мир, 1972. - 292с.
2. Михотин В.Д., Чувькин Б.В. Использование сплайнов для восстановления дискретизированных сигналов // Измерение, контроль, автоматизация. - 1982. - №3 (43). - С.17-24.
3. Чувькин Б.В. Фinitные функции. Теория и инженерные приложения // Монография: Под ред. Э.К. Шахова. - Пенза: Изд-во Пенз. гос. ун-та, 1999. - 100 с.

СИНТЕЗ АВТОМАТА МИЛИ С ПРЕОБРАЗОВАНИЕМ МИКРООПЕРАЦИЙ В КОДЫ СОСТОЯНИЙ

А.А.Баркалов, Эль-Джейби Ахмад Камаль, А.А.Красичков

Пропонується метод оптимізації логічної схеми мікропрограмного автомата, що базується на перетворенні кодів мікрооперацій на коди стану. Визначені умови застосування методу, та приведений приклад його використання.

Предлагается метод оптимизации логической схемы микропрограмного автомата, основанный на преобразовании кодов микроопераций в коды состояния. Определены условия применения метода, и приведен пример его использования.

The method of optimization of the logic circuit of the microprogram automaton is proposed. The method is based on transformation of sets of microoperations to codes of conditions. The method and the example of synthesis of the automaton with proposed transformation is given. The area of effective application of the given method is determined.

ВВЕДЕНИЕ

Одной из актуальных проблем, стоящих перед специалистами в области синтеза цифровых устройств, является проблема оптимизации аппаратных затрат в схемах устройств [1]. Эта же проблема является одной из основных при синтезе схем устройств управления (УУ) цифровых систем на программируемых логических устройствах (ПЛУ) [2, 3]. Практическая ценность решения этой задачи заключается в получении УУ с меньшей стоимостью, что повышает их конкурентоспособность. Пути решения этой проблемы при реализации УУ в виде микропрограммного автомата широко освещены в литературе [3, 4]. Один из путей решения этой проблемы - увеличение числа уровней схемы [4], что приводит к появлению дополнительных переменных. При этом актуальным является разработка методов, приводящих к схемам с минимальным числом дополнительных переменных.

В настоящей работе предлагается метод многоуровневой реализации схемы автомата с минимальным числом дополнительных переменных. В основе метода находится преобразование микроопераций в коды состояний автомата.

1 ПОСТАНОВКА ЗАДАЧИ

Пусть автомат Мили задан прямой структурной таблицей (ПСТ) со столбцами [5]:

a_m - исходное состояние автомата, $a_m \in A$, где $A = \{a_1, \dots, a_m\}$ - множество состояний автомата;

$K(a_m)$ код состояния a_m разрядности $R = \lceil \log_2 M \rceil$, для кодирования используются внутренние переменные $T_r \in T = \{T_1, \dots, T_R\}$;

a_s , $K(a_s)$ - состояния перехода и его код;

X_k - входной сигнал, определяющий переход $\langle a_m, a_s \rangle$ и равный конъюнкции некоторых элементов (или их отрицаний) множества логических условий $X = \{x_1, \dots, x_L\}$;

Y_k - набор микроопераций, формируемых на переходе $\langle a_m, a_s \rangle$, $Y_k \subseteq Y$, где $Y = \{y_1, \dots, y_N\}$ - множество микроопераций;

Φ_k - набор функций возбуждения памяти автомата, принимающих единичные значения для переключения памяти из $K(a_m)$ в $K(a_s)$, $\Phi_k \subseteq \Phi$, где $\Phi = \{\phi_1, \dots, \phi_R\}$ - множество функций возбуждения; $h = \overline{1, H}$ - номер перехода. Эта таблица является основной для синтеза одноуровневой схемы автомата Мили (рис. 1), в которой комбинационная схема КС строится по системам функций

$$\phi_r = \bigvee_{h=1}^H \text{Crh} \text{Am}^h X_h \quad (r = 1, \dots, R), \quad (1)$$

$$y_n = \bigvee_{h=1}^H \text{Cnh} \text{Am}^h X_h \quad (n = 1, \dots, N). \quad (2)$$

Здесь Crh (Cnh) - булева переменная, равная единице, если и только если в h -й строке ПСТ записана переменная ϕ_r (y_n); Am^h - конъюнкция внутренних переменных $T_r \in T$, соответствующая коду исходного состояния $K(a_m)$ из h -й строки ПСТ.

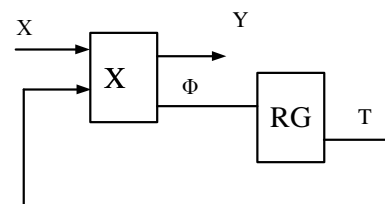


Рисунок 1 - Одноуровневая схема микропрограммного автомата Мили

Основной недостаток одноуровневой реализации - значительное дублирование входных и выходных переменных на разных микросхемах схемы автомата, что приводит к увеличению аппаратных затрат [4]. При переходе к многоуровневым схемам [3,4] появляются дополнительные переменные, реализация которых требует дополнительных ресурсов ПЛУ.

В настоящей работе предлагается метод оптимизации автомата Мили, приводящий к двухуровневой схеме с минимальным числом дополнительных переменных, в предельном случае - с нулевым.

Пусть состояние перехода a_s из h -й строки ПСТ однозначно определяется парой

$$b_h = \langle a_m, Y_h \rangle, \quad h = \overline{1, H}. \quad (3)$$

Под однозначностью понимается тот факт, что в ПСТ не существует пар с одинаковыми $a_m \in A$ и $Y_h \subseteq Y$ различными состояниями перехода $a_s \in A$. В этом случае для реализации логической схемы автомата Мили предлагается структура (рис. 2), названная ТУ-автоматом.

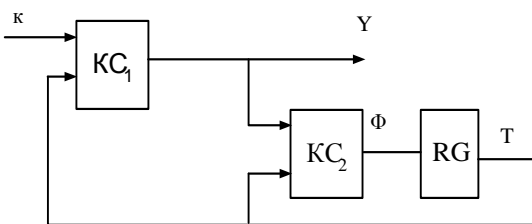


Рисунок 2 - Структурная схема ТУ-автомата Мили

В ТУ-автомате схема KC_1 реализует систему (2), а схема KC_2 основана на (3) и реализует систему функций

$$\varphi_r = \bigvee_{h=1}^H C_r h A_m^h Z_h \quad (r = 1, \dots, R), \quad (4)$$

где Z_h - конъюнкция микроопераций избы h -й строки ПСТ.

Будем рассматривать Y_h как N -компонентный вектор $Y_h = \langle l_{h1}, \dots, l_{hN} \rangle$ где $l_{hN} \in \{0, 1\}$ - значение микрооперации $y_n \in Y$ в наборе $Y_h \subseteq Y$. Тогда

$$Z_h = \bigwedge_{n=1}^N y_n^{l_{hn}} \quad (h = 1, \dots, H), \quad (5)$$

где $y_n^1 = y_n, y_n^0 = 1$.

В общем случае пара (3) может соответствовать нескольким состояниям перехода. В этом случае необходимо ввести специальные идентификаторы $I_k \in I = \{I_1, \dots, I_k\}$, позволяющие однозначно идентифицировать состояние перехода совместно с парой (3). Закодируем идентификаторы $I_k \in I$ двоичными кодами $K(I_k)$ разрядности $Q = \lceil \log_2 K \rceil$, используя для кодирования переменные $V_q \in V = \{V_1, \dots, V_Q\}$. В этом случае автомат Мили будет иметь следующую структуру (рис. 3), называемую ТУИ-автоматом.

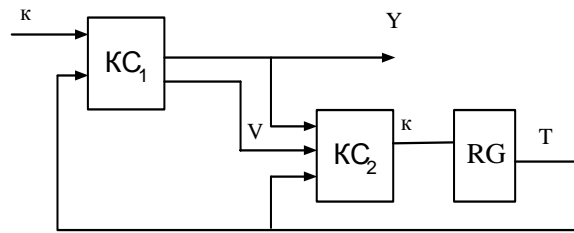


Рисунок 3 - Структурная схема ТУИ-автомата Мили

Здесь схема KC_1 реализует систему (2) и систему функций $V_q \in V$

$$V_q = \bigvee_{h=1}^H C_q h A_m^h X_h \quad (q = 1, \dots, Q). \quad (6)$$

Здесь $C_q h$ - булева переменная, равная единице, если и только если в h -й строке ПСТ записана переменная $V_q = 1$. Схема KC_2 реализует систему (4), трансформируемую в

$$\varphi_r = \bigvee_{h=1}^H C_r h A_m^h Z_h B_h \quad (r = 1, \dots, R), \quad (7)$$

где B_h - конъюнкция переменных $V_q \in V$, соответствующая коду идентификатора $K(I_k)$ из h -й строки ПСТ.

2 МЕТОД РЕШЕНИЯ ЗАДАЧИ

Для синтеза ТУИ-автомата предлагается следующая методика, иллюстрируемая на примере автомата Мили S_1 (табл. 1).

Таблица 1 - Прямая структурная таблица автомата Мили S_1

| a_m | $K(a_m)$ | a_s | $K(a_s)$ | X_h | Y_h | Φ_h | h |
|-------|----------|-------|----------|-----------------------|-----------|-----------|-----|
| a_1 | 00 | a_2 | 01 | $x_1 x_2$ | $y_1 y_2$ | D_2 | 1 |
| | | a_3 | 10 | $x_1 \bar{x}_2$ | y_3 | D_1 | 2 |
| | | a_4 | 11 | $\bar{x}_1 x_3$ | $y_1 y_2$ | $D_1 D_2$ | 3 |
| | | a_1 | 00 | $\bar{x}_1 \bar{x}_3$ | $y_1 y_3$ | - | 4 |
| a_2 | 01 | a_3 | 10 | x_4 | y_3 | D_1 | 5 |
| | | a_4 | 11 | $\bar{x}_4 x_5$ | $y_1 y_2$ | $D_1 D_2$ | 6 |
| | | a_2 | 01 | $\bar{x}_4 \bar{x}_5$ | y_3 | D_2 | 7 |
| a_3 | 10 | a_3 | 10 | x_6 | y_1 | D_1 | 8 |
| | | a_4 | 11 | \bar{x}_6 | $y_1 y_2$ | $D_1 D_2$ | 9 |
| a_4 | 11 | a_1 | 00 | 1 | - | - | 10 |

1. Проанализировать подтаблицы $\pi_m(m = \overline{1, M})$ исходной ПСТ. Если в подтаблице π_m имеется I_k^h пар вида $\langle a_m, Z_h \rangle$ с одинаковыми Z_h , определяющие различные состояния перехода $a_m \in A$, то поставить им в соответствие I_k^h идентификаторов.

2. Сформировать множество идентификаторов $I = \{I_1, \dots, I_k\}$, где $K = \max(I_1^1, \dots, I_M^H)$. В нашем случае пара $\langle a_1, y_1y_2 \rangle$ определяет состояние перехода a_2 и a_4 , пара $\langle a_2, y_3 \rangle$ - состояния a_3 и a_2 . Следовательно, $I = \{I_1, I_2\}$.

3. Закодировать идентификатор $I_h \in I$ двоичными кодами $K(I_h)$ разрядности Q . Для автомата S_1 $K = 2$, $Q = 1$, $V = \{V_1\}$. Пусть $K(I_h) = 0$, $K(I_h) = 1$.

4. Построить таблицу микроопераций, удалив из ПСТ столбцы a_m , $K(a_s)$, Φ_h и введя столбцы I_h , V_h . Столбец I_h содержит идентификатор состояния a_s для пары b_h ($h = 1, H$), столбец V_h - переменные $v_q \in V$, равные единице в коде $K(I_h)$. Если для строки h идентификатор отсутствует, то в столбце V_h записывается "*".

Для автомата S_1 таблица микроопераций приведена в табл. 2. Отметим, что эта таблица имеет $H = 9$ строк, так как в 10-й строке ПСТ (Табл. 1) отсутствуют микрооперации и функции возбуждения памяти автомата.

Таблица 2 - Таблица микроопераций ТУИ-автомата S_1

| a_m | $K(a_m)$ | X_h | Y_h | I_h | V_h | h |
|-------|----------|----------------------|----------|-------|-------|-----|
| a_1 | 00 | x_1x_2 | y_1y_2 | I_1 | - | 1 |
| | | $\bar{x}_1\bar{x}_2$ | y_3 | - | * | 2 |
| | | \bar{x}_1x_3 | y_1y_2 | I_2 | v_1 | 3 |
| | | $\bar{x}_1\bar{x}_3$ | y_1y_3 | - | * | 4 |
| a_2 | 01 | x_4 | y_3 | I_1 | - | 5 |
| | | \bar{x}_4x_5 | y_1y_2 | - | * | 6 |
| | | $\bar{x}_4\bar{x}_5$ | y_3 | I_2 | v_1 | 7 |
| a_3 | 10 | x_6 | y_1 | - | * | 8 |
| | | \bar{x}_6 | y_1y_2 | - | * | 9 |

5. Построить по таблице микроопераций системы функций (2) и (6). Для автомата S_1 из табл. 2 имеем, например, $y_3 = A_1X_2 \vee A_2X_5 \vee A_2X_7 = \bar{T}_1\bar{T}_2\bar{x}_1\bar{x}_2 \vee \dots \vee \bar{T}_1\bar{T}_2\bar{x}_4\bar{x}_5$; $V = A_1X_3 \vee A_2X_7 = \bar{T}_1\bar{T}_2\bar{x}_1\bar{x}_2 \vee \dots \vee \bar{T}_1\bar{T}_2\bar{x}_4\bar{x}_5$.

6. Построить таблицу функций возбуждения памяти ТУИ-автомата, удалив из исходной ПСТ столбцы $a_s - X_h$ и введя столбец V_h из таблицы микроопераций. Для ТУИ-автомата S_1 таблица функций возбуждения памяти (Табл. 3) имеет $H = 8$ строк, так как для строк 4 и 10 исходной ПСТ $a_s = a_1$.

Таблица 3 - Таблица функций возбуждения памяти ТУИ-автомата S_1

| a_m | $K(a_m)$ | Y_h | V_h | Φ_h | h |
|-------|----------|----------|-------|----------|-----|
| a_1 | 00 | y_1y_2 | - | D_2 | 1 |
| | | y_3 | * | D_1 | 2 |
| | | y_1y_2 | v_1 | D_1D_2 | 3 |
| a_2 | 01 | y_3 | - | D_1 | 4 |
| | | y_1y_2 | * | D_1D_2 | 5 |
| | | y_3 | v_1 | D_2 | 6 |
| a_3 | 10 | y_1 | * | D_1 | 7 |
| | | y_1y_2 | * | D_1D_2 | 8 |

7. Построить по таблице функций возбуждения памяти систему (7). Для нашего примера имеем, например, $D_1 = \bar{T}_1\bar{T}_2y_3 \vee \bar{T}_1\bar{T}_2y_1y_2V_1 \vee \bar{T}_1T_2y_3\bar{V}_1 \vee \bar{T}_1T_2y_1y_2 \vee T_1\bar{T}_2y_1 \vee T_1\bar{T}_2y_1y_2$. Отметим, что если в столбце V_h записана "*", то для соответствующей строки $b_h = 1$.

8. Построить схему ТУИ-автомата по системам функций (2), (6) - (7). Функциональная схема ТУИ-автомата S_1 приведена на рис. 4. Здесь схемы КС1 и КС2 реализованы на программируемых логических матрицах, сигнал "Пуск" используется для установки регистра RG в начальное состояние, сигнал "Clock" является импульсом синхронизации.

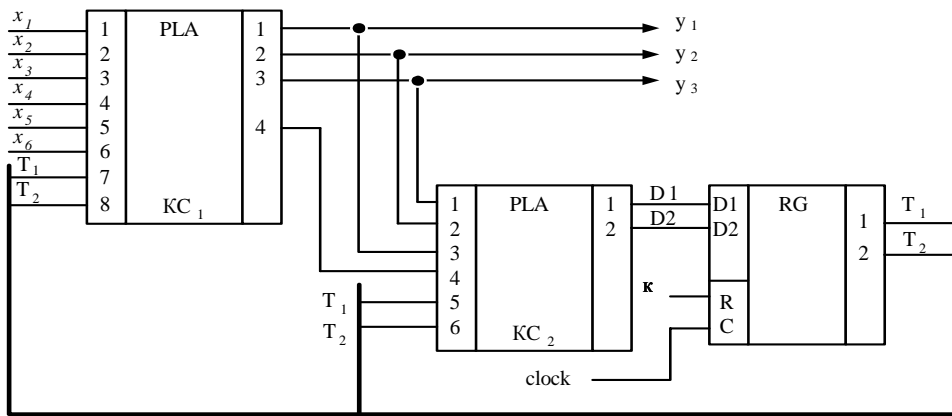


Рисунок 4 - Функциональная схема ТУІ-автомата S_1

ЗАКЛЮЧЕНИЕ

Пусть S, t - соответственно число входов и выходов ПЛУ, используемых для синтеза схем KC_1 и KC_2 . Как показали исследования, применение предложенной методики целесообразно при выполнении условий

$$N + R + Q \leq S, \tag{8}$$

$$N + Q \leq t, \tag{9}$$

$$N + R > t, \tag{10}$$

$$L + R > S. \tag{11}$$

Если (8) не выполняется, то применение предложенного метода приводит к многоуровневой схеме KC_2 , что равно увеличивает аппаратные затраты. При выполнении условий (9) и (10) в одноуровневой схеме требуется расширение ПЛМ по выходам, что не требуется в ТУІ-автомате. Если (9) не выполняется, то выигрыш от предложенного метода уменьшается и необходимо синтезировать оба автомата для выбора лучшей структуры. Если (10) не выполняется, то одноуровневая схема обладает лучшими характеристиками. Если (11) не выполняется, то аппара-

турные затраты в одноуровневой схеме будут не более, чем в ТУІ-автомате. Если условия (5)-(11) выполняются, то аппаратные затраты в ТУІ-автомате на 25-32% меньше, чем в одноуровневой схеме. Выигрыш растёт по мере увеличения числа логических условий исходной граф-схемы алгоритма.

Высокая результативность предложенного метода свидетельствует о целесообразности продолжения исследований в этом направлении. При этом исследованию подлежит возможность применения предложенной методики при синтезе микропрограммных автоматов Мура и композиционных микропрограммных устройств управления.

ПЕРЕЧЕНЬ ССЫЛОК

1. Грушвицкий Р.И., Мурсаев А.Х., Угрюмов Е.П. Проектирование систем на микросхемах программируемой логики. - СПб.: БХВ-Петербург, 2002. - 608с.
2. Баркалов А.А Палагин А.В. Синтез микропрограммных устройств управления. - Киев: ИК НАН Украины, 1997. - 156 с.
3. Соловьёв В.В. Проектирование цифровых систем на основе программируемых логических интегральных схем. - М.: Горячая линия - Телеком, 2001. - 636 с.
4. Баркалов А.А. Синтез устройств управления на программируемых логических устройствах. - Донецк: ДонНТУ, 2002. - 262 с.
5. Баранов С.И. Синтез микропрограммных автоматов. - Л.: Энергия, 1979. - 232 с.

УДК 004.681.32

КОМПЬЮТЕРНАЯ ПСИХОДИАГНОСТИЧЕСКАЯ СИСТЕМА ДЛЯ ОЦЕНКИ ПРОФПРИГОДНОСТИ ПЕРСОНАЛА РАЗДЕЛИТЕЛЬНЫХ ПРОИЗВОДСТВ

О.Г.Берестнева, В.А.Власов, А.А.Орлов

Предлагается система компьютерных психодиагностических тестов для оценки профпригодности персонала разделительных производств. Разработаны гибкая система оценки профессиональной пригодности, а также оригинальная методика тестирования и удобная для пользователя компьютерная программа оценки индивидуальных особенностей и

правильности выбора будущей профессии школьников и абитуриентов.

The computer system of psychodiagnostics tests for an evaluation of occupational aptitude of staff of separating productions is offered. The flexible system of an evaluation of professional

suitability is developed. The original technique of testing and computer program, for evaluation of individual features and regularities of choice of the future trade of the schoolboys and entrants is developed.

ВВЕДЕНИЕ

Современные производства по разделению изотопов представляют собой сложные технические и технологические комплексы, оснащенные дорогостоящим, точным, наукоемким оборудованием. Эти производства характеризуются высокой энергонапряженностью всех элементов технологической схемы, режимом работы оборудования на номинальных параметрах, взаимодействием специалистов различного профессионального уровня, высокой централизацией управления, повышенными требованиями к надежности оборудования и ядерной безопасности производства. Наличие таких особенностей требует проведения проверки профессиональной пригодности принимаемых на работу специалистов, периодической оценки подготовки работающего технологического персонала разделительных производств, и в случае необходимости, прохождения курсов переподготовки.

Разработанная авторами компьютерная система тестирования предназначена для оценки профессиональной пригодности различных категорий технологического персонала разделительных производств (инженеров-исследователей, инженеров-технологов, аппаратчиков и т.д.).

1 ПОСТАНОВКА ЗАДАЧИ

Профессиональная пригодность - это вероятностная характеристика, отражающая возможности человека в овладении какой-либо профессиональной деятельностью. Она может оцениваться по нескольким критериям:

- медицинским показателям (в том числе по показателям физической подготовленности);
- по данным образовательного ценза или конкурсным экзаменам;
- с помощью психологического обследования (психологический отбор);
- с учетом достигнутого уровня профессиональной адаптации и др.

В нашем случае все оцениваемые качества были объединены в три основных блока:

- личностные (психологические);
- психофизиологические;
- навыки и умения, необходимые для успешной профессиональной деятельности.

Как известно, задача профотбора решается в несколько этапов.

1. Психологическое изучение профессии с целью выявления требований к человеку. При этом должна быть раскрыта внутренняя структура деятельности и дан не просто перечень психических и психомоторных процессов, находящихся под наибольшей нагрузкой, а целостная картина их взаимосвязи и воплощения в профессиональной эффективности. Такое изучение (профессиография) завершается подробной характеристикой профессии - профессиограммой и психограммой.

2. Выбор психодиагностических методов исследования (тестов), в наибольшей мере характеризующих те психические процессы и профессиональные действия, в отношении которых надлежит оценивать профессиональную пригодность.

3. Психодиагностика - психологическое изучение обследуемых, оценка их общего развития, направленности на овладение профессией, степени развития у них качеств, наиболее важных для успехов в овладении профессией, и оценка способности выполнения ими элементов целостной деятельности в различных условиях.

4. Психологический прогноз успешности профессиональной деятельности на основе сопоставления сведений о требованиях к профессии и полученных психодиагностических данных, с акцентом на оценку личностных характеристик. Заключение о возможности целенаправленного совершенствования и компенсации профессионально значимых качеств.

Исходным пунктом для исследования профессиональной пригодности является, построение профессиограммы и, в частности, ее психологической части - психограммы. Психограмма определяет "центр тяжести" профессии в отношении нагрузки на психофизические функции и требования к личностным чертам работника-специалиста в данной профессии. Перечень таких требований для специальностей разделительных производств был составлен авторами на основе анализа литературных источников [1-3] с учетом мнения экспертов [4]. В качестве экспертов выступали ведущие специалисты разделительных производств и профессорско-преподавательский состав физико-технического факультета Томского политехнического университета, осуществляющий подготовку специалистов данного профиля.

При выборе методик для компьютерного тестирования авторы исходили из того, что, с одной стороны, набор тестов должен быть достаточно полным, позволяющим оценить не только каждое качество в отдельности, но и индивидуальный стиль деятельности в целом, а с другой стороны, выбранные методики должны представлять собой экспресс-методы психодиагностики.

2 ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ СИСТЕМЫ

В разработанной авторами компьютерной системе психодиагностического профессионального тестирования реализована схема, приведенная на рис. 1.

Выбор опросника Кеттелла для решения нашей задачи обусловлен тем, что все определяемые в нем факторы личности входят в число профессионально значимых качеств персонала разделительных производств. В компьютерную систему профотбора был включен адаптированный вариант опросника Кеттелла, разработанный сотрудниками Института психологии академии наук [5].

С целью повышения качества оценки профессионально значимых психологических качеств в компьютерную систему был включен также тест для определения социотипа личности.

Установлено, что существует тесная взаимосвязь между социотипом человека и его профессиональной успеш-

ностью в той или иной сфере деятельности [6]. Нами разработаны диагностические таблицы, содержащие данные о том, при каких социотипах можно прогнозировать успешную профессиональную деятельность по интересующим нас категориям персонала, а при каких нет [4].

Многие из профессионально значимых качеств являются одинаковыми для разных профессий. Поэтому при решении задачи оценки профессиональной пригодности необходимо определять также диагностические коэффициенты ("веса") данных качеств (либо пороговые значения исследуемых характеристик для каждой профессии [7]).

Эта задача может быть решена двумя путями:

- величина диагностических коэффициентов (или пороговых значений) задается эмпирически, если имеется большой опыт профотбора на данные специальности или на основе анализа литературных источников;
- диагностические коэффициенты (или пороговые значения) определяются на основе результатов статистических исследований, заключающихся в сопоставлении тестовых показателей с показателями внешних критериев профессиональной эффективности.

На данном этапе работы был реализован первый подход к решению данной задачи. На основе анализа литературных источников [1] и результатов экспертного оценивания нами пороговые значения личностных качеств (определяемых по тесту Кеттелла) для специальностей, включенных в компьютерную систему тестирования. Пороговые значения взяты из литературного источника и скорректированы с привлечением экспертов [4].

Для повышения надежности результатов тестирования в состав программного обеспечения системы включена методика, построенная на основе метода независимых характеристик К.К. Платонова. Метод независимых характеристик заключается в том, что в соответствии с разработанной программой опрашиваются коллеги по работе, непосредственное начальство, другие работники, в силу тех или иных обстоятельств хорошо знающие изучаемого специалиста, - всего 5-6 человек. Предложена шкала, по которой оценивается та или иная черта личности [8]: 5 - названная в карте черта личности развита очень хорошо, ярко выражена и проявляется часто и в различных видах деятельности; 4 - заметно выражена, но проявляется не постоянно, хотя противоположная ей черта проявляется крайне редко; 3 - как и противоположная ей черта личности, выражена не резко, и в проявлениях они уравновешивают одна другую, хотя обе проявляются нечасто; 2 - заметно более выражена, и чаще проявляется противоположная названной черта личности; 1 - противоположная названной черта личности проявляется часто и в различных видах деятельности. Оценка каждой черты личности должна опираться на знание "жизненных показателей", т.е. типичных случаев из жизни оцениваемого человека и его деятельности (действий, поступков), в которых проявляется оцениваемое качество.

В состав нашей компьютерной системы, включен блок, реализующий методику независимых характеристик и представляющий собой модификацию методик Платонова [7], Фукуямы [3], Борейши [2]. Уровень владения тем или иным качеством оценивается следующим образом: 5 - очень высокий; 4 - высокий; 3 - средний; 2 - низкий; 1 -

очень низкий; 0 - данное качество отсутствует.

На первом этапе тестирования испытуемый сам оценивает себя по всем профессионально значимым качествам. Кроме того, для каждого из этих качеств он указывает, необходимо ли оно для успешной профессиональной деятельности по выбранной специальности. Эти данные необходимы для определения того, насколько полными являются представления испытуемого о требованиях, предъявляемых к данной специальности. Результатом данного теста является средний балл по всем качествам, а также средний балл отдельно по психологическим (личностным) и психофизиологическим качествам, а также по группе качеств, представляющих собой профессиональные навыки и умения. В выходном документе указывается на развитие каких качеств испытуемому следует обратить внимание (это качества, по которым испытуемый оценил себя ниже трех баллов).

На следующем этапе тестирования испытуемого по этим же качествам оценивают эксперты (работники хорошо знающие данного специалиста). Оптимальное количество экспертов 5-6 человек, минимальное 2-3 человека. Затем подсчитывается усредненная экспертная оценка по каждому качеству и выдается заключение о степени соответствия испытуемого данной специальности (как по всем качествам в целом, так и отдельно по трем группам качеств: психологическим, психофизиологическим и профессиональным).

Если средний балл 5, то выдается заключение: "очень высокая степень соответствия специальности", если средний балл от 4 до 5 - "высокая степень", если от 3 до 4 - "средняя степень", если от 2 до 3 - "низкая степень", если меньше 2 - "очень низкая".

В качестве приемов исследования и оценки психофизиологических функций используются методики субъективной и объективной оценки личности [7-11]. Первые основаны на получении сведений о тех или иных свойствах человека, проявляющихся в различных в жизненных или экспериментальных ситуациях, либо от самого индивида (самонаблюдение и самоанализ), либо непосредственно экспериментатором (анализ внешних проявлений этих свойств у конкретного человека или данных, полученных от него). В нашем случае это методика независимых характеристик, описанная выше.

Объективная оценка личности проводится на основании анализа результатов выполнения определенных (как правило, стандартных) заданий или функциональных проб, направленных на изучение основных познавательных психических процессов и состояний различных психофизиологических функций индивида.

Наряду с субъективными были использованы также и объективные оценки, характеризующие психофизиологические функции. Из имеющегося сравнительно большого количества методик были отобраны такие, которые, по мнению большинства специалистов, считаются наиболее адекватными [7,8,10] - это время реакции на звук, время реакции на свет, тепшинг-тест, тест быстроты и точности реакции на движущийся объект.

Время простой сенсомоторной реакции человека на звук и на свет характеризует возбудимость анализаторов, временные показатели нервных процессов, являющиеся

основными в управляющих действиях человека [10]. Выбор теста "время реакции на световые раздражители" для оценки функционального состояния человека в процессе адаптации-деадаптации обусловлен его информативностью как показателя функции "целостного мозга".

Кроме того, для оценки общих свойств нервной системы (силы процессов возбуждения и торможения; их уравновешенности и подвижности) в систему включен тест на определение типа темперамента. Нами разработаны таблицы "совместимости" интересующих нас специальностей и типов темперамента, полученные на основе анализа литературных источников [10] и мнения экспертов [4].

Заключение по блоку качеств "Навыки и умения, необходимые для успешной профессиональной деятельности" формируется на основе результатов тестирования и экспертной оценки по методике независимых характеристик [4].

На рисунке 1 приведена структурная схема программного обеспечения разработанной компьютерной системы. Как видно из рисунка 1, в системе наряду с формированием общего заключения об уровне профпригодности предусмотрена выдача заключения о степени соответствия той или иной специальности по каждому отдельному блоку.

3 ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ СИСТЕМЫ

Основные данные программы разделены на три больших раздела:

1. Данные о пользователе.
2. Данные о тесте.
3. Данные о специальности.

Каждый из крупных блоков данных представляет

иерархическую структуру, состоящую из базовых типов данных.

Информация о пользователе делится на две части:

1. Общие данные:

1. Фамилия, Имя и Отчество тестируемого.
2. Дата рождения.
3. Образование.
4. Название учебного заведения.
5. Название специальности.
6. Должность и дата назначения на должность.
7. Квалификация.
8. Дата приема на работу.
9. Дата поступления и окончания ВУЗа (для абитуриентов и студентов).
10. Почтовый адрес.
11. Почтовый адрес родителей (для абитуриентов и студентов).
12. Примечания.

2. Результаты тестирования.

Результаты тестирования объединяются в группы. Каждая группа состоит из даты проведения тестов, результатов тестов и результатов экспертной оценки по методике независимых характеристик.

Данные о тестах также представлены в виде иерархической структуры. Каждый тест состоит из:

1. Названия.
2. Уникального номера теста.
3. Массива категорий тестирования.

Категории тестирования в свою очередь состоят из:

4. Названия категории тестирования.
5. Номера категории (выставляется автоматически).
6. Массива вопросов, относящихся к данной категории.

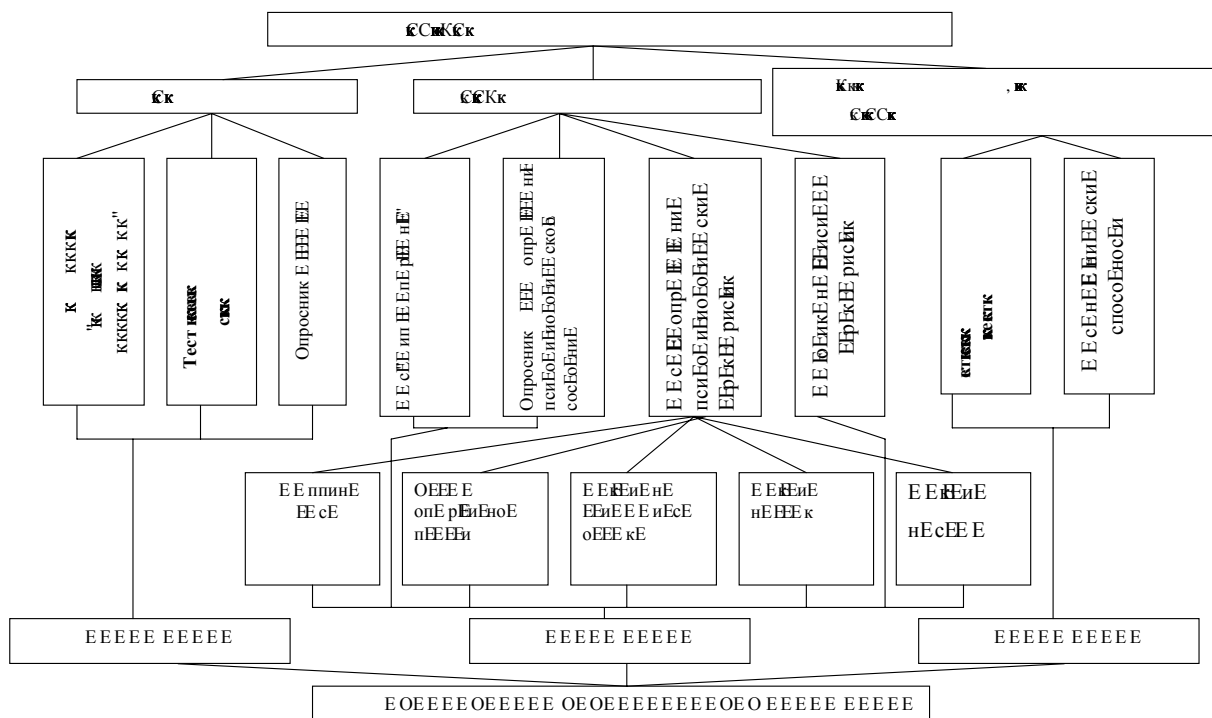


Рисунок 1 - Схема программного обеспечения системы профтестирования

Каждый вопрос состоит из следующих данных:

7. Номера вопроса, уникального в пределах категории тестирования.

8. Текста вопроса.

9. Массива вариантов ответов в виде:

а) текст варианта ответа;

б) количество баллов, получаемое при выборе данного варианта ответа.

10. Результата ответа на текущий вопрос (это поле используется для возврата данных в программу после тестирования).

Данные о специальности:

Название специальности.

Номер специальности (уникальное число).

Массив пороговых значений для теста Кеттелла.

Массив пороговых значений для методики независимых характеристик.

Каждая из описанных структур данных была сведена в классы, содержащие данные и методы их обработки.

Массивы данных были реализованы, используя классы массивов-шаблонов (templates). Данные массивы позволяют программисту легко оперировать данными, которые в них содержатся и обеспечивают динамическое изменение размеров массивов.

В качестве хранилища данных для баз перечисленных тестов, специальностей и пользователей применяются .ini-файлы, работа с которыми закодирована на уровне системных функций Windows.

Выбор данного типа файлов обусловлен независимостью их реализации от используемых драйверов для доступа к данным (ODBC), что позволяет использовать программу на любой машине с Microsoft Windows 9X/Windows NT™.

Для реализации программного обеспечения был выбран язык C++ [4], который позволяет конструировать абстрактные типы данных и реализовывать на их основе работу с требуемыми данными.

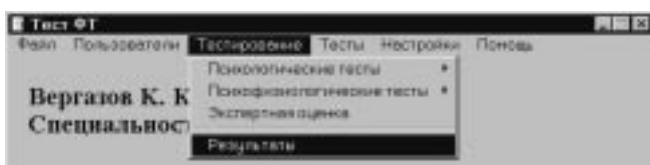


Рисунок 2 - Главное окно и меню программы

В основе языка C++ лежит концепция классов, которые содержат в себе как данные, так и процедуры работы с этими данными. Внешние по отношению к классам процедуры могут работать с данными классов только через четко определенные процедуры-интерфейсы. Это позволяет реализовать принцип "черного ящика", когда внешние части не знают, каким образом производится обработка данных классов, что позволяет легко менять разные части программы, не внося изменений в другие части.

Вторым важным достоинством C++ является возмож-

ность использования наследования классов от уже определенных классов. Это позволяет реализовать разную обработку данных в различающихся классах. А механизм виртуальных функций позволяет единообразно производить вызовы функций разных классов, унаследованных от одного источника. Этот механизм использовался при реализации обработки результатов тестирования. Структура разработанного программного обеспечения и направления потоков данных в программе показаны на рисунке 2.

Программа "Тест ФТ" разработана с использованием интегрированной среды разработки Sybase Power++ 2. Эта программа предназначена для проведения тестов различной сложности и размерности, с последующим сохранением результатов тестирования, а также для создания новых тестов.

Система состоит из двух основных блоков:

1. Тестирование.

2. Создание и модификация тестов.

Программа соответствует современным стандартам, основными критериями которых являются:

- Удобный пользовательский интерфейс, который представляет собой стандартные Windows-компоненты (пример интерфейса представлен на рис. 2).

- Модифицируемость блоков программы, то есть возможность изменять отдельные модули без изменения всей программы.

- Контекстно-зависимая справочная система, позволяющая получить справку о том или ином программном компоненте, независимо от работы основной программы.

Разработанный программный комплекс "Тест ФТ" представляет собой стандартное Windows-приложение, отвечающее всем современным требованиям программного продукта для операционных систем Windows 95/98/NT.

4 РЕЗУЛЬТАТЫ

Разработанная компьютерная система позволяет проводить оценку уровня профпригодности специалистов разделительных производств. При разработке программного обеспечения системы нами был предварительно сформирован набор тестов для выявления личностных, и психофизиологических качеств, характеризующих профессиональную пригодность человека для работы на разделительном производстве, а также его навыки и умения необходимые для успешной профессиональной деятельности. Определен набор качеств, которыми должны обладать различные категории персонала разделительного производства, и разработана гибкая система оценки профессиональной пригодности. Предусмотрен вариант использования компьютерной системы для решения задач профориентации.

В настоящее время система прошла апробацию в ПО "Электрохимический завод" г. Зеленогорск. В Томском политехническом университете с использованием данной системы ежегодно проводится профориентационное тестирование абитуриентов физико-технического факультета и оценка уровня профпригодности выпускников специальности "Физика кинетических явлений".

ПЕРЕЧЕНЬ ССЫЛОК

1. Комплексная социально - психологическая методика изучения личности инженера. // Л.: Изд-во ЛГУ - 1991 - 178с.
2. Борейша В. Б. ТомПИКС - ловушка для подходящих профессий. - Томск, 1994. - 32 с.
3. Фукуяма С. Теоретические основы профессиональной ориентации. - М.: Изд-во МГУ, 1989.- 108 с.
4. Власов В. А., Орлов А. А., Тихонов Г. С., Отт А. Я., Берестнева О. Г., Еремина В. В. Компьютерная система оценки профессиональной пригодности персонала для разделительных производств // Научный отчет. Рег.№02200005350, г.Томск, ТПУ, 2000. - 92 с.
5. Русалов В. М., Гусева О. В. Сокращенный вариант личностного опросника Кэттелла (8 PF) // Психологический журнал, Том 11, №1, 1990. - С.34-48
6. Гуленко В. В., Тыщенко В. П. Соционика - межвозрастной педагогике. - Новосибирск: Изд-во НГУ, 1998. - 269 с.
7. Маришук В. Л. Критерии профессиональной пригодности в отборе // Методология исследований по инженерной психологии и психологии труда. - Л., 1974. - Ч.1.
8. Дмитриева М. А., Крылов А. А., Нафтульев А. И. Психология труда и инженерная психология. - Л.: Изд-во Ленингр.ун-та, 1979.-224с.
9. Ольхов О.Г. Опросник психофизиологического состояния рабочих горячих цехов металлургической промышленности.// Психологический журнал, Том 11, № 5, 1990. - С.44
10. Психофизиологические функции человека и операторский труд.-Киев: Наук. Думка, 1991. - 216 с.
11. Белоус В. В. К исследованию влияния типов темперамента на эффективность индивидуальной и совместной деятельности.// Вопросы психологии, № 3, 1986.- С.113-120.

УДК 519.7:556

ФОРМИРОВАНИЕ СТРУКТУРЫ ПОЛИНОМИАЛЬНОЙ НЕЙРОННОЙ СЕТИ С ПОМОЩЬЮ ГЕНЕТИЧЕСКОГО АЛГОРИТМА

Е.Ф.Васечкина, В.Д.Ярин

Пропонується новий метод побудови архітектури односпрямованої поліноміальної нейронної мережі з одночасною ідентифікацією вагових коефіцієнтів. Алгоритм базується на сполученні Методу групового обліку аргументів і генетичного алгоритму. Метод застосовується для рішення задач математичного моделювання параметрів водного середовища. Результати порівнюються з отриманими іншими методами на тих же даних.

Предлагается новый метод построения архитектуры однонаправленной полиномиальной нейронной сети с одновременной идентификацией весовых коэффициентов. Алгоритм базируется на сочетании метода группового учета аргументов и генетического алгоритма. Метод применяется для решения задач математического моделирования параметров водной среды. Результаты сравниваются с полученными другими методами на тех же данных.

The new method of evolving a structure of feed-forward polynomial neural network with simultaneous weights identification is proposed. The algorithm is based on combination of Group Method of Data Handling and Genetic Algorithm. The method is applied to tasks of mathematical simulation of marine environment. Results of practical application of the method are compared with those obtained by other methods on the same data.

ВВЕДЕНИЕ

Разработка фундаментальных основ морской экологии в настоящее время еще далека от завершения в связи с тем, что объект исследования является так называемой "большой системой" и для его сколько-нибудь полного описания не существует готового математического аппарата. Между тем прогрессирующее ухудшение экологической ситуации требует разработки моделей, способных при существующем уровне знаний оценивать текущую ситуацию и осуществлять прогнозы, на основе которых можно принимать те или иные решения в области хозяйственных и природоохранных мероприятий [1]. В таких задачах, по нашему мнению, может с успехом

применяться технология нейронных сетей (НС) [2,3], поскольку их универсальность позволяет строить модели без требования полного понимания и математического описания взаимодействий в системе.

Целью настоящей работы является разработка алгоритма построения функциональной зависимости между компонентами вектора состояния произвольной системы. Эта задача при отсутствии четкого представления о структуре связей в системе лучше поддается решению с помощью многослойных НС, чем классическими методами. При этом определение архитектуры НС имеет первостепенное значение, поскольку она влияет на эффективность процесса обучения и производительность сети при решении конкретных задач. В последнее время появилось большое количество работ, в которых предлагались различные способы построения структуры НС. Для этого в основном применяются различные модификации генетических алгоритмов (ГА), и проблема построения архитектуры сети рассматривается как задача поиска экстремума в многомерном пространстве, в котором каждая точка представляет собой некоторую сетевую архитектуру [4,5,6]. Применение этих методов предполагает обучение сети как отдельный процесс, который необходимо осуществить для всех вариантов структуры сети, чтобы оценить их производительность. Для обучения сети чаще всего применяется метод обратного распространения ошибки. При использовании этого метода весовые коэффициенты и пороговые значения узлов сети отыскиваются с помощью алгоритма градиентного спуска, но структура связей между узлами должна при этом быть задана априори. Такие расчеты требуют значительных вычислительных затрат.

Представляется более перспективным одновременное определение структуры связей НС и соответствующих весовых коэффициентов. Такие алгоритмы были рассмотрены в ряде работ, где предлагалось расширение генотипа путем включения в него весовых коэффициентов тем или

инным способом, например [7,8]. Однако пространство возможных решений при этом значительно расширяется, что существенно затрудняет поиск оптимальной архитектуры с помощью эволюционного алгоритма. При этом особое значение приобретает разработка специальных эволюционных операторов, зависящих от типа кодирования, поскольку установлено их существенное влияние на эффективность поиска.

В настоящей работе предлагается более простой метод построения архитектуры однонаправленной полиномиальной НС с одновременным оцениванием весовых коэффициентов. Он основывается на сочетании Метода группового учета аргументов (МГУА), разработанного А.Г.Ивахненко [9,10] и ГА. МГУА обладает рядом отличительных особенностей:

- способностью выбирать действительно важные предикторы из десятков представленных на входе, существенно понижая размерность задачи;
- автоматически формировать структуру связей между узлами;
- не требует обучения, как отдельного процесса, поскольку весовые коэффициенты рассчитываются в процессе формирования структуры сети;
- пороговые уровни рассчитываются одновременно с весовыми коэффициентами с помощью МНК;
- использует линейную функцию активации;
- позволяет получить удовлетворительные результаты при ограниченном количестве исходных данных, когда применение других методов вообще проблематично;
- позволяет записывать полученные модели в аналитическом виде.

Несмотря на то, что МГУА предполагает линейность активационных функций нейронов, получаемые зависимости не являются линейными, поскольку на вход каждого нейрона поступают как линейные, так и нелинейные комбинации сигналов от других нейронов. Формирование структуры сети в МГУА осуществляется путем сокращенного перебора возможных вариантов с оценкой эффективности каждого. Авторами было предложено для совершенствования этой процедуры использовать ГА. Простые расчеты показывают, что использование ГА при большой размерности входного вектора позволяет существенно сократить объем вычислений, а также увеличить точность получаемого решения, поскольку ГА обеспечивает значительно большую вероятность отыскания глобального минимума целевой функции, чем сокращенный перебор, традиционно используемый в МГУА.

1 АЛГОРИТМ ПОСТРОЕНИЯ НС

В общем случае архитектура однонаправленной НС определяется числом слоев, количеством узлов в каждом слое и конфигурацией связей между узлами. Сеть строится последовательным наращиванием числа слоев, причем для каждого узла конфигурация его связей с другими узлами определяется с помощью ГА. Генетический алгоритм относится к группе так называемых эволюционных алгоритмов, в основе которых лежит идея поиска решения задачи путем имитации его "биологической" эволюции

[11,12]. Алгоритм генерирует и поддерживает "популяцию" потенциальных решений (индивидуумов) исследуемой проблемы. Некоторые из этих решений используются для создания новых решений путем применения к ним специальных операторов, имитирующих процессы биологической эволюции. Это операторы селекции, скрещивания (кроссовера) и мутации. Таким образом, в алгоритме реализуется дарвиновская триада движущих сил эволюции: отбор, наследственность, изменчивость. Потенциальные решения для оператора скрещивания отбираются на основе оценки их функции качества согласно введенным критериям, определяемым конкретной задачей. Решения, соответствующие введенным критериям, имеют большие шансы "дать потомков" в новое поколение. Генерация новых поколений популяции потенциальных решений осуществляется до выполнения некоторого условия, обычно связанного с глубиной найденного экстремума функции качества. Эффективность алгоритма зависит от многих аспектов, среди которых важное место занимает формализация задачи и способ кодирования ее потенциального решения в виде символьной строки - "генотипа" экземпляра популяции. Решение этого вопроса зависит от специфики задачи. Выбранный вид кодирования определяет в дальнейшем алгоритмические особенности применения эволюционных операторов кроссовера и мутаций. Генотип должен однозначно преобразовываться в обычное представление решения задачи, удобное и понятное исследователю. Такое представление в терминах эволюционного моделирования называется "фенотипом" индивидуума популяции. Функция качества индивидуума может быть рассчитана только после преобразования его генотипа в фенотип с помощью соответствующей процедуры.

В работе используется стандартный ГА, оперирующий с популяцией индивидуумов, являющейся некоторым подмножеством множества всех возможных конфигураций связей одного из узлов в слое, начиная с первого (набор входных узлов принимается за нулевой слой), с узлами предыдущего слоя сети. Чтобы найти наилучшую структуру связей между слоями НС, в каждом узле сети с помощью ГА осуществляется поиск экстремума функции качества, характеризующей успешность подобранной конфигурации связей для данного узла.

Для обучения сети сформируем таблицу данных, столбцы которой представляют собой набор факторов, функционально связанных между собой. Ими могут быть составляющие вектора состояния системы, а также внешние по отношению к системе параметры. В столбцы таблицы записывается последовательность их значений, согласованных определенным образом. В частности, это могут быть временные ряды параметров исследуемой динамической системы с различной степенью запаздывания. Строку такой таблицы, представляющую собой набор значений факторов в один момент времени, будем называть точкой таблицы данных. Например, для вектора состояния системы $X = (X_1, X_2, \dots, X_n)$ и вектора внешних параметров $B = (B_1, B_2, \dots, B_i)$, каждая строка таблицы может иметь вид

$$X_{1,k} \dots X_{2,k} X_{2,k-1} X_{2,k-2} \dots X_{n,k} \dots X_{n,k-m} \dots, \\ \dots B_{i,k} \dots B_{i,k-m}$$

где k - момент времени, m - максимальное запаздывание.

Обозначим через Z_k последовательность известных на обучающей выборке значений искомой переменной, а соответствующий им вектор входных параметров - U_k .

Поскольку принципиально существует возможность в первом же слое сети получить решение задачи, мы можем рассматривать последовательность значений на выходе любого n -го узла сети в слое m как некоторое ее частное решение вида

$$Y_{nk} = a_{n0} + \sum_i^F a_{ni} X_{ik} + \sum_i^F \sum_j^F a_{nij} X_{ik} X_{jk} + \dots, \quad (1)$$

где F - количество узлов в слое $m-1$;

a_{n0} - пороговый уровень нейрона;

a_{ni} - весовые коэффициенты входных узлов;

X_{ik} - численные значения, полученные на выходе нейронов предыдущего слоя сети, с которыми связан данный узел.

Такое представление является достаточно универсальным, поскольку соответствует полиному Колмогорова-Габор [9]. Если ограничиться в правой части линейными членами, получим обычное уравнение множественной регрессии. Учет членов разложения более высокого порядка существенно расширяет пространство поиска, зато дает возможность получить более точную модель выходного параметра системы. Практическое применение алгоритма показало, что для задач идентификации сложных динамических систем, особенно, природных, линейного представления оказывается недостаточно и требуется более полное описание выходного сигнала нейрона с учетом по крайней мере квадратичных членов.

Определим генотип экземпляра популяции в виде битовой строки вида $L = (001100010\dots01)$, где "1" означает наличие соответствующего члена в уравнении (1), считая по порядку слева направо ($a_{nij} \neq 0$), а "0" - отсутствие фактора либо произведения факторов ($a_{nij} = 0$). Для линейно-го варианта сети выход нейрона можно представить в виде

$$Y_{nk} = a_{n0} + \sum_j^F a_{nj} X_{jk}. \quad (2)$$

Тогда положение "1" в строке генотипа обозначает наличие связи данного узла с соответствующим узлом предыдущего слоя, а "0" - ее отсутствие. Преобразование генотипа экземпляра популяции в фенотип (частное решение задачи) осуществляется с помощью (1) или (2). В общем случае отличие от нуля коэффициента a_{nij} перед произведением нескольких факторов в (1) порождает

возникновение сразу нескольких связей между нейронами соседних слоев. Схема результирующей нейронной сети представлена в графическом виде на рис.1.

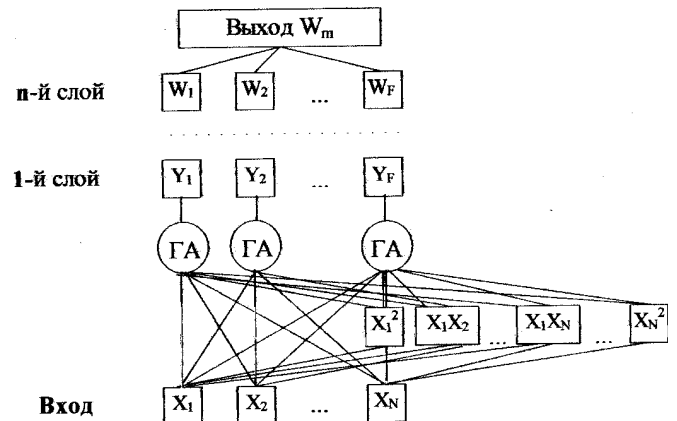


Рисунок 1 - Нейронная сеть на основе МГВА. Все последующие слои сети идентичны первому

Для вычисления весовых коэффициентов в формуле (1) потребуем, чтобы значения на выходе нейрона Y_{nk} были как можно ближе к точным значениям искомого параметра Z_k , заданным для обучения сети, по критерию минимума среднеквадратической ошибки. Подставляя в (1) данные из выборки и осредняя полученные соотношения, получим линейную систему нормальных уравнений Гаусса. Решением системы нормальных уравнений является набор весовых коэффициентов, соответствующих ненулевым позициям в генотипе L . Заметим, что при оценке коэффициентов первого слоя X_{ik} являются числовыми значениями входных узлов сети U_{ik} , а в последующих слоях X_{ik} - это последовательность значений, полученных на выходе нейронов предыдущего слоя сети. Таким образом, таблица данных для обучения сети при наращивании числа слоев последовательно переписывается замещением предыдущих значений вновь полученными в последнем из добавленных слоев. Многослойность алгоритма позволяет строить модель рекурсивно и многократно использовать одни и те же точки выборки данных для все более точной подгонки уравнения множественной регрессии.

Для обеспечения способности сети к обобщению (т.е. экстраполяции), необходимо оценивать эффективность подбора конфигурации связей нейрона на данных, которые не использовались для идентификации весовых коэффициентов. Поэтому вся имеющаяся для обучения сети выборка делится на три части: обучающую, контрольную и экзаменационную последовательности, причем контрольная должна быть не меньше обучающей. На обучающей части выборки с помощью метода наименьших квадратов определяются весовые коэффициенты, соответствующие "1" в генотипе. На контрольной части выборки оценивается функция качества такого частного решения (фенотипа индивидуума популяции) как среднеквадратическая

ошибка этого решения относительно известных величин выходного параметра Z_k . Эта процедура повторяется для каждого индивида популяции, и в результате применения стандартных операторов кроссовера, мутации и селекции через определенное небольшое число поколений отыскивается локальный оптимум конфигурации связей данного узла. Поскольку для каждого из узлов требуется отыскать лишь локальный оптимум в пространстве связей, для его нахождения обычно бывает достаточно небольшой численности популяции (≤ 150) и ограниченного числа поколений (≤ 30).

Повторив эту процедуру для каждого из узлов, количество которых ограничивается заданным максимальным числом, получим конфигурацию потенциальных связей данного слоя с предыдущим. Та же последовательность действий повторяется в последующих слоях сети, число которых также ограничивается максимально допустимым. Для слоя в целом определяется минимальное значение функции качества в узлах. Процесс наращивания числа слоев продолжается до тех пор, пока эта величина не достигнет своего минимума. В последнем слое сети имеется попрежнему набор узлов, каждый из которых представляет собой вариант решения задачи с соответствующей ему сетевой архитектурой. Окончательный выбор решения осуществляется на экзаменационной последовательности, которая до этого момента не использовалась в процедуре поиска. Последовательность Y_{mk} , дающая минимальную ошибку относительно Z_k на экзаменационной последовательности, определяет выбор выходного узла с номером m , а вместе с ним и окончательный выбор архитектуры сети.

Для инициализации процедуры построения структуры НС необходимо определить следующие параметры:

- число входных узлов и соответствующую им таблицу данных, на которой будет обучаться сеть;
- максимальные размеры сети, т.е. максимально допустимое число слоев и нейронов в каждом слое;
- численность популяции и максимально допустимое число поколений генетического алгоритма.

Учет квадратичных членов в представлении (1) влечет за собой резкий рост объема выборки данных, необходимых для обучения сети. Это обстоятельство представляет значительную проблему, особенно для природных систем, где данных, как правило, весьма ограниченное количество. Предположим, что максимально допустимое число узлов в слое равно F . Количество слагаемых в описании (1) при учете квадратичных членов будет равно $F(F+3)/2 + 1$. Учтем, что для сколь угодно репрезентативной оценки весовых коэффициентов необходимо, чтобы каждое нормальное уравнение строилось осреднением как минимум по 5 точкам таблицы данных. Таким образом, для типичных условий ($F = 20$) одновременное оценивание такого количества коэффициентов потребовало бы $231 \cdot 5 = 1155$ точек в таблице данных. Однако в практических задачах длина выборки, предназначенной для обучения сети, часто значительно меньше. Чтобы обойти эту трудность, необходимо ограничить возможное число битовых единиц

в генотипе до одной пятой от длины обучающей последовательности. Таким образом, на вход каждого слоя сети будут поступать лишь частные описания, в которых большая часть коэффициентов a_{nij} равна нулю. Поскольку схемы связей для каждого из узлов в слое различаются между собой, каждый слой в целом сохраняет необходимое разнообразие схем, а сам процесс поиска в многомерном пространстве осуществляется поэтапно, при последовательном усложнении структуры сети.

Численные эксперименты показали, что эффективность построенной сети зависит от того, какие именно точки таблицы данных использовались для расчета весовых коэффициентов и какие для оценки функции качества частных решений. В связи с этим возникает вопрос об оптимальном упорядочивании входной последовательности. В работе А.Г.Ивахненко [10] было предложено ранжировать все точки таблицы данных по дисперсии, которая рассчитывалась по формуле

$$D_k = \frac{1}{N} \sum_i \frac{(U_{ik} - [U_i])^2}{[U_i]^2},$$

где U_{ik} - значение i -того входного параметра в k точке таблицы данных,

N - число входных параметров,

$[U_i]$ - среднее значение i -того входного параметра, полученное на имеющейся выборке данных, исключая экзаменационную последовательность.

При этом разбиение выборки данных на обучающую, контрольную и экзаменационную последовательности осуществлялось после ранжирования. Однако практическое применение алгоритма показало, что далеко не всегда эта процедура существенно повышает качество сети, более того, отмечены случаи, когда точность результатов снижается. Поэтому в настоящей работе предлагается осуществлять селекцию точек в таблице данных также с помощью ГА, но с расчетом существенно ограниченного числа поколений. С этой целью вся выборка данных за исключением экзаменационной последовательности была пронумерована, и в качестве генотипа экземпляра популяции принималась рандомизированная последовательность номеров точек. Фенотипом индивида популяции в данном случае являлось решение задачи, полученное на НС заданной архитектуры с весовыми коэффициентами, идентифицированными по выборке, определенной генотипом.

Сформированная в соответствии с генотипом выборка данных была разделена на две равные части. На первой половине - обучающей последовательности - рассчитывались весовые коэффициенты, а на второй - контрольной - оценивалась среднеквадратическая ошибка полученного решения, т.е. значение функции качества экземпляра популяции. Для определения оптимальной последовательности точек для расчета весовых коэффициентов НС заданной архитектуры достаточно лишь нескольких поколений ГА, поскольку значение имеет лишь попадание точки таблицы данных в ту или иную последовательности, но не порядок следования точек.

Необходимо отметить, что применение предлагаемого метода уточнения сетевой структуры с помощью подбора точек для обучающей и контрольной последовательностей при учете квадратичных членов в формуле (1) в ряде конкретных задач снижало способность НС к обобщению (экстраполяции). Очевидно, что предпочтительной будет менее точная аппроксимация на контрольной последовательности, но дающая прогноз на экзаменационной последовательности с ошибкой, не превышающей допустимую. При применении алгоритма в линейном варианте такого явления не наблюдалось.

2 РЕЗУЛЬТАТЫ ТЕСТИРОВАНИЯ АЛГОРИТМА

Предлагаемый алгоритм тестировался на двух задачах. В первой из них - подборе формулы расчета условной плотности по измеренным температуре и солёности, т.е. определении уравнения состояния морской воды - функциональная зависимость была известна. Для второй тестовой задачи - построение регрессионной модели прогноза стока воды - результаты сравнивались с полученными ранее другими авторами, а также с расчетами, выполненными с помощью стандартных методов.

2.1 Идентификация уравнения состояния морской воды

В этой задаче были использованы профили температуры и солёности, полученные в Черном море с помощью комплекса МГИ 4102 (Исток), имеющиеся в банке данных МГИ НАН Украины [13]. Результаты зондирований представлены в виде таблиц значений глубины (H), температуры (T) и солёности (S). Эти параметры являлись входными для алгоритма. Для расчета эталонной условной плотности использовалось уравнение УС-80 [14]

$$\begin{aligned} \sigma_t = & -0,157436 + 6,793952 \cdot 10^{-2}T - 9,09529 \cdot 10^{-3}T^2 + \\ & + 1,001685 \cdot 10^{-4}T^3 - 1,120083 \cdot 10^{-6}T^4 + \\ & + 6,536332 \cdot 10^{-9}T^5 + (8,24493 \cdot 10^{-1} - 4,0899 \cdot 10^{-3}T + \\ & + 7,6438 \cdot 10^{-5}T^2 - 8,2467 \cdot 10^{-7}T^3 + 5,3875 \cdot 10^{-9}T^4)S + \\ & + (-5,72466 \cdot 10^{-3} + 1,0227 \cdot 10^{-4}T - \\ & - 1,6546 \cdot 10^{-6}T^2)S^{3/2} + 4,8314 \cdot 10^{-4}S^2. \end{aligned} \quad (3)$$

Требовалось как можно более точно аппроксимировать последовательность значений σ_t . Выборка данных для обучения сети составляла 200 точек - один гидрологический профиль до глубины 200м - распределенных следующим образом: по 90 точек на тренировочную и контрольную последовательности и 20 на экзаменационную. После нескольких прогонов алгоритма был выбран наилучший

вариант НС, дающий меньшую ошибку аппроксимации на экзаменационной последовательности. Полученную структуру НС можно легко преобразовать к обычному алгебраическому виду

$$\begin{aligned} \sigma_t = & -7,169027 \cdot 10^{-1} - 8,425296 \cdot 10^{-2}T + \\ & + 9,923865 \cdot 10^{-1}S - 1,7624174 \cdot 10^{-2}S^2 + \\ & + 1,257996 \cdot 10^{-2}TS + 7,446529 \cdot 10^{-5}T^3 - \\ & - 4,347284 \cdot 10^{-4}TS^2 - 8,929592 \cdot 10^{-4}T^2S + \\ & + 6,591923 \cdot 10^{-4}S^3 - 5,955412 \cdot 10^{-7}T^4 + \\ & + 2,580612 \cdot 10^{-5}T^2S^2 - 8,340646 \cdot 10^{-6}S^4. \end{aligned} \quad (4)$$

Применение уравнения (4) вместо (3) для расчета аномалии плотности на 50 контрольных станциях, выполненных в разное время года, показало, что относительная среднеквадратическая ошибка аппроксимации (отношение среднеквадратической ошибки к стандартному отклонению реализации, представляющей собой точное решение задачи) варьируется от 0.0005% до 0.005% (абсолютная - от $2 \cdot 10^{-12}$ до $2 \cdot 10^{-10}$). Такую ошибку можно признать вполне удовлетворительной.

2.2 Регрессионная модель для прогноза стока воды из озера

Цель расчетов состояла в подборе регрессионного уравнения, отражающего зависимость стока воды из озера Таупо (Новая Зеландия) от величины осадков.

Данные (рис. 2) были взяты из работы [15] и представляли собой временные ряды осадков, измеренных на станции недалеко от озера, и стока за 1971 год с дискретностью в один день. Статистический анализ данных показывает, что их распределение далеко от нормального - осадки имеют одномодальное распределение, сдвинутое влево от нормального, а сток - трехмодальное. В работе [15] для построения одношаговой прогностической модели было выделено три режима функционирования системы (нормальный, увеличение объема стока и уменьшение) и для каждого из них подбирались регрессионные уравнения. Параллельно рассчитывалась вероятность наступления каждого из режимов. Прогноз представлял собой взвешенную сумму вычисленных по трем уравнениям величин, с весовыми коэффициентами, пропорциональными вероятности наступления того или иного режима. Ошибку прогноза можно оценить по приведенным в работе таблицам, минимум ее составляет 78% от стандартного отклонения искомого ряда. Представляет интерес протестировать на этих рядах многослойный перцептрон (МСП) [2], НС на основе МГУА и уравнение множественной линейной регрессии, полученное традиционным способом, и сравнить их эффективность.

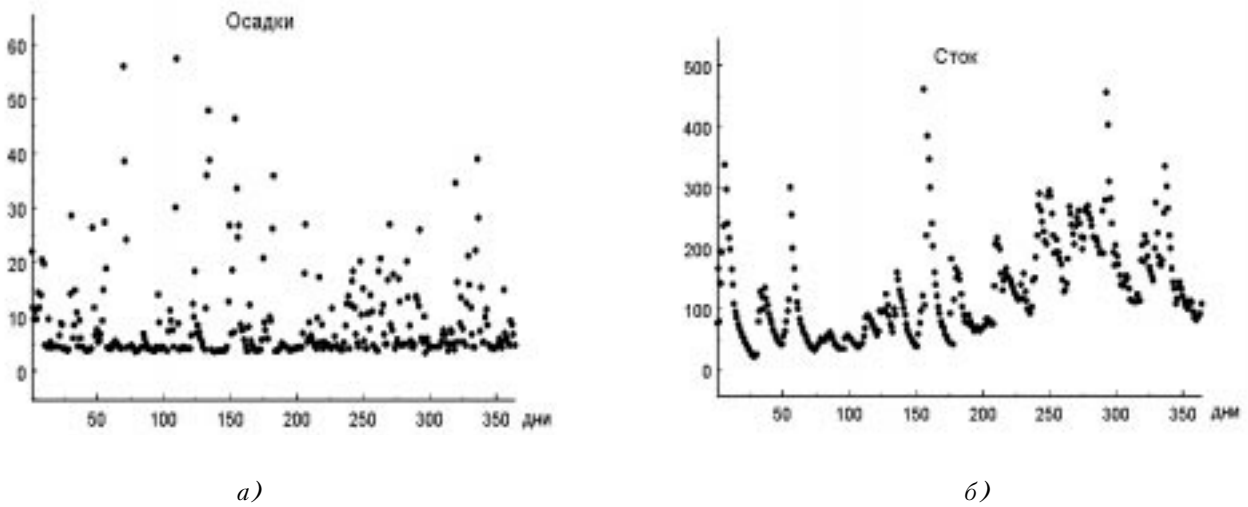


Рисунок 2 - Суточные данные об осадках и стоке за 1971 год. Озеро Таупо, Новая Зеландия

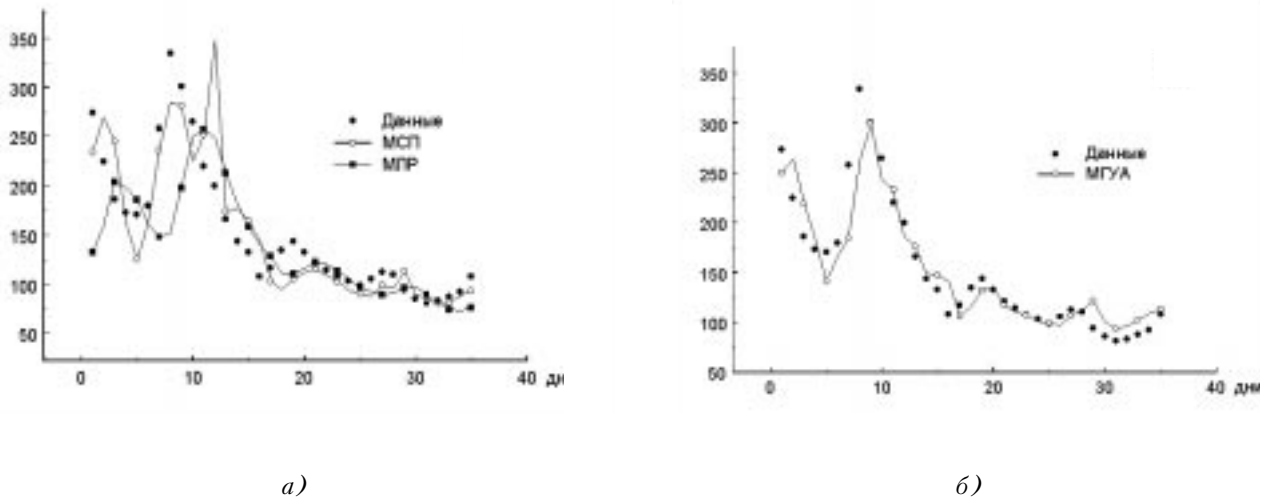


Рисунок 3 - Сравнение результатов модельных расчетов реальными данными. а) множественная линейная (МСП); б) НС на основе МГУА

С этой целью рассчитывался одношаговый прогноз для точных значений стока и прогноз среднего стока за пять суток с упреждением в 5 дней. Для одношагового прогноза в качестве предикторов рассматривались следующие переменные: осадки и сток с запаздыванием от одних до пяти суток, средняя величина осадков за 5, 10, 15, 20, 25 и 30 дней до текущего момента, а также средний за 5 дней сток с запаздыванием 1, 5 и 7 суток, всего 19 переменных. Длина рядов составляла 335 отсчетов, из которых 35 последних точек были выделены на контрольную последовательность и не использовались в расчетах. По выборке из 300 точек с помощью МНК были рассчитаны коэффициенты уравнения множественной

линейной регрессии, полученное уравнение на контрольной последовательности (рис.3а) дало относительную ошибку 52% (к стандартному отклонению искомого ряда).

Для построения МСП данные разделялись на тренировочную и проверочную последовательности, длина которых варьировалась. Структура МСП подбиралась методом проб и ошибок. Наилучшая аппроксимация с помощью МСП была получена при использовании шести слоев со следующим распределением нейронов по слоям, считая от нулевого (входного слоя): 19, 25, 16, 12, 10, 1. Средняя относительная ошибка прогноза (рис. 3а) на контрольной последовательности составила 46%.

При построении НС на основе МГУА выборка данных делилась в пропорции 3:1, оптимальная модель получалась в первом слое сети и имела среднюю относительную ошибку на контрольной последовательности 30% (рис. 36). Построение прогностической модели для среднего стока с пятидневным упреждением оказалось более сложной задачей, полученные средние ошибки прогноза составили 55%, 50% и 38% для моделей линейной регрессии, МСП и МГУА соответственно.

3 ЗАКЛЮЧЕНИЕ

Предложенный метод поиска функциональных зависимостей между компонентами вектора состояния природной системы даже на ограниченных выборках данных позволяет находить решение с ошибкой, не превышающей 25-35% от среднеквадратического отклонения искомой переменной. Вопросы использования НС в моделировании экологических и экологоэкономических систем требуют дальнейшей разработки, поскольку такой подход является безусловно перспективным. Это подтверждается множеством примеров успешного использования НС в других областях, сходных по ряду признаков с экологическим моделированием, а также опытом применения НС в гидрологии и океанологии. Приведенные в работе примеры построения регрессионных моделей иллюстрируют лишь одно из возможных применений НС в моделировании сложных систем.

УДК 519.2

МОДЕЛИРОВАНИЕ ПРОЦЕССА ОБУЧЕНИЯ

А.И.Вершина, Б.Т.Солдатов

Запропонована модель процесу навчання, яку засновано на припущенні, що імовірність одержання знань в нескінченно малому проміжку часу пропорційна розміру цього проміжку. Розглянуто умови, при яких навчання можна описати як марківський процес. Отримано формули для очікуваних витрат часу на навчання та їхні дисперсії.

Предложена модель процесса обучения, основанная на предположении, что вероятность получения знаний в бесконечно малом промежутке времени пропорциональна размеру этого промежутка. Рассмотрены условия, при которых обучение можно описать как марковский процесс. Получены выражения для ожидаемых затрат времени на обучение и их дисперсии.

The mode of process of training based on the assumption is offered that the probability of reception of knowledge in infinitesimal an interval of time is proportional to the size of this interval. The conditions are considered, at which training is possible to describe as process of Markov. The expressions for expected expenses of time for training and their dispersion are received.

ПЕРЕЧЕНЬ ССЫЛОК

1. Беляев В.И., Кондуфорова Н.В., Математическое моделирование экологических систем шельфа. - Киев, "Наукова думка", 1990, 240с.
2. Заенцев И.В. Нейронные сети: основные модели. - Изд. Воронежского ун-та, 1999, 76с.
3. Muller B., Reinhardt J. Neural Networks. An Introduction. - Berlin, Springer-Verlag, 1991, 266с.
4. Arena P., Caponetto R., Fortuna L., Xibilia M. (M.L.P.) Optimal Topology via Genetic Algorithms // Proceedings, International Conference on Artificial Neural Nets and Genetic Algorithms, 1993, P. 670- 674.
5. Miller G., Todd P., and Hegde S. Designing Neural Networks Using Genetic Algorithms // Proceedings, Third International Conference on Genetic Algorithms, 1989, P. 379- 384.
6. Stepniewski S. and Keane A. Topology Design of Feed-Forward Neural Networks // Proceedings, Parallel Problem Solving from Nature, 1996, P. 771- 780.
7. Maniezzo V. Genetic Evolution of the Topology and Weight Distribution of Neural Networks // IEEE Transactions on Neural Networks, 1994, Vol. 5, №1, P. 39- 53.
8. Pujol J. Evolution of Artificial Neural Networks Using a Two-Dimensional Representation. - Ph.D. Thesis, School of Computer Science, University of Birmingham, 1999. 177p.
9. Ивахненко А.Г. Самоорганизация прогнозирующих моделей. - Киев, "Техника", 1985, 223с.
10. Ивахненко А.Г.. Системы эвристической самоорганизации в технической кибернетике. - Киев, "Техника", 1971, 372с.
11. Beasley D., Bull D.R., Martin R.R., An Overview of Genetic Algorithms: Part I, Fundamentals // University Computing, 1993, 15, №2, P. 58-69.
12. Beasley D., Bull D.R., Martin R.R., An Overview of Genetic Algorithms: Part II, Research Topics // University Computing, 1993, 15, №4, P. 170-181.
13. Банк данных МГИ НАН Украины. URL: <http://www.mhi.iuf.net/DEPTS/>.
14. Обработка данных океанографической станции. Сост. ред. коллегией ОГОТС. - ЮНЕСКО-МГИ, Севастополь, МГИ, ISBN 5-7702-0643-8, 1991, 87с.
15. Lu Z.Q., Berliner L.M. Markov Switching Time Series Models with Application to Daily Runoff Series // Water Resour. Res., 1999, V.35, № 2, P. 523-534.

ВВЕДЕНИЕ

Эффективность обучения зависит от многих факторов. Поиск новых методов обучения является в настоящее время актуальной задачей. Наряду с этим, возникает проблема определения целесообразности введения новых методов в конкретных условиях. Эта проблема связана с необходимостью создания математической модели процесса обучения. В данной работе предлагается один из возможных путей создания этой модели.

ПОСТАНОВКА ЗАДАЧИ

Процесс обучения состоит в усвоении определенного объема знаний по некоторой дисциплине. Его можно разбить на два этапа, первый из которых состоит в передаче информации обучающемуся, а второй - в проверке результатов усвоения материала с целью определения достигнутого уровня знаний и, в случае необходимости, повторной передаче обучающемуся не усвоенной им информации. Результаты выполнения

этапов зависят от многих субъективных и объективных факторов и, в общем случае, представляет собой случайный процесс.

Если считать, что существует оценка качества усвоения знаний как вероятность того, что данный материал усвоен, и оценка качества проверки знаний как вероятность того, что в процессе проверки выявлен достоверный уровень знаний, то процесс обучения можно описывать вероятностными характеристиками. Это позволяет поставить задачу построения математической модели процесса обучения как случайного процесса.

МЕТОДЫ РЕШЕНИЯ

Пусть вероятность $\Delta k(t)$ усвоения элемента знаний в малом промежутке времени Δt пропорциональна величине этого промежутка, то можно записать

$$\Delta k(t) = k(t + \Delta t) - k(t) = [1 - k(t)]\lambda \Delta t, \quad (1)$$

где $k(t)$ - вероятность того, что элемент знаний за время t усвоен;

λ - коэффициент пропорциональности, который отражает интенсивность усвоения знаний.

Переходя к пределу $\Delta t \rightarrow 0$, получим дифференциальное уравнение

$$dk(t) = [1 - k(t)]\lambda dt, \quad (2)$$

решение которого имеет вид

$$k(t) = 1 - e^{-\lambda t}. \quad (3)$$

Плотность распределения времени усвоения элемента знаний определяется выражением

$$p_k(t) = \frac{dk(t)}{dt} = \lambda e^{-\lambda t}, \quad (4)$$

то есть подчиняется экспоненциальному закону.

Усвоение определенного объема знаний, представляющего собой совокупность α элементов знаний приводит к композиции законов распределения элементов знаний. Так как показательная плотность является частным случаем гамма-распределения [1, 2], а семейство гамма-плотностей замкнуто относительно операции свертки, то плотность распределения времени на усвоение некоторого объема знаний имеет вид

$$p(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}, \quad (5)$$

где $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$ - гамма-функция Эйлера.

Для произвольного времени T , затраченного на усвоение определенного объема знаний, вероятность $K(T)$ усвоения знаний определяется выражением

$$K(T) = \int_0^T p(t) dt = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^T t^{\alpha-1} e^{-\lambda t} dt. \quad (6)$$

Ожидаемое время на усвоение знаний равно

$$T_{yc} = \int_0^\infty t p(t) dt = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty t^\alpha e^{-\lambda t} dt = \frac{\alpha}{\lambda}. \quad (7)$$

Как правило, время на обучение T_0 ограничено, и оно пропорционально ожидаемому времени на усвоение знаний T_{yc} :

$$T_0 = \mu T_{yc} = \mu \frac{\alpha}{\lambda}, \quad (8)$$

где μ - коэффициент пропорциональности.

В этом случае имеем

$$K(T_0) = \int_0^{\mu \frac{\alpha}{\lambda}} p(t) dt = \int_0^{\mu \frac{\alpha}{\lambda}} t^{\alpha-1} e^{-\lambda t} dt. \quad (9)$$

После внесения λ^α под знак интеграла, замены переменной $z = \lambda t$ и изменения пределов интегрирования получим

$$K(T_0) = \frac{1}{\Gamma(\alpha)} \int_0^{\mu \alpha} z^{\alpha-1} e^{-z} dz = K(\mu, \alpha). \quad (10)$$

Из этого следует, что вероятность усвоения знаний в случае выделения времени на обучение пропорционально ожидаемому времени \bar{T} не зависит от значения λ .

Если за заданное время T_0 объем знаний не усвоен, то необходимо дополнительное время T_d на усвоение. Это время разумно также выделять пропорционально ожидаемому времени на усвоение знаний. Если считать, что усвоение знаний в одинаковой мере затрагивает каждый элемент знания, то количество элементов α останется прежним, а изменится значение λ . В соответствии с полученным выражением (10), вероятность усвоения знаний за дополнительное время не изменится. Можно также сказать, что время, выделяемое на усвоение определенного объема знаний, определяется заданным уровнем качества обучения.

Оценим ожидаемое дополнительное время T_d , которое необходимо выделить на усвоение материала:

$$T_d = \int_{T_0}^\infty (t - T_0) p(t) dt = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_{T_0}^\infty (t - T_0) t^{\alpha-1} e^{-\lambda t} dt = \frac{\lambda^\alpha}{\Gamma(\alpha)} \left[\int_{T_0}^\infty t^\alpha e^{-\lambda t} dt - T_0 \int_{T_0}^\infty t^{\alpha-1} e^{-\lambda t} dt \right]. \quad (11)$$

Используя для первого слагаемого в квадратных скобках метод интегрирования по частям, вводя функции $u = t^\alpha$ и $dv = e^{-\lambda t} dt$, для которых имеем $du = \alpha t^{\alpha-1} dt$ и $v = \int e^{-\lambda t} dt = -\frac{1}{\lambda} e^{-\lambda t}$, с учетом того, что

$$\frac{\lambda^\alpha}{\Gamma(\alpha)} \int_{T_0}^{\infty} t^{\alpha-1} e^{-\lambda t} dt = 1 - K(T_0), \text{ получим}$$

$$T_d = \frac{\lambda^\alpha}{\Gamma(\alpha)} \left[-t^\alpha \frac{1}{\lambda} e^{-\lambda t} \Big|_{T_0}^{\infty} + \frac{\alpha}{\lambda} \int_{T_0}^{\infty} t^{\alpha-1} e^{-\lambda t} dt - T_0 \int_{T_0}^{\infty} t^{\alpha-1} e^{-\lambda t} dt \right] = \frac{\lambda^\alpha}{\Gamma(\alpha)} T_0 \frac{1}{\lambda} e^{-\lambda T_0} + \left[\frac{\alpha}{\lambda} - T_0 \right] [1 - K(T_0)]. \quad (12)$$

Подставляя T_0 из формулы (8) в данное выражение, получим

$$T_d = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{1}{\lambda} \left[\mu \frac{\alpha}{\lambda} e^{-\lambda \mu \frac{\alpha}{\lambda}} + (1 - \mu) \frac{\alpha}{\lambda} [1 - K(\mu \frac{\alpha}{\lambda})] \right] = \frac{1}{\lambda \Gamma(\alpha)} (\mu \alpha)^\alpha e^{-\mu \alpha} + (1 - \mu) \frac{\alpha}{\lambda} [1 - K(\mu, \alpha)]. \quad (13)$$

Будем считать, что дополнительное время выделяется также пропорционально ожидаемому времени, то есть,

$$T_{0d} = \mu T_d. \quad (14)$$

Отсюда отношение δ дополнительного времени на усвоение материала к времени, выделенному на первоначальное усвоение, равно

$$\delta = \frac{T_{0d}}{T_d} = \frac{\mu \left\{ \frac{1}{\lambda \Gamma(\alpha)} (\mu \alpha)^\alpha e^{-\mu \alpha} + (1 - \mu) \frac{\alpha}{\lambda} [1 - K(\mu, \alpha)] \right\}}{\mu \frac{\alpha}{\lambda}} = \frac{(\mu \alpha)^\alpha}{\Gamma(\alpha + 1)} e^{-\mu \alpha} + (1 - \mu) [1 - K(\mu, \alpha)]. \quad (15)$$

Из полученного выражения следует, что отношение дополнительного времени на обучение к первоначально выделяемому времени будет постоянной величиной, если считать, что $\mu = \text{const}$ и $\alpha = \text{const}$.

Аналогичные рассуждения приводят к выражению для качества проверки знаний как вероятности $A(T_{\Pi})$ того, что в результате проверки в течение времени T_{Π} получен достоверный результат об усвоенном знании:

$$A(T_{\Pi}) = \int_0^{T_{\Pi}} p(t) dt = \frac{\lambda_{\Pi}^\alpha}{\Gamma(\alpha)} \int_0^{T_{\Pi}} t^{\alpha-1} e^{-\lambda_{\Pi} t} dt. \quad (16)$$

Следует отметить, что на практике вероятность качественной проверки усвоенного и неусвоенного знания дает различные значения, поэтому целесообразно для проверки не усвоенного знания ввести вероятность $\bar{A}(T_{\Pi})$. Величина $A(T_{\Pi})$ близка к единице, то есть при проверке усвоенного знания получается достоверный результат.

Ожидаемое время на проверку знаний определяется выражением

$$T_{\text{ож.п}} = \int_0^{\infty} t p(t) dt = \frac{\lambda_{\Pi}^\alpha}{\Gamma(\alpha)} \int_0^{\infty} t^\alpha e^{-\lambda_{\Pi} t} dt = \frac{\alpha}{\lambda_{\Pi}}, \quad (17)$$

где λ_{Π} - интенсивность проверки знаний.

Время T_{Π} , выделяемое на проверку знаний, берется по аналогии с (8) пропорционально ожидаемому времени $T_{\text{ож.п}}$, то есть

$$T_{\Pi} = \mu_{\Pi} T_{\text{ож.п}}, \quad (18)$$

где μ_{Π} - коэффициент пропорциональности при выделении времени на проверку знаний.

Качество проверки знаний определяется выражением

$$\bar{A}(T_{\Pi}) = \frac{1}{\Gamma(\alpha)} \int_0^{\mu_{\Pi} \alpha} z^{\alpha-1} e^{-z} dz. \quad (19)$$

Отношение дополнительного времени на проверку усвоенного материала к времени, выделенному на первоначальную проверку знаний, равно

$$\delta_{\Pi} = \frac{(\mu_{\Pi} \alpha)^\alpha}{\Gamma(\alpha + 1)} e^{-\mu_{\Pi} \alpha} + (1 - \mu_{\Pi}) [1 - A(\mu_{\Pi}, \alpha)]. \quad (20)$$

В основе описанной модели лежит гипотеза о гамма-распределении времени получения знаний. Проверка этой гипотезы в реальных условиях встречает определенные трудности.

В данной работе ограничимся простейшим процессом получения знаний, в основе которого лежит запоминание некоторой последовательности символов. Испытуемым предлагалось запоминание наборов символов с регистрацией времени запоминания. Карточка с наборами символов показывалась испытуемому, который после просмотра символов заявлял об их запоминании. После этого карточка переворачивалась, фиксировалось время T_i , и испытуемый пытался воспроизвести этот набор символов. В случае, когда воспроизведение выполнялось с ошибкой, карточка вновь предоставлялась испытуемому, и выделялось дополнительное время, в течение которого он вновь заявлял о готовности воспроизведения, и т.д.

Реализация подобного процесса не представляет особых трудностей и позволяет получить достаточно большой статистический материал. Результаты 100 экспериментов по запоминанию случайных наборов символов, получаемых программой генерации паролей из 5 символов, сведены в таблицу 1.

Таблица 1 - Время запоминания наборов символов

| № | T_i, c | № | T_i, c | № | T_i, c | № | T_i, c |
|----|----------|----|----------|----|----------|-----|----------|
| 1 | 23.1 | 26 | 8.4 | 51 | 33.7 | 76 | 9.8 |
| 2 | 11.4 | 27 | 8.0 | 52 | 9.0 | 77 | 9.9 |
| 3 | 9.3 | 28 | 7.6 | 53 | 22.0 | 78 | 3.5 |
| 4 | 9.1 | 29 | 20.4 | 54 | 10.2 | 79 | 10.1 |
| 5 | 8.3 | 30 | 15.7 | 55 | 17.2 | 80 | 8.9 |
| 6 | 15.2 | 31 | 8.1 | 56 | 13.7 | 81 | 13.2 |
| 7 | 23.0 | 32 | 9.1 | 57 | 9.8 | 82 | 19.4 |
| 8 | 24.2 | 33 | 12.4 | 58 | 4.7 | 83 | 11.4 |
| 9 | 8.5 | 34 | 12.0 | 59 | 10.2 | 84 | 12.9 |
| 10 | 7.6 | 35 | 7.6 | 60 | 7.2 | 85 | 11.0 |
| 11 | 4.9 | 36 | 9.1 | 61 | 9.1 | 86 | 15.6 |
| 12 | 8.0 | 37 | 18.8 | 62 | 3.6 | 87 | 3.9 |
| 13 | 8.7 | 38 | 10.2 | 63 | 12.0 | 88 | 19.1 |
| 14 | 19.2 | 39 | 11.4 | 64 | 7.7 | 89 | 4.5 |
| 15 | 9.6 | 40 | 11.8 | 65 | 4.1 | 90 | 11.5 |
| 16 | 20.7 | 41 | 9.2 | 66 | 5.1 | 91 | 5.6 |
| 17 | 11.5 | 42 | 4.1 | 67 | 21.1 | 92 | 11.3 |
| 18 | 8.7 | 43 | 4.4 | 68 | 12.1 | 93 | 3.6 |
| 19 | 13.7 | 44 | 3.4 | 69 | 8.2 | 94 | 9.1 |
| 20 | 4.1 | 45 | 4.7 | 70 | 13.3 | 95 | 10.1 |
| 21 | 13.5 | 46 | 9.1 | 71 | 12.7 | 96 | 18.0 |
| 22 | 15.4 | 47 | 17.7 | 72 | 7.7 | 97 | 8.1 |
| 23 | 19.1 | 48 | 8.3 | 73 | 9.6 | 98 | 4.3 |
| 24 | 7.7 | 49 | 25.7 | 74 | 5.4 | 99 | 7.6 |
| 25 | 13.5 | 50 | 17.9 | 75 | 21.2 | 100 | 13.2 |

В таблице 2 приведены частоты n_i попадания времен в частичные интервалы $t_i - t_{i+1}$.

Таблица 2 - Частоты попадания в частичные интервалы

| $t_i - t_{i+1}$ | n_i | $t_i - t_{i+1}$ | n_i |
|-----------------|-------|-----------------|-------|
| 0 - 5.0 | 14 | 15.0 - 20.0 | 13 |
| 5.0 - 10.0 | 37 | 20.0 - 25.0 | 8 |
| 10.0 - 15.0 | 26 | >25 | 2 |

Гистограмма частот приведена на рисунке 1.

Вид гистограммы напоминает кривую гамма-распределения. Оценку параметров гамма распределения будем производить исходя из выражений для математического ожидания и дисперсии, которые соответственно равны

$$M(t) = \frac{\alpha}{\lambda}, \quad D(t) = \frac{\alpha}{\lambda^2}. \quad (21)$$

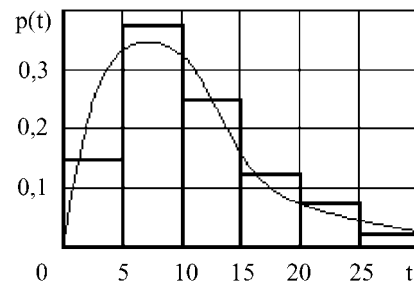


Рисунок 1 - Гистограмма частот времени запоминания

На основании данных, приведенных в таблице 1, получим следующие уравнения:

$$\frac{\alpha}{\lambda} = 11,413 \quad \text{и} \quad \frac{\alpha}{\lambda^2} = 33,221. \quad (22)$$

совместное решение которых дает значения

$$\lambda = 0,344 \text{ и } \alpha = 3,921. \quad (23)$$

Отсюда предполагаемая плотность распределения равна [3,4]

$$p(t) = \frac{0,344^{3,921}}{\Gamma(3,921)} t^{2,921} e^{-0,344t} = 0,0028 t^{2,921} e^{-0,344t}. \quad (24)$$

Для сравнения теоретического и экспериментального распределений воспользуемся критерием Пирсона [5]. Определим вероятности попадания значений времени в границы выделенных интервалов:

$$P(t_i \leq t < t_{i+1}) = \int_t^{t_{i+1}} p(t) dt. \quad (25)$$

Находим ожидаемое количество событий n_i , приходящихся на каждый интервал

$$n_i' = NP(t_i \leq t < t_{i+1}), \quad (26)$$

где N - количество проведенных наблюдений.

Определим значение $\chi_{\text{набл}}^2$, которое находится по формуле

$$\chi_{\text{набл}}^2 = \sum_i \frac{(n_i - n_i')^2}{n_i'}. \quad (27)$$

Результаты расчетов сведены в таблицу 3.

Таблица 3 - Оценка наблюдаемого $\chi_{\text{набл}}^2$

| $t_i - t_{i+1}$ | n_i | $P(t_i \leq t < t_{i+1})$ | n_i' | $\frac{(n_i - n_i')^2}{n_i'}$ |
|-----------------|-------|---------------------------|--------|-------------------------------|
| 0.0 - 5.0 | 14 | 0.104 | 10.4 | 1.246 |
| 5.0 - 10.0 | 37 | 0.363 | 36.3 | 0.013 |
| 10.0 - 15.0 | 26 | 0.302 | 30.2 | 0.584 |
| 15.0 - 20.0 | 13 | 0.149 | 14.9 | 0.242 |
| 20.0 - 25.0 | 8 | 0.057 | 5.7 | 0.928 |
| >25 | 2 | 0.035 | 2.5 | 0.100 |

$$\chi_{\text{набл}}^2 = 3,113$$

Число степеней свободы определяется выражением

$$k = i - r - 1 = 3, \quad (28)$$

где $i = 6$ - количество интервалов после объединения;

$r = 2$ - число параметров, оцениваемых при выборке.

Отсюда имеем $k = 3$.

Из таблицы критических точек распределения $\chi_{\text{кр}}^2$ по уровню значимости 0,05 и числу степеней свободы, рав-

ном 3, находим критическую точку правосторонней критической области:

$$\chi_{\text{кр}}^2(0,05, 3) = 7,8. \quad (29)$$

Так как $\chi_{\text{набл}}^2 < \chi_{\text{кр}}^2$, то нет оснований отвергать гипотезу о гамма-распределении и данные наблюдений согласуются с этой гипотезой.

Следует отметить, что особенности гамма-распределения нашли место, например, для описания творческой деятельности человека. Так, в работе [6] предложено теоретико-вероятностное описание научного труда, исходя из случайного распределения знаний и проблем в тезаурусе ученого. При этом делаются допущения, которые позволяют рассматривать порождение знаний и проблем как дискретный процесс Пуассона. После обобщения автор приходит к гамма-распределению, которое достаточно хорошо согласуется с экспериментальными данными науковедения о производительности научного труда.

Приведенные выше допущения могут быть положены в основу описания процесса обучения с помощью цепей Маркова.

Марковская модель обучения представлена следующими состояниями процесса обучения:

- состояние 1 - исходное состояние;
- состояние 2 - усвоение определенного объема знаний;
- состояние 3 - отсутствие удовлетворительного усвоения знаний;
- состояние 4 - результат усвоения знаний после проведения проверки знаний;
- состояние 5 - результат отсутствия удовлетворительного усвоения знаний после их проверки.

Если вероятности переходов между состояниями являются постоянными величинами, то такой процесс можно описать поглощающей цепью Маркова, так как существуют состояния 4 и 5, которыми завершается процесс обучения.

Структура матрицы переходов для поглощающей цепи Маркова имеет вид [7]:

$$P = \begin{bmatrix} Q & R \\ O & E \end{bmatrix}, \quad (30)$$

где Q - подматрица, описывающая поведение процесса до попадания в поглощающее состояние;

R - подматрица переходов в поглощающие состояния;

O, E - нулевая и единичные подматрицы.

Для процесса обучения 4 и 5 состояния являются поглощающими.

Матрицы Q и R соответственно равны:

$$Q = \begin{bmatrix} 0 & K_0 & 1 - K_0 \\ 1 - A & 0 & 0 \\ \bar{A} & 0 & 0 \end{bmatrix}; R = \begin{bmatrix} 0 & 0 \\ A & 0 \\ 0 & 1 - \bar{A} \end{bmatrix}, \quad (31)$$

где K_0 - вероятность усвоения знания;

A и \bar{A} - вероятности качественной проверки усвоенных и неусвоенных знаний соответственно.

Использование фундаментальной матрицы $N=(E-Q)$ позволяет получить ряд важнейших характеристик исследуемого процесса:

$$N = \frac{1}{1 - (1 - K_0)\bar{A} - K_0(1 - A)} \times \begin{bmatrix} 1 & K_0 & 1 - K_0 \\ 1 - A & 1 - (1 - K_0)\bar{A} & (1 - K_0)(1 - A) \\ \bar{A} & K_0\bar{A} & 1 - K_0(1 - A) \end{bmatrix}. \quad (32)$$

Элемент $n_{i,j}$ матрицы N дает ожидаемое количество моментов времени, которое проводит процесс в состоянии j до попадания в поглощающее состояние при условии, что он начался в состоянии i .

Матрица $B = NR$ позволяет оценить вероятность попадания в соответствующее поглощающее состояние:

$$B = \frac{1}{1 - (1 - K_0)\bar{A} - K_0(1 - A)} \times \begin{bmatrix} K_0A & (1 - K_0)(1 - \bar{A}) \\ [1 - (1 - K_0)\bar{A}]A & (1 - K_0)(1 - A)(1 - \bar{A}) \\ K_0\bar{A}A & [1 - K_0(1 - A)](1 - \bar{A}) \end{bmatrix}. \quad (33)$$

Для предлагаемой модели представляют интерес элементы $n_{1,1}$ и $b_{1,1}$ матриц N и B соответственно:

$$n_{1,1} = \frac{1}{1 - F}; \quad b_{1,1} = \frac{K_0A}{1 - F}, \quad (34)$$

где $F = (1 - K_0)\bar{A} + K_0(1 - A)$.

Значение элемента $b_{1,1}$ соответствует вероятности усвоения знаний или качеству знаний на выходе процесса обучения:

$$K = b_{1,1} = \frac{K_0A}{1 - F}. \quad (35)$$

Величина $n_{1,1}$ определяет ожидаемое количество интервалов времени, необходимых для усвоения материала.

Для оценки общего ожидаемого времени на усвоение материала рассмотрим вероятности следующих событий:

- вероятность завершения процесса обучения за выделенное время T_0 $G_0 = K_0A + (1 - K_0)(1 - \bar{A})$, где первое слагаемое определяет вероятность того, что материал усвоен и проверка усвоения дала положительный резуль-

тат, а второе слагаемое - это вероятность того, что материал не усвоен, а проверка не выявила это;

- вероятность выделения первого дополнительного времени на усвоение как события, противоположного завершению процесса за выделенное время $F_1 = F = 1 - G_0$;

- вероятность завершения процесса обучения после выделения первого дополнительного времени $G_1 = FG_0$;

- вероятность выделения второго дополнительного времени на усвоение $F_2 = F^2$;

- вероятность завершения процесса обучения после выделения второго дополнительного времени $G_2 = F^2G_0$; и т.д.

Таким образом, вероятность выделения j -го дополнительного времени на обучение определяется выражением

$$F_j = F^j, \quad (36)$$

а вероятность завершения процесса обучения после выделения j -го дополнительного времени на обучение получается умножением (36) на:

$$G_j = G_0F^j. \quad (37)$$

Для ограниченного количества n моментов времени, выделяемых на обучение, вероятность того, что процесс обучения будет завершен, определяется суммой

$$G_{(n)} = G_0 + G_0F + G_0F^2 + \dots = G_0 \prod_{j=1}^n F^j = G_0 \frac{1 - F^{n+1}}{1 - F}. \quad (38)$$

Предел этой вероятности при $n \rightarrow \infty$ равен единице:

$$G = \lim_{n \rightarrow \infty} \left[G_0 \frac{1 - F^{n+1}}{1 - F} \right] = \frac{G_0}{1 - F} = 1. \quad (39)$$

Проводя аналогичные рассуждения для вероятности усвоения материала, получим выражения, которые подобны (38) и (39). В этих выражениях G_0 будет заменено произведением K_0A , которое представляет собой вероятность того, что материал усвоен и проверка знаний дала положительный результат:

$$K_{(n)} = K_0A \frac{1 - F^{n+1}}{1 - F}, \quad (40)$$

при $n \rightarrow \infty$

$$K = \frac{K_0A}{1 - F}. \quad (41)$$

Последнее выражение совпадает с (35), полученным при рассмотрении обучения как марковского процесса.

Оценим время, затрачиваемое на обучение.

Затраченное на обучение время $T_{(n)}$ определяется суммой основного времени на обучение и n раз выделяемого

дополнительного времени:

$$T_{(n)} = T_0 + T_1 + \dots + T_i + \dots + T_n, \quad (42)$$

где T_i - выделяемое i -ое дополнительное время.

В соответствии с выражением (15) отношение $\frac{T_{i+1}}{T_i} = \delta$,

отсюда получим

$$T_i = \delta^i T_0 \quad (43)$$

и

$$T_{(n)} = T_0 \int_{i=0}^n \delta^i = T_0 \frac{1 - \delta^{n+1}}{1 - \delta}. \quad (44)$$

Ожидаемое время $\bar{T}_{(n)}$ при количестве выделяемых дополнительных моментов времени, не превышающих n , с учетом выражений (37) и (44), равно

$$\begin{aligned} \bar{T}_{(n)} &= \int_{i=0}^n T_{(i)} G_i = \frac{T_0 G_0}{1 - \delta} \int_{i=0}^n F^i (1 - \delta^{n+1}) = \\ &= \frac{T_0 G_0}{1 - \delta} \left[\frac{1 - F^{n+1}}{1 - F} - \delta \frac{1 - (\delta F)^{n+1}}{1 - \delta F} \right]. \end{aligned} \quad (45)$$

При неограниченном количестве выделяемых дополнительных интервалов времени получим значение ожидаемого времени \bar{T} на обучение:

$$\bar{T} = \lim_{n \rightarrow \infty} \left\{ \frac{T_0 G_0}{1 - \delta} \left[\frac{1 - F^{n+1}}{1 - F} - \delta \frac{1 - (\delta F)^{n+1}}{1 - \delta F} \right] \right\} = \frac{T_0}{1 - \delta F}. \quad (46)$$

Исходя из общих затрат времени на обучение при выделении n раз дополнительного времени (44) и ожидаемого значения времени на обучение \bar{T} (46), выражение для дисперсии D имеет вид

$$D = M \left\{ T_0 \frac{1 - \delta^{n+1}}{1 - \delta} - T_0 \frac{1}{1 - \delta F} \right\}^2, \quad (47)$$

где M - знак математического ожидания.

После возведения выражения в скобках в квадрат и вынесения $\frac{T_0 \delta}{1 - \delta}$ за знак математического ожидания, имеем

$$D = \left\{ \frac{T_0}{1 - \delta} \right\}^2 M \left[\left\{ \frac{1 - F}{1 - \delta F} \right\}^2 - 2 \frac{1 - F}{1 - \delta F} \delta^n + \delta^{2n} \right]. \quad (48)$$

Раскрывая выражение для математического ожидания, с учетом формулы (37), получим

$$\begin{aligned} D &= \left\{ \frac{T_0 \delta}{1 - \delta} \right\}^2 \left[\left\{ \frac{1 - F}{1 - \delta F} \right\}^2 G_0 \int_{i=0}^n F_i - 2 \frac{1 - F}{1 - \delta F} G_0 \int_{i=0}^{\infty} \delta F^i + \right. \\ &\quad \left. + G_0 \int_{i=0}^{\infty} (\delta^2 F)^i \right]. \end{aligned} \quad (49)$$

После суммирования членов геометрической прогрессии это выражение приобретает вид

$$\begin{aligned} D &= \left\{ \frac{T_0 \delta}{1 - \delta} \right\}^2 \left[\left\{ \frac{1 - F}{1 - \delta F} \right\}^2 \frac{G_0}{1 - F} - 2 \frac{1 - F}{1 - \delta F} \frac{G_0}{1 - \delta F} + \frac{G_0}{1 - \delta^2 F} \right] = \\ &= \frac{T_0^2 \delta^2 F (1 - F)}{(1 - \delta^2 F)(1 - \delta F)^2}. \end{aligned} \quad (50)$$

При рассмотрении T_0 как случайной величины с дисперсией D_0 , получение выражения для дисперсии описанным выше подходом представляет определенные трудности. Представим время (43) как результат произведения двух случайных величин - времени T_0 и коэффициента $\frac{1}{1 - \delta F}$, дисперсия которого может быть получена из выражения (50) после подстановки вместо T_0 единицы.

Формула для дисперсии произведения двух случайных величин, математические ожидания и дисперсии которых соответственно равны m_1, m_2, D_1 и D_2 , имеет вид

$$D = D_1 D_2 + m_1^2 D_2 + m_2^2 D_1, \quad (51)$$

откуда

$$D = (D_0 + T_0^2) \frac{\delta^2 F (1 - F)}{(1 - \delta^2 F)(1 - \delta F)^2} + D_0 \frac{1}{(1 - \delta F)^2}. \quad (52)$$

В полученных выражениях T_0 , представляет собой только время обучения. Учет времени проверки знаний производится по аналогичным формулам, однако, если выделение времени на обучение и проверку брат равным ожидаемым, то есть в выражениях (15) и (20) $\mu = \mu_n = 1$, то величина δ будет одинаковой как для обучения, так и для проверки знаний:

$$\delta = \frac{\alpha^\alpha}{\Gamma(\alpha + 1)} e^{-\alpha}. \quad (53)$$

Это позволяет использовать выражения (46), (50) и (52) для получения общих временных затрат. При этом T_0 будет представлять сумму ожидаемых затрат времени на обучение и проверку знаний, а D_0 - дисперсию этой суммы.

РЕЗУЛЬТАТЫ

Получены выражения для законов распределения времени на усвоение элементов знаний и некоторого объема знаний как совокупности этих элементов, которые описывается соответственно экспоненциальным законом (4) и гамма-распределением (5). На простейшем примере продемонстрировано согласование этого утверждения с экспериментальными данными.

Показано, что выделение времени на обучение пропорционально ожидаемому времени усвоения приводит к постоянству качества обучения (10) как для основного объема знаний, так и для случая, когда выделяется дополнительное время при отсутствии полного усвоения основного объема знаний. Это позволяет описывать обучение как марковский процесс (30, 31).

Получены выражения для вероятностей качественного обучения (35, 40, 41), времени ожидаемых затрат на обучение (46) и дисперсии этой величины (50, 52).

ВЫВОДЫ

Гипотеза о том, что вероятность получения элемента знаний в малом промежутке времени пропорциональна величине этого промежутка, приводит к экспоненциальному закону распределения времени на усвоение элемента знаний. Композиция экспоненциальных законов при постоянстве интенсивности обучения приводит к гамма-распределению времени на усвоение некоторого объема знаний как совокупности элементов знаний. К такому же результату приводит процесс проверки знаний.

Проведенный эксперимент по проверке гипотезы о гамма-распределении времени на запоминание определенного объема информации показывает, что полученные данные согласуются с этой гипотезой.

Выделение времени на обучение пропорционально

ожидаемому времени на усвоение знаний приводит к постоянству качества обучения и позволяет описывать его как марковский процесс, что может быть положено в основу построения математической модели процесса обучения.

Дальнейшие исследования в этом направлении должны включать себя определения условий описания процесса в виде цепей Маркова в случае, когда при выделении дополнительного времени на обучение изменяется параметр α , либо λ и одновременно, что повлечет за собой изменения выражений (9-20). Для достижения постоянства вероятностей переходов между состояниями процесса обучения необходимо будет определить закон изменения коэффициента μ в выражениях (8) и (14).

Использование цепей Маркова позволит построить модель обучения для каждой дисциплины и для специальности в целом, а затем перейти к модели учебного заведения, в которой влияние всех служб на качество обучения будет выражаться через параметры моделей дисциплин и специальностей.

ПЕРЕЧЕНЬ ССЫЛОК

1. Феллер В. Введение в теорию вероятностей и ее приложения. В 2-х томах. Пер. с англ. - М.: Мир, 1984. - 1266с.
2. Вентцель Е.С. Теория вероятностей. М.: Наука, 1964. - 576с.
3. Справочник по специальным функциям / Под ред. М. Абрамовица и И. Стиган. -М.: Наука, 1979. - 832с.
4. Янке Е., Эмде Ф., Леш Ф. Специальные функции. М.: Наука, 1977. - 344с.
5. Гмурман В.Е. Руководство к решению задач о теории вероятностей и математической статистике. М.: Высшая школа, 1975. - 336с.
6. Muller Fritz. Ein Versuch zur wahrscheinlichkeitstheoretischen Erklärung der wissenschaftlichen Produktivität. - Elektron. Informationsverarb. und Kybern., 1974, 10, №1, s.53-64.
7. Кемени Дж., Снелл Дж. Конечные цепи Маркова. М.: Наука, 1970. - 272с.

УДК 517.977.1

ЗАДАЧА КОММИВОЯЖЕРА И ГЕНЕТИЧЕСКИЕ АЛГОРИТМЫ

В.М.Гавриляко

Изучается генетический алгоритм решения задачи коммивояжера. Акцент делается на возможности его практического применения.

Вивчається генетичний алгоритм розв'язання задачі коммивояжера. Акцент робиться на можливості його практичного застосування.

The author study a Genetic Algorithm for solution of Traveling Salesman Problem. The emphasis is done on the possibility of its practical applications.

В данной работе рассматривается известная задача оптимизации - задача коммивояжера (ЗК), которая принадлежит к числу NP - полных задач, являющихся трудно-разрешимыми с вычислительной точки зрения. Известно,

что время нахождения точного решения путем перебора всех возможных вариантов имеет астрономический порядок. Поэтому один из возможных подходов к практическому решению этой задачи состоит в использовании эвристических алгоритмов. В этом случае, как известно, отсутствуют привычные обоснования применимости метода. "Правдоподобные рассуждения, наша интуиция, опыт и машинный эксперимент - вот пока те оправдания эвристики, которыми мы располагаем" [1]. Изучаемый в работе эвристический алгоритм позволяет найти, вообще говоря, локальный минимум в ЗК с помощью моделирования на компьютере законов генетики и естественного отбора Ч. Дарвина. Впервые идея этого алгоритма была предложена Дж. Холландом [2] и названа им "генети-

ческий алгоритм (ГА)". ГА является алгоритмом прямого поиска (методом оптимизации на основе сопоставления значений функций в пробных точках). В настоящее время с помощью генетического алгоритма решен ряд практических задач [2,4,5,7,8].

Основная концепция классической генетики - ген - была введена Менделем с целью объяснения наблюдаемой статистики наследования [3]. Носителями генов в клеточном ядре являются нитевидные тела, называемые хромосомами. Место, занимаемое геном в хромосоме, носит название локус. Схематически можно представить себе хромосому как прямолинейный отрезок, а локусы - как последовательные участки на которые этот отрезок разбит. Статистически действия генов проявляются в достаточно больших сообществах организмов, скрещивающихся между собой. Такие сообщества биологи называют популяциями. Популяция существует не только в пространстве, но и во времени. Основой этого существования является репродукция путем скрещивания. Часто можно считать, что вся популяция состоит в пространстве и времени из дискретных поколений F_0, F_1, F_2, \dots . Поколение F_{i+1} - это совокупность особей, родители которых принадлежат поколению F_i ; поколение F_0 фиксируется априори. Мы будем рассматривать такие популяции, у которых для любого состояния z в поколении F однозначно определено состояние z' , возникающее в следующем поколении F' путем скрещивания и отбора. В данной работе понятия "особь, организм" мы отождествляем с понятием "хромосома". Отображение V , определяемое уравнением $z' = Vz$, называют эволюционным оператором [3]. Обычно требуют, учитывая биологический смысл эволюционного оператора, непрерывность отображения V , для того, в частности, чтобы обеспечить устойчивость при замене вероятностей близкими к ним частотами. Тогда рассматриваемую популяцию можно считать динамической системой на компактном подмножестве конечномерного пространства [3]. Неподвижные точки эволюционного оператора V называют равновесными состояниями популяции. Они существуют в силу известной топологической теоремы Брауэра. Генетический алгоритм позволяет находить эти состояния популяции: "Работу алгоритма прекращаем при достижении популяцией состояния адаптации, идентифицируемому по стягиванию ядра популяции сначала в плотное облачко, а затем в точку. Кроссинговер как механизм изменчивости теряет в таких условиях свою силу - при скрещивании идентичных родителей потомок ничем не будет отличаться ни от одного из них." [5]. По сути дела этот эвристический алгоритм опирается на используемую в динамике отбора восходящую к R. A. Fisher [6] фундаментальную теорему о возрастании средней приспособленности популяции в ходе естественного отбора. Если определить для каждой популяции функционал качества (фитнес), то неподвижные точки эволюционного оператора будут соответствовать точкам его локального экстремума. К сожалению, нельзя с уверенностью утверждать, что найденное решение дает глобальный экстремум.

В случае ЗК основные понятия генетического алгоритма имеют следующий смысл: ген - это номер вершины графа городов (номер города), через который проходит маршрут коммивояжера; хромосома - последовательность ген, определяющих замкнутый маршрут, проходящий через все заданные города (вершины графа); популяция - набор из нескольких возможных маршрутов. Состояние популяции - множество возможных маршрутов, удовлетворяющих условиям задачи. Функционал качества - длина маршрута.

На эффективность метода существенно влияет способ задания первой популяции, т.е. систематическое перечисление вершин графа. Нами реализован так называемый поиск в ширину, при котором вершины обходятся в следующем порядке: выбирается случайный город в качестве точки старта. Он рассматривается как текущий. Затем сравниваются длины путей, ведущих из текущего города в еще не посещенные города, и из них выбирается кратчайший путь. Полученный таким образом следующий город становится текущим. Эта процедура повторяется до тех пор, пока все города не будут перебраны. Если брать различные точки старта, то можно получить до n различных маршрутов (хромосом). Остальные хромосомы (до заданного числа M) заполняем случайными маршрутами.

В теории ГА существенным является выбор кодировки чисел (номеров городов). Для задачи коммивояжера кодировка выбрана десятичной. Эффективность применения ГА к ЗК зависит также и от настройки операторов мутации и кроссинговера. Мутация, как незначительное (точечное) случайное изменение в хромосоме может осуществляться несколькими способами. Наиболее распространенным является выбор двух случайных ген и их обмен местами [2,4,5,7]. Нами была использована другая схема: если в хромосоме случайным образом выбран ген с номером i , то он меняется местом с геном с номером $n+1-i$. Цель оператора мутации - расширение пространства поиска за счет случайных точечных изменений в хромосоме. В классическом варианте [4] действие оператора кроссинговера состоит в следующем: выбираются две определенные хромосомы (случайным образом), далее также случайным образом определяется точка кроссинговера и происходит обмен участками хромосом.

Пример

| | | |
|-----------|---|-----------|
| 3 2 1 4 5 | { | 3 2 4 1 3 |
| 5 2 4 1 3 | { | 5 2 1 4 5 |

Однако полученные хромосомы не всегда являются допустимыми маршрутами коммивояжера (как в данном примере). Возможный вариант осуществления кроссинговера - брать не две хромосомы, а одну. В этом случае действие оператора кроссинговера будет состоять в повороте участка хромосомы либо в его циклической перестановке. Нами был выбран двухточечный кроссинговер, при этом последовательность чисел между этими двумя точками "поворачивалась на 180 градусов".

Пример:

| | | | | |
|---------|--|---------|--|------------|
| 3 2 1 6 | | 4 7 8 5 | | 9 12 10 11 |
| 3 2 1 6 | | 5 8 7 4 | | 9 12 10 11 |

На компьютере был реализован следующий алгоритм:

Шаг 1. Задается размер популяции M , вероятность кроссинговера $P_{\tilde{n}}$, вероятность мутации P_m , определяется критерий конца вычислений: количество итераций N (т.е. количество популяций P). Шаг 2. Задается начальная популяция $P(0)$. Шаг 3. $k=0$.

Шаг 4. К каждой хромосоме популяции $P(k)$ применяются операторы кроссинговера и мутации. Шаг 5. $k=k+1$. Шаг 6. Если $k < N$ то возвращаемся к шагу 4, иначе

Шаг 7. Вывод на печать результатов вычислений

Приведем результаты некоторых численных экспериментов. На рис.1 приведен пример маршрута коммивояжера по 100 городам.

```
TSP report for current task. Population Size =400
First population The ending populations
Bad travel: 273759 50695
Average travel: 251 076 50684
Best travel: 221144 49696
Using time: 374368 mlsec
equal 49696 km
```

На рис.2 представлен результат работы этой же программы для 11 городов.

В левой части рис.2 представлен маршрут коммивояжера, определенный генетическим алгоритмом, в правой части этот же маршрут найден глобальным перебором всех вариантов. В данном случае с помощью генетического алгоритма найден глобальный минимум и при этом за существенно меньшее время.

Генетический алгоритм может работать в двух режимах - обычный (по вышеописанной схеме) и "жадный". "Жадный" отличается тем, что в процессе кроссинговера и мутации "выживают" только те хромосомы, для которых применение этих операторов привело к некоторому улучшению целевой функции. При этом размер популяции уменьшается на каждой итерации. Применение "жадного" алгоритма приводит к более быстрой сходимости, но популяция в этом алгоритме быстро вырождается и обычные меры для расширения пространства поиска в виде операторов мутации не помогают. В этом варианте наблюдается быстрая сходимость к локальному минимуму.

Интересен крайний случай, когда популяция состоит только из одной особи (рис.3-5).

На рис.3 представлено расположение 35 городов, на рис. 4 итоговый маршрут коммивояжера, на рис. 5 представлен график изменения функционала качества (фитнес).

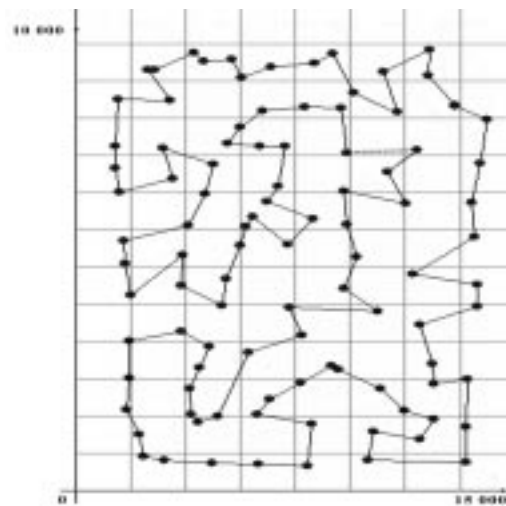


Рисунок 1

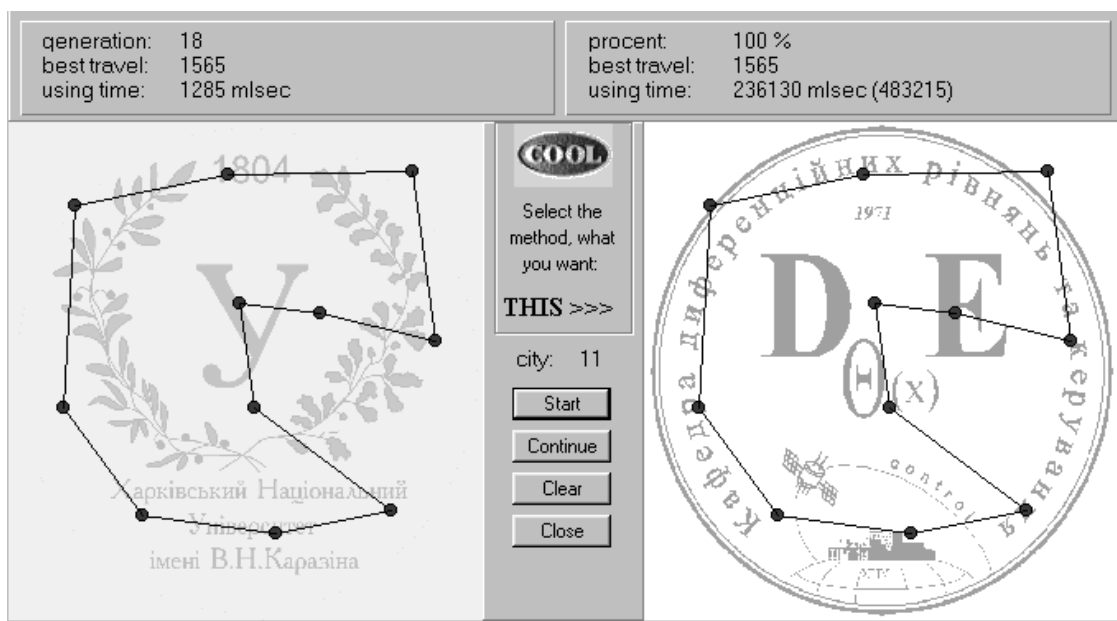


Рисунок 2

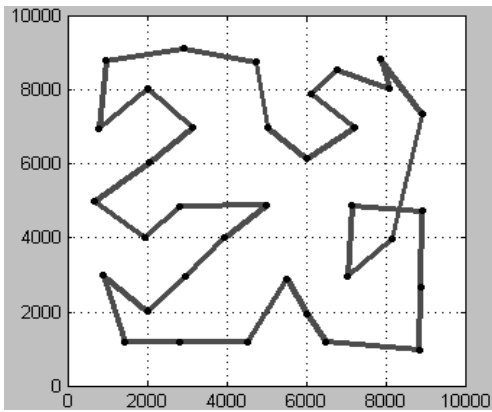


Рисунок 3

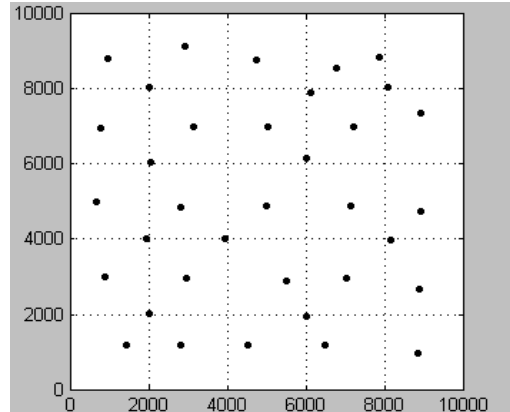


Рисунок 4

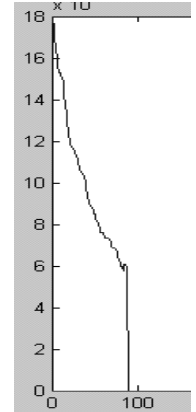


Рисунок 5

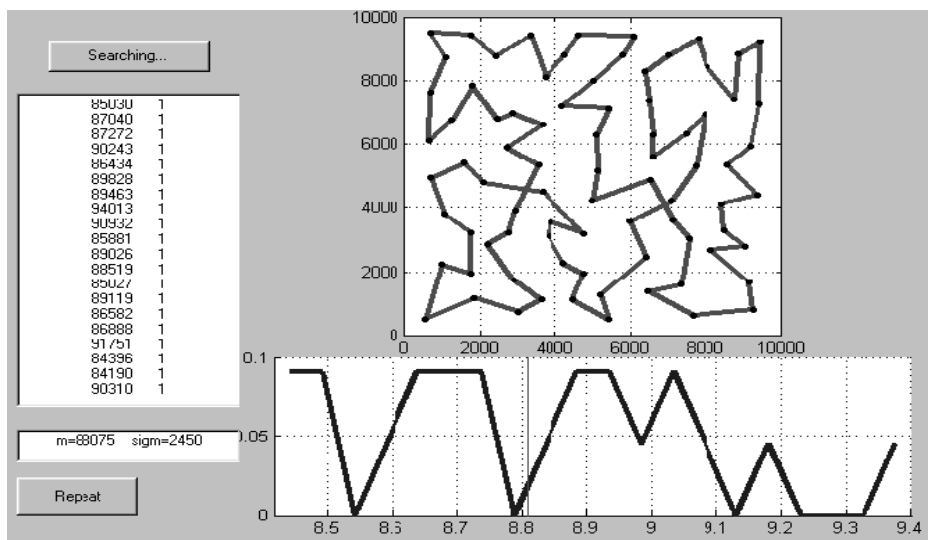


Рисунок 6

Рассмотренные генетические алгоритмы могут быть условно разбиты на два класса: на работающие с множеством решений (популяцией, объем которой больше одной хромосомы) и на работающие с одной хромосомой. Второй класс алгоритмов имеет более высокую скорость сходимости, но имеет меньше шансов нахождения глобального экстремума функционала качества, так как менее способен к выходу из локальных экстремумов.

Выполняя расчеты с помощью генетического алгоритма заданное число раз с одним и тем же набором городов, мы имеем возможность выбрать минимальный из локальных минимумов. Как показывают эксперименты, при $n < 20$ достаточно 5-6 запусков программы, для того чтобы с 95-процентной вероятностью получить глобальный минимум. Однако при увеличении количества городов вопрос о точности решения снимается и возникает вопрос о величине погрешности генетического алгоритма. К сожалению, данная проблема практически не поддается точному анализу. Результаты экспериментов (размер популяции равен одной хромосоме) приведены также на рис.6, где размещены элементы контроля и исследуемый набор из 80

городов. По результатам испытаний построена гистограмма и найдено математическое ожидание и среднее квадратическое отклонение. Математическое ожидание в данном случае равно 88075, а среднее квадратическое отклонение равно 2450, что составляет 2.78% от математического ожидания. Для городов, количество которых не превышает 50, большое количество экспериментов показало, что среднее квадратическое отклонение не превосходит 5% математического ожидания. Это дает возможность оценить погрешность результатов генетического алгоритма как практически хорошую. Метод генетических алгоритмов позволил нам также численно решить ряд принципиально разных по своей природе прикладных задач: найти решение системы негладких уравнений, возникающих в задачах оптимального управления при использовании принципа максимума Понтрягина [8], решить некоторые задачи целочисленного программирования. Однако ГА, как и другие методы эволюционных вычислений, не гарантируют обнаружение глобального экстремума за полиномиальное время. Эти алгоритмы хороши для получения "достаточно хорошего" решения задачи за "достаточно хорошее" время.

ПЕРЕЧЕНЬ ССЫЛОК

1. Моисеев Н.Н. Элементы теории оптимальных систем. М.: Наука, 1975. - 526 с.
2. Holland J.H. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Application to Biology, Control and Artificial Intelligence. University of Michigan, 1975.
3. Любич Ю.И. Математические структуры в популяционной генетике. К.: Наук.думка, 1983. - 296 с.
4. Goldberg D.E. Genetic Algorithms in Search, Optimization and Machine Learning. Addison - Wesley Publishing Company, Inc., 1989.

5. Вороновский Г.К., Махотило К.В., Петрашев С.А., Сергеев С.А. Генетические алгоритмы, искусственные нейронные сети и проблемы виртуальной реальности. Харьков: Основа, 1997. - 112 с.
6. Fisher R.A. The genetical theory of natural selection. -Oxford: Clarendon press, 1930. - 362 p.
7. Батищев Д.И. Генетические алгоритмы решения экстремальных задач. Воронеж: ВГТУ, 1995
8. Гавриляко В.М., Соколов А.В. Численное решение негладких уравнений в задаче быстрогодействия // Вісник Харківського університету, сер. Математика, прикладна математика і механіка. - 2002. - № 542. - С.145 - 148.

УДК 004.518

РАЗРАБОТКА И ИССЛЕДОВАНИЕ ИЕРАРХИЧЕСКИХ АГЛОМЕРАТИВНЫХ АЛГОРИТМОВ РАЙОНИРОВАНИЯ

А.В.Заварзин, В.И.Месюра

В статье рассматривается задача районирования, т.е. задача классификации многомерных пространственных данных при учете ограничений на географическую смежность получаемых классов-районов. Приводятся классические агломеративные методы иерархической классификации, для которых разработаны и реализованы их аналоги - агломеративные алгоритмы районирования.

На данных в разрезе субъектов Российской Федерации (89 объектов) и в разрезе избирательных округов Российской Федерации (порядка 2500 объектов) проведены многочисленные эксперименты. Исследованы оптимальные варианты реализации и ход районообразования агломеративных иерархических алгоритмов районирования.

In this article the problem of regionalization is examined, i.e. the problem of classification of the multivariate spatial data with the conditions on a geographical contiguity of classes (regions). Classical agglomerative methods of hierarchical classification are given. Agglomerative algorithms of regionalization are developed and realized for the classical methods .

The number of experiments are carried out on the attributive data of subjects of the Russian Federation (89 objects) and of constituencies of the Russian Federation (about 2500 objects). Optimum ways of realization and interim classes (regions) properties of agglomerative hierarchical regionalization algorithms are investigated.

ПОСТАНОВКА ЗАДАЧИ РАЙОНИРОВАНИЯ

Объектом классификации в географии является ОТЕ - операционно-территориальная единица (Трофимов и др., 1985, стр. 13). В качестве ОТЕ могут выступать, например, административно-территориальные единицы, населенные пункты, ячейки регулярной или нерегулярной сетки, наложенной на исследуемую территорию. Обозначим все множество ОТЕ символом $O = \{o_1, \dots, o_N\}$, где o_i - i -я ОТЕ, N - количество ОТЕ.

Целью классификаций в географии является получение некоторого заранее заданного или не заданного количества групп ОТЕ (классов ОТЕ). В пределах каждого класса ОТЕ должны быть максимально "похожи" друг

на друга в некотором смысле, однородны, а ОТЕ из разных классов - максимально "отличаться".

Будем называть j -м классом подмножество ОТЕ

$$S_j = \{o_{ij} \in O | j \in \{1, \dots, N\} \subset O,$$

где $N_j = |S_j| \leq N$ - количество ОТЕ i -го класса.

Под системой классов, получение которой является целью классификации, будем понимать множество S со следующими свойствами:

1. $S = \{S_i | i \in \{1, \dots, K\}\}$, т.е. S состоит из K классов.

2. $S_i \cap S_j = \emptyset \quad \forall i, j \in \{1, \dots, K\}$, где \emptyset - пустое

множество.

3. $S_i \neq \emptyset$ т.е. $|S_i| > 0 \quad \forall i \in \{1, \dots, K\}$.

$$4. \bigcup_{i=1}^K S_i = O, \quad \bigcap_{i=1}^K N_i = N.$$

В географическом пространстве ОТЕ описываются различными показателями и метриками, основными из которых можно считать способы расчета расстояния между ОТЕ (т.е. коэффициентов "сходства" или "отличия" ОТЕ). Помимо географического пространства, исследуемая совокупность ОТЕ фиксирована и в пространстве M атрибутивных показателей (или пространстве M атрибутивных признаков). В этом пространстве ОТЕ теряют свою географичность и независимо от своей первоначальной природы становятся M -мерными точками.

Алгоритмы районирования направлены на получение территориально (географически) нерасчлененных районов (классов), выделяемых по критерию их несхожести в при знаковом пространстве с учетом выбранной метрики или матрицы близостей ОТЕ-ОТЕ. Многие алгоритмы районирования, с одной стороны, практически полностью дублируют алгоритмы классификации с добавлением дополнительной процедуры проверки условия на географи-

ческую смежность районов. С другой стороны, существуют специфические алгоритмы районирования, для которых отсутствуют аналоги в классификации (например, метод барьеров максимальных различий, описанный в работе Monmonier, 1973).

КЛАССИЧЕСКИЕ ИЕРАРХИЧЕСКИЕ АГЛОМЕРАТИВНЫЕ АЛГОРИТМЫ КЛАССИФИКАЦИИ

Классический агломеративный алгоритм иерархической классификации начинает свою работу с формирования $K_1 = N$ классов (при этом каждая ОТЕ на нулевом шаге представляет отдельный класс) и проводит в общем случае $I = N - 1$ итерацию. На каждом шаге алгоритма происходит объединение двух "ближайших" классов в один, т.е. $K_n - 1 = K_{n+1}$. Последний $(N - 1)$ -й шаг алгоритма характеризуется объединением двух сформированных на предыдущих этапах классов в один класс, включающий в себя все имеющиеся (поступившие на вход анализа) ОТЕ. Выбор расстояния настолько влияет на результат классификации, что зачастую оно вносится в название алгоритма (например, "агломеративный алгоритм средней связи").

Способы вычисления степени "близости" классов (расстояния между классами D) иногда называются стратегиями объединения классов. Пусть $D: S \times S \rightarrow R$ - функция расстояния между классами, $S_i = \{o_{i1}, \dots, o_{iN_i}\}$ - i -й класс, N_i - число ОТЕ в i -м классе, $i \in \{1, \dots, K\}$.

Ниже приведены наиболее известные виды расстояния между классами.

1. Метод ближнего соседа. Расстояние между двумя классами рассчитывается как расстояние между двумя ближайшими ОТЕ этих двух классов:

$$D_{min}(S_i, S_j) = \min\{d(o_{ix}, o_{jy}) | o_{ix} \in S_i, o_{jy} \in S_j\}.$$

В качестве недостатка метода можно отметить тот факт, что при наличии в выборке из N ОТЕ аномальных наблюдений (т.е. таких, которые существенно отличаются по своим значениям показателей от остальных ОТЕ), они будут помещены в отдельные классы. Основная же группа ОТЕ "сольется" в один большой класс. В литературе по геоинформатике (Тикунов, 1997) предлагается учитывать в вычислении расстояния по методу ближнего соседа количества ОТЕ в каждом из классов:

$$D_{min}(S_i, S_j) = (N_i + N_j) \times \min\{d(o_{ix}, o_{jy}) | o_{ix} \in S_i, o_{jy} \in S_j\}.$$

Такое расстояние гораздо более устойчиво к образованию одного многочисленного класса в агломеративной процедуре, т.к. значительно "подправляет" в сторону увеличения расстояния между "большими" классами, что не дает им объединяться.

Важной разновидностью расстояния между классами по методу ближнего соседа является географическая смежность между классами g^S , используемая в алгоритмах ра-

ионирования. Как и в случае атрибутивного пространства, это расстояние основано на смежности отдельных ОТЕ классов. Если в двух классах существует хотя бы по одной смежной друг с другом ОТЕ, то (следуя принципу ближнего соседства) два класса также смежны:

$$g^S: S \times S \rightarrow \{0, 1\},$$

$$g^S(S_i, S_j) = \begin{cases} 1, & \exists x \in \{1, \dots, N_i\}, \exists y \in \{1, \dots, N_j\}: g(o_{ix}, o_{jy}) = 1, \\ 0, & \forall x \in \{1, \dots, N_i\}, \forall y \in \{1, \dots, N_j\}: g(o_{ix}, o_{jy}) = 0. \end{cases}$$

2. Метод дальнего соседа. Расстояние между двумя классами рассчитывается как расстояние между двумя самыми дальними ОТЕ этих классов:

$$D_{max}(S_i, S_j) = \max\{d(o_{ix}, o_{jy}) | o_{ix} \in S_i, o_{jy} \in S_j\}.$$

Данный метод более устойчив к аномальным наблюдениям при использовании в агломеративном алгоритме. Кроме того, получаемые с его помощью классы обычно соразмеримы (т.е. число ОТЕ в них примерно одинаково).

3. Центроидный метод. Расстояние между двумя классами рассчитывается как расстояние между центрами классов:

$$D_{cen}(S_i, S_j) = d(\bar{o}_i, \bar{o}_j),$$

$$\text{где } \bar{o}_i = \frac{1}{N_i} \int_{x=1}^{N_i} o_{ix} = \frac{1}{N_i} \int_{x=1}^{N_i} o_{ix}^{(1)}, \dots, \frac{1}{N_i} \int_{x=1}^{N_i} o_{ix}^{(M)}.$$

4. Метод группового среднего (средней связи). Расстояние между двумя классами рассчитывается как среднее расстояние между ОТЕ двух классов:

$$D_{avg}(S_i, S_j) = \frac{1}{N_i N_j} \int_{x=1}^{N_i} \int_{y=1}^{N_j} d(o_{ix}, o_{jy}).$$

Если число классов K , которые нужно получить, известно заранее, достаточно провести $I = N - K$ итераций, в результате которых и будет сформировано ровно K классов. Если количество классов заранее неизвестно, анализируются либо значения функционала качества разбиения для $K \in \{2, \dots, K_{max}\}$, либо применяются другие методы (см., например, метод анализа сложности группировочного дерева в работе Бланца, 1993, стр. 94). Информацию о количестве классов может дать и визуальный анализ дендрограммы.

Ниже приведена общая схема работы агломеративного алгоритма классификации.

1. Инициализировать начальные значения алгоритма.

1.1) Задаться метрикой d и стратегией объединения D .

1.2) Сформировать первую систему S^1 из N классов:

$$S^1 = \left\{ S_1^1, \dots, S_N^1 \right\}, \quad S_1^1 = \{o_1\}, \dots, S_N^1 = \{o_N\}.$$

Верхний индекс в обозначении класса будет указывать

шаг алгоритма n , нижний - номер класса в текущей системе.

1.3) Положить $n = 1$.

2. Пусть на шаге $n \in \{1, \dots, N-1\}$ получена система классов

$$S^n = \{S_1^n, \dots, S_K^n\}, K = N - n + 1. \text{ Тогда:}$$

2.1) Вычислить расстояния между классами

$$D(S_i^n, S_j^n) \quad \forall i, j \in \{1, \dots, K\}.$$

2.2) Найти $x \neq y, x, y \in \{1, \dots, K\}$:

$$D(S_x^n, S_y^n) = \min \left\{ D(S_i^n, S_j^n) \mid i \neq j \right\}.$$

2.3) Не ограничивая общности $i < j$. Положить

$$S_i^{n+1} = \begin{cases} S_i^n, & i \in \{1, \dots, x-1\} \cup \{x+1, \dots, y-1\}, \\ S_x^n \cup S_y^n, & i = x, \\ S_{i-1}^n, & i \in \{y+1, \dots, K\}, \end{cases}$$

сформировав, таким образом, новую систему классов

$$S^{n+1} = \{S_1^{n+1}, \dots, S_{K-1}^{n+1}\}.$$

3. Если $n = N - 1$, конец, иначе положить $n = n + 1$ и перейти к шагу 2.

Детальное описание процедур агломеративных иерархических классификаций можно найти в работе (Жамбю, 1989).

АДАПТАЦИЯ КЛАССИЧЕСКИХ ИЕРАРХИЧЕСКИХ АГЛОМЕРАТИВНЫХ АЛГОРИТМОВ КЛАССИФИКАЦИИ ДЛЯ ЦЕЛЕЙ РАЙОНИРОВАНИЯ

Во многих работах по геоинформатике встречаются упоминания об использовании для целей районирования иерархических агломеративных алгоритмов классификации. Автору, однако, не удалось найти описание обобщенного алгоритма районирования и экспериментов по исследованию свойств алгоритмов для каждого типа межклассового расстояния.

Отличие агломеративных алгоритмов районирования от соответствующих алгоритмов классификации состоит в формировании матрицы географической смежности и проверке на каждом шаге граничности объединяемых районов. Для классических агломеративных алгоритмов это соответствует условию нахождения минимума расстояния между районами (или максимум функции качества при объединении) только для географически смежных районов.

Примем следующую схему работы обобщенного агломеративного алгоритма районирования.

1. Задаться метрикой d и стратегией объединения D . "Обобщенность" алгоритма относится к возможности использования произвольной стратегии объединения D .

2. Сформировать первую систему S^1 из N районов:

$$S^1 = \{S_1^1, \dots, S_N^1\}, S_1^1 = \{o_1\}, \dots, S_N^1 = \{o_N\}.$$

Верхний индекс в обозначении районов будет указывать шаг алгоритма, нижний - номер района в текущей системе. Положить $n = 1$.

3. Положить $K = N - n + 1$ и получить матрицу смежности G_{KK}^n :

$$g^n(i, j) = \begin{cases} 1, & \exists o_{ix} \in S_i^n, \exists o_{jy} \in S_j^n : (o_{ix}, o_{jy}) = 1; \\ 0, & \text{иначе} \end{cases}$$

$i, j \in \{1, \dots, K\}$.

4. Пусть на шаге $n \in \{1, \dots, N-1\}$ получена система районов

$$S^n = \{S_1^n, \dots, S_K^n\}. \text{ Тогда:}$$

4.1) Вычислить расстояния между районами

$$D(S_i^n, S_j^n) \quad \forall i, j \in \{1, \dots, K\}.$$

4.2) Найти $x \neq y, x, y \in \{1, \dots, K\}$:

$$D(S_x^n, S_y^n) = \min \left\{ D(S_i^n, S_j^n) \mid i \neq j, g^n(i, j) = 1 \right\}.$$

4.3) Не ограничивая общности $i < j$. Положить

$$S_i^{n+1} = \begin{cases} S_i^n, & i \in \{1, \dots, x-1\} \cup \{x+1, \dots, y-1\}, \\ S_x^n \cup S_y^n, & i = x, \\ S_{i-1}^n, & i \in \{y+1, \dots, K\}, \end{cases}$$

сформировав, таким образом, новую систему районов

$$S^{n+1} = \{S_1^{n+1}, \dots, S_{K-1}^{n+1}\}.$$

5. Если $n + 1 = N$, конец, иначе положить $n = n + 1$ и перейти к шагу 3.

Вполне может получиться, что для $n < N$ на шаге 4.2 не удастся найти два смежных класса, которые требуется объединить. Эта ситуация соответствует случаю несвязности графа G географических расстояний, построенного на основе матрицы смежности G_{NN}^1 . Если c - число компонент связности графа G , то критерием останова алгоритма на 5-м шаге является сравнение n с числом $(N - c + 1)$.

ПРОСТРАНСТВЕННАЯ СТРУКТУРА РАЙОНОВ, ПОЛУЧАЕМЫХ РАЗЛИЧНЫМИ АГЛОМЕРАТИВНЫМИ АЛГОРИТМАМИ РАЙОНИРОВАНИЯ

В результате проведения многочисленных экспериментов на данных в разрезе субъектов Российской Федерации (89 объектов) и в разрезе избирательных округов Российской Федерации (порядка 2500 объектов) исследована зависимость мощности образуемых классов от номера шага алгоритма при учете ограничений на пространственную нерасчлененность классов и (для сравнения) без учета таких ограничений.

Номер шага n агломеративного алгоритма классификации или районирования напрямую связан с количеством получаемых на данном шаге классов или районов K соотношением

$$K = N - n + 1.$$

Обратимся к эксперименту классификации и районирования с помощью агломеративных алгоритмов 89-ти субъектов Российской Федерации. В качестве атрибутивных данных использовались пять признаков, отражающих здоровье населения России (Прохоров, 1996): уровень младенческой смертности, среднеожидаемая продолжительность жизни мужчин, среднеожидаемая продолжительность жизни женщин, стандартизированный коэффициент смертности мужчин, стандартизированный коэффициент смертности женщин.

На первых шагах все агломеративные алгоритмы дают примерно одинаковые результаты. Серьезные расхождения возникают на итерациях $n > \frac{N}{2}$.

Для примера рассмотрим итерацию

$$n = 67 = N - \left\lceil \frac{N}{4} \right\rceil = 89 - \left\lceil \frac{89}{4} \right\rceil$$

и продемонстрируем следующие результаты:

- алгоритмы районирования тяготеют к более неравномерному распределению объектов по районам, нежели их аналоги в классификации;
- наиболее нестабильным алгоритмом районирования является метод ближайшего соседа, тяготеющий к образованию одного представительного района и порядка $K = N - n$ районов, содержащих по одному два объекта;
- наилучшие результаты показывают алгоритмы дальнего соседа и взвешенного (количеством объектов в объединяемых классах) ближнего соседа, которые позволяют получать примерно соразмеримые районы.

Примем следующие обозначения алгоритмов на графиках:

- × - Дальний сосед
- ⊖ - Центроидный
- + - Групповое среднее
- * - Ближний X количество

Классические агломеративные алгоритмы на приведе-

нных выше данных на 67-м шаге формируют 23 класса объектов. Будем считать, что классы заданы их номерами и отсортированы в порядке возрастания количества объектов. Отбросив 4 класса, содержащих для всех алгоритмов по одному объекту, получим следующее распределение объектов по классам:

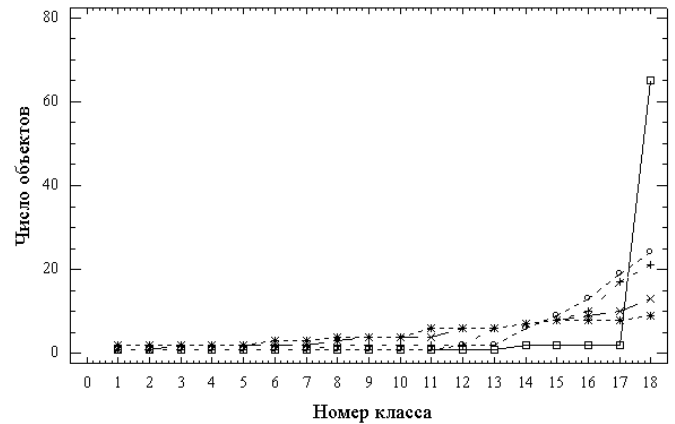


Рисунок 1 - Мощности классов для алгоритмов классификации

Как следует из графика, алгоритмы можно отсортировать в соответствии с соразмерностью получаемых классов следующим образом (от худшего к лучшему):

- алгоритм ближнего соседа;
- центроидный алгоритм и алгоритм группового среднего;
- метод дальнего соседа и видоизмененный алгоритм ближнего соседа.

Следующий график демонстрирует результаты, полученные при наложении условий на географическую смежность. Очевидно, что все алгоритмы, кроме метода ближнего соседа, потеряли в соразмерности классов. Результат работы метода ближнего соседа остался неизменно неудовлетворительным.

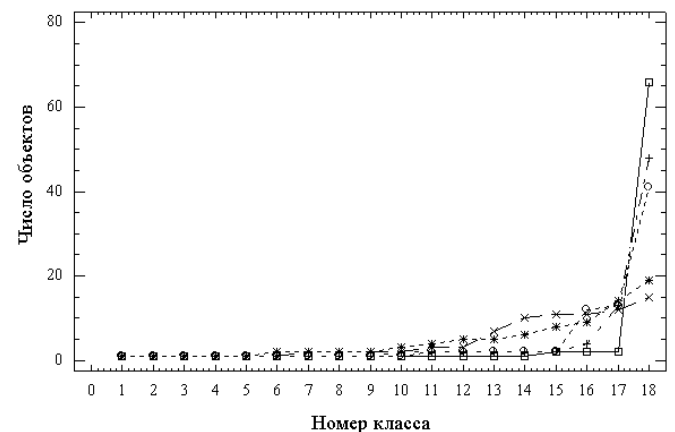


Рисунок 2 - Мощности районов для алгоритмов районирования

Естественно, что на разных данных (картографических и атрибутивных) получаются различные результаты, однако их различие невелико. В среднем приведенный выше пример иллюстрирует общие закономерности. Можно, однако, подобрать примеры хорошо разделимых районов как в пространстве признаков, так и в географическом пространстве, для которых все методы будут давать сходные результаты.

ОПТИМАЛЬНАЯ РЕАЛИЗАЦИЯ ИЕРАРХИЧЕСКИХ АГЛОМЕРАТИВНЫХ АЛГОРИТМОВ РАЙОНИРОВАНИЯ

Наиболее трудоемким этапом алгоритма районирования является пересчет расстояний между районами на каждом шаге n :

$$D(S_i^n, S_j^n) \forall i, j \in \{1, \dots, K\}.$$

При реализации алгоритмов можно рассмотреть два альтернативных подхода. Первый анализируемый подход основан на использовании формулы Жамбю. Эта формула позволяет вычислять новые расстояния между классами на основе уже посчитанных. Для расчета расстояний от вновь полученного (например, объединением S_x и S_y) класса до произвольного класса S_i необходимо произвести следующие вычисления (Айвазян, Мхитарян, 2001, стр. 497):

$$D(S_i, S_x \cup S_y) = \alpha D(S_i, S_x) + \beta D(S_i, S_y) + \gamma D(S_x, S_y) + \delta |D(S_i, S_x) - D(S_i, S_y)|. \quad (1)$$

При различных значениях параметров $\alpha, \beta, \gamma, \delta$ эта формула соответствует вычислению наиболее часто используемых межклассовых расстояний. Значения коэффициентов для каждого вида расстояния известны.

Реализация классических агломеративных алгоритмов классификации в статистических пакетах на базе формулы (1) на порядок увеличивает быстродействие анализа. Однако для ее использования в алгоритмах районирования необходимо вычислять все межрайонные расстояния для вновь созданного района на каждом шаге. Для преодоления указанного недостатка предлагается использовать иной подход, сокращающий количество вычислений межклассовых расстояний.

1. Определить константу D с максимальным положительным значением, которой помечаются несмежные районы в матрице межрайонных расстояний. Районам, не имеющим общей границы, присвоить заведомо недостижимые максимальные расстояния

$$D = \left\{ D(S_a, S_b) \mid g^n(a, b) = 0 \right\}.$$

2. На очередном шаге, после нахождения минимума в матрице расстояний между двумя районами S_x и S_y и объединения их в один район $S_x \cup S_y$, пересчитывать только расстояния вида

$$D(S_x \cup S_y, S_z), \text{ где } g^n = (xy, z) = 1.$$

3. Последнее условие соответствует либо случаю

$$g^n(x, z) = 1, (D(S_x, S_z) < D)$$

либо случаю

$$g^n(y, z) = 1, (D(S_y, S_z) < D),$$

поэтому при пересчете расстояний в формуле (1) некоторые слагаемые известны. Пересчет остальных расстояний формулы (1) осуществляется с использованием M -мерных векторов объектов каждого из классов.

Как видно из описания, количество вычислений межклассовых расстояний снижается, например, на первом шаге с $N-2$ до $\Theta(S_{xy})$. Под символом понимается число районов, соседних с данным районом S_{xy} . Можно считать, что $\Theta(S_{xy}) \leq 15$, хотя на практике наиболее часто $\Theta(S_{xy}) \in \{4, 5, 6\}$.

Результаты экспериментов показали, что время работы алгоритма районирования сильно зависит от пространственной структуры получаемых районов и их соразмерности. Например, программная реализация методов ближнего и дальнего соседей абсолютно идентична, однако время работы алгоритмов на одних и тех же данных сильно варьируется.

Расчеты проводились на персональном компьютере с процессором типа РПД-500 и оперативной памятью 128 мегабайт. Ниже приведены результаты некоторых замеров времени $T(n)$ в зависимости от шага алгоритма n , потребовавшегося для обработки количества объектов $N = 2500$ при количестве признаков $M = 5$.

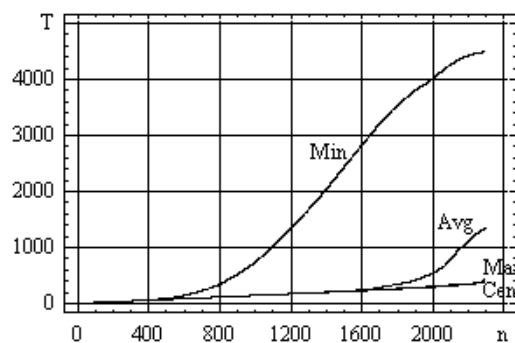


Рисунок 3 - Вид временной функции для разработанных алгоритмов районирования

Обработка модельных данных с помощью статистических пакетов и разработанных автором программ показала, что реализация некоторых алгоритмов (ближнего соседа, центроидного) на базе предложенного автором подхода почти вдвое снижает время анализа.

ПЕРЕЧЕНЬ ССЫЛОК

1. Трофимов А.М., Заботин Я.И., Панасюк М.В., Рубцов В.А. Количественные методы районирования и классификации.- Казань: изд-во Казанск. ун-та, 1985, 120 с.
2. Monmonier M.S. Maximum-difference barriers: alternative numerical regionalization method. - Geogr. Anal., 1973, vol.5, №3, p. 245-261.
3. Тикунов В.С. Классификации в географии: ренессанс или увядание? - Смоленск, Изд-во СГУ, 1997, 367 с.
4. Блануца В.И. Интегральное экологическое районирование:

5. концепция и методы. - Новосибирск, ВО "Наука", 1993, 159 с.
6. Жамбю М. Иерархический кластер-анализ и соответствия.- М: Финансы и статистика, 1989, 342 с.
7. Прохоров Б.Б. Медико-экологическое районирование и региональный прогноз здоровья населения России. - М: Изд-во МНЭПУ, 1996, 70 с.
8. Айвазян С.А., Мхитарян В.С. Основы эконометрики. Теория вероятностей и прикладная статистика. - М., Юнити, 2001, 656 с.

УДК 658.512

ОБЪЕКТНО-ОРИЕНТИРОВАННАЯ РЕАЛИЗАЦИЯ АЛГОРИТМА КЛОНАЛЬНОЙ СЕЛЕКЦИИ

В.И.Литвиненко, А.А.Фефелов, С.П.Горавский

У роботі подані результати проектування об'єктно-орієнтованої бібліотеки для розв'язання задач оптимізації за допомогою алгоритму клональної селекції. Наводиться модель алгоритму у вигляді його крокової реалізації. Структура бібліотеки і відносини між класами показані на схемах та пояснюються у тексті статті.

В работе представлены результаты проектирования объектно-ориентированной библиотеки для решения оптимизационных задач при помощи алгоритма клональной селекции. Приводится модель алгоритма в виде его пошаговой реализации. Структура библиотеки и отношения между классами показаны на схемах и разъясняются в тексте статьи.

The paper deals with the model of clonal selection algorithm that is used to solve mathematical optimization problems and represented there as the object-oriented library of classes. The model of algorithm is shown as its step-by-step realization. The structure of library and relationships between its classes

ВВЕДЕНИЕ

Иммунная система (ИС) позвоночных является одной из запутанных систем организма, а ее сложность сравнима только со сложностью нервной системы позвоночных. Развитие генетики, молекулярной биологии значительно увеличило наше понимание работы иммунной системы. Знание механизмов работы иммунной системы показывает, что они интересны не только с биологической, но и с вычислительной точки зрения. Подобно тому, как понимание работы нервной системы привело к созданию и разработке искусственных нейронных сетей, понимание работы иммунной системы привело к созданию иммунных систем - новой вычислительной парадигмы.

Искусственные иммунные системы (ИИС) определяются как метафоры или абстракции разрабатываемых вычислительных систем, использующие иммунологические механизмы и теории [1]. До недавнего времени модели иммунных систем предназначались в основном для исследования биологических аспектов иммунной системы, чтобы как можно лучше и глубже понять их поведение и таким образом содействовать развитию биологической

науки [2, 3]. В настоящий момент возрос интерес к разработке ИИС для решения прикладных вычислительных задач, таких как оптимизация, распознавание образов, классификация, data mining, оптимальное управление, идентификация систем, противовирусная защита компьютеров и компьютерных сетей. Существует множество программных реализаций предназначенных указанных задач.

Данная работа посвящена проблемам объектно-ориентированной реализации ИИС, а именно использованию принципов клонального отбора для решения задач нелинейной оптимизации.

1 БИОЛОГИЧЕСКИЕ ОСНОВЫ ИСКУССТВЕННЫХ ИММУННЫХ СИСТЕМ

Все позвоночные имеют иммунную систему, которая интересна нам с вычислительной точки зрения. Иммунная система выработалась в процессе эволюции как средство защиты от заражения микроорганизмами и более крупными паразитами. У позвоночных ИС состоит их множества разнообразных молекул и клеток, рассредоточенных и циркулирующих по всему организму. Характерной особенностью ИС, является то, что отсутствует какой-либо центральный орган, который осуществляет управление, и функционирование ИС при этом имеет несколько элементов, которые постоянно находятся в движении и выполняют различные дополнительные функции [4]. Основной задачей ИС является защита организма от живых тел и веществ, несущих на себе признаки генетически чужеродной информации, называемых антигенами. Каждый антиген специфичен, т.е. имеет в своей структуре, такие особенности, которые отличают его от других антигенов, и вызывают специфическую реакцию иммунной системы, направленную только против него одного. Т.о. с помощью сложных механизмов распознавания, которые отличают "свое" от "чужого" на основе химической структуры, иммунный процесс порождает клетки и молекулы, связывающие и уничтожающие "чужое". Образование защитных клеток и молекул называется иммунной

реакцией. Одной из привлекательных свойств ИС является наличие молекул рецепторов на поверхности иммунных клеток, способных к распознаванию разнообразных антигенов (образов). Обычно выделяют две большие группы иммунных клеток, известных как В-клетки и Т-клетки. Эти два типа довольно схожи между собой, но отличаются тем, как они распознают антигены и своими функциональными ролями: В-клетки вырабатывают антитела, а Т-клетки осуществляют иммунные реакции клеточного типа [5]. Механизмы защиты организма хозяина включают в себя большое количество взаимосвязанных процессов, одним из важных является процесс образования молекул, называемых антителами. Главными особенностями иммунной реакции является ее специфичность и наличие своего рода памяти. Специфичность означает, что антитела, образованные в ответ на некоторый антиген, сильно связывают именно этот антиген и не столь сильно другие антигены. Память проявляется в том, что в течение длительного времени после первичной реакции на антиген повторное введение того же антигена вызывает более быструю и мощную вторичную реакцию. Распознавание антигенов является первой предпосылкой для активации иммунной системы и осуществления иммунного ответа. Существует два основных типа иммунного ответа. Гуморальный ответ состоит в выработке антител, которые циркулируют в крови и специфически связываются с вызвавшими их синтез чужеродными молекулами. Иммунный ответ клеточного типа включает образование специализированных клеток, реагирующих с чужеродным антигеном главным образом на поверхности собственных клеток организма. Распознавание должно удовлетворять нескольким критериям. Антитела или иммуноглобулины играют весьма важную роль в иммунном процессе. Они представляют собой молекулы, которые с помощью активных центров могут связываться с некоторыми, обладающими определенной химической структурой функциональными группами антигенов. Эти группы называются антигенными детерминантами или эпитопами. Равновесная константа ассоциации для связывания с определенной антигенной детерминантой называется аффинитетом или сродством молекулы антитела к соответствующему антигену. Первоначально, рецептор клетки распознает антиген с некоторой аффинностью, и осуществляется связывание между рецептором и антигеном с силой, пропорциональной этой аффинности. Если аффинность больше, чем данная пороговая, то иммунная система активизируется. На результат столкновения между антигеном и рецептором клетки влияет природа антигена и тип распознающего сайта. Связывание антител с антигеном облегчает его удаление, которое осуществляется специализированными клетками организма (фагоцитами и макрофагами). Если антигены представляют собой молекулы, то комплексы антигена и антител (иммунные комплексы) способны образовывать агрегаты больших размеров, обладающие цитофильными свойствами. Кроме того, защитное действие антител может осуществляться путем прямой нейтрализации активности антигена. Если антиген является проникшей в организм бактерией или вирусом, взаимодействие с антителами может привести к лизису "агрессора". После налипания антитела к поверхности

бактерии запускается цепь ферментативных реакций, в которые последовательно вступают компоненты комплемента и вызывают разрыв бактериальной стенки. Функционально этот процесс представляет собой все большее увеличение концентрации фермента за счет мультипликативных управляющих воздействий который, в конце концов, разрушает клеточную мембрану. Иммунная система человека содержит орган, называемый тимусом или вилочковой железой, расположенный в грудной клетке в области верхнего межплеврального пространства переднего средостения на уровне верхней части грудной кости, которая выполняет главную роль в созревании Т-клеток. После того как Т-клетки сгенерированы, они поступают в тимус, где происходит их созревание. В течение этого созревания, все Т-клетки, которые распознают антигены своего организма, удаляются из популяции Т-клеток, данный процесс называется отрицательной селекцией. Если В-клетка сталкивается с чужим антигеном с достаточной аффинностью, то такие клетки пролиферируют и дифференцируют в эффекторные клетки и клетки памяти, а данный процесс называется клональной селекцией. В противоположность этому, если В-клетки распознают свои антигены, то это как предполагается в соответствии с теорией иммунной сети, может привести к супрессии. В данной статье мы не будем затрагивать алгоритмы на основе отрицательной селекции и иммунной сети, а остановимся на аналоге вычислительного алгоритма клональной селекции. Но перед этим более подробно рассмотрим теорию клональной селекции.

2 КЛОНАЛЬНАЯ СЕЛЕКЦИЯ

Клональная селекция является теорией, которая используется для объяснения механизмов иммунного ответа при распознавании иммунными клетками образов чужих антигенов В-клетками. Первую, чисто биологическую, селективную теорию образования антител сформулировал в 1955 году Нильс Эрне [4]. Он предположил, что в организме действительно синтезируется полный набор антител, но каждое из них образуется в небольшом количестве и поступает в кровь в виде "естественных антител". Эти антитела избирательно связываются с соответствующим антигеном, и доставляет этот антиген неким клеткам организма, для которых антитела служат сигналом к воспроизведению таких же молекул, т.е. к образованию большого количества специфических антител. С этой точки зрения вторичный антительный ответ объяснялся как результат появления после первой иммунизации увеличенного числа антител-"носителей", присутствие которых будет одновременно способствовать селекции антигеном антител с более высоким сродством, что объясняет изменение качества антител при многократной иммунизации. Гипотеза, высказанная Н. Эрне, согласно которой роль антигена сводится к отбору преадаптированных к нему клеток, получила дальнейшее развитие в теории селекции клонов Франка Бернета [6, 4], постулирующей четыре основных положения: 1) лимфоидная ткань организма содержит большое количество клеток; 2) популяция лимфоидных клеток гетерогенна и состоит из большого количества клонов, возникших в результате мутаций;

постоянные процессы мутаций лимфоидных клеток обеспечивают достаточное количество отдельных клонов лимфоидных клеток, специфических по отношению к возможному количеству антигенных детерминант; клонирование может быть обусловлено генетическим кодом; 3) малое количество антигена стимулирует специфический клон лимфоидных клеток к размножению и дифференцировке в плазмциты, чем и обеспечивается иммунный ответ - выработка антител; 4) большое количество антигена элиминирует соответствующий клон лимфоидных клеток, чем обуславливается гибель возникших в эмбриональном периоде лимфоидных клеток, способных реагировать против собственных антигенов, и формируется естественная толерантность к ним.

Теория Бернета постулирует, что иммунокомпетентная клетка (малый лимфоцит) содержит в своем геноме генетическую информацию к ответу на один антиген до вовлечения ее в иммунный ответ. В процессе иммунного ответа клетки пролиферируют, образуя клон. Вторая клетка отвечает за второй антиген и т.д. Пул лимфоцитов индивидуума запрограммирован для иммунного ответа на различные антигены [7, 8]. Таким образом, лимфатическая система организма имеет набор предшествующих клонов, запрограммированных отвечать на различные антигены. Установлено, что одна клетка может продуцировать антитела нескольких классов. Возникают клоны лимфоидных клеток как следствие экспрессии различных генов, содержащихся в геноме, и в результате соматических мутаций. Каждый В-лимфоцит генетически запрограммирован синтезировать только один вид антител, содержащих один вид рецепторов для антигена. Этот рецептор содержится на поверхностной мембране лимфоцитов. Присоединение антигена к рецепторам активирует лимфоцит к пролиферации и дифференцировке, в результате чего возникает клон плазматических клеток, синтезирующий антитела к данному антигену. Согласно клонально-селекционной теории иммунитета, способность популяций лимфоидных клеток вырабатывать антитела определенной специфичности не зависит от антигена, а детерминирована генетически. Роль антигена состоит в том, что он стимулирует (отбирает, селекционирует) соответствующий ему клон клеток лимфоидной ткани, являющийся носителем рецептора, с которым антиген может специфически реагировать.

3 МОДЕЛЬ АЛГОРИТМА КЛОНАЛЬНОЙ СЕЛЕКЦИИ

Алгоритм клональной селекции, моделирующий основные принципы теории клонального отбора иммунной системы, относится к категории систем вычислительного интеллекта [1], куда входят системы способные к адаптации своего поведения, при достижении целей, т.е. обладающие способностью обучаться даже в отсутствии человека, в отличие от систем искусственного интеллекта и так называемых "символьных вычислений", последние из которых неспособны к самообучению вообще.

Формально клональный алгоритм для задач оптимизации можно представить следующим образом [9]:

$$CLONALG = (Ab^0, f, L, N, n, \beta, d, \epsilon), \quad (1)$$

где Ab^0 - исходная популяция антител; f - целевая функция; N - количество антител в популяции; L - длина рецептора антитела; n - количество антител, отбираемых для клонирования (с самой высокой аффинностью); β - множительный фактор, регулирующий количество клонов отобранных антител; d - количество антител, подлежащих замене новыми (т.е. имеющие самую низкую аффинность); ϵ - критерий останова.

Пошаговая реализация алгоритма приведена ниже.

Шаг 1. Инициализация. Генерировать начальную популяцию Ab^0 антител

$$Ab^0 = \emptyset, j \in \{1..N\}, t = 0.$$

Случайно выбрать $i_j \in I$.

$$Ab^0 = Ab^0 + i_j,$$

где I - пространство индивидуумов, т.е. множество всех возможных структур, представляющих антитело, $i \in I$ - подмножество индивидуумов, составляющих популяцию, t - номер поколения.

Шаг 2. Определение аффинности. Для каждого антитела $Ab_j \in Ab^t$ вычислить значение целевой функции $y_i = f(Ab_j)$ и определить аффинность $g_j = affinity(y_j)$, $j \in \{1..N\}$.

Шаг 3. Селекция. Выбрать подмножество антител с самой высокой аффинностью ($Ab_{\{n\}}$).

$$Ab_{\{n\}} = \{Ab_j \in Ab^t | select(Ab_j, Ab^t, n) = 1\}, \text{ где}$$

$$select(Ab_j, Ab^t, n) = \begin{cases} 1, & rank(Ab_j) < n; \\ 0, & rank(Ab_j) \geq n; \end{cases}$$

$$rank(Ab_j) = j, \text{ если}$$

$$\forall j \in \{1..N-1\}: affinity(f(Ab_j)) \geq affinity(f(Ab_{j+1})).$$

Шаг 4. Клонирование. Получить популяцию клонов $C_{\{N_c\}}$ из $Ab_{\{n\}}$.

$$C_{\{N_c\}} = \emptyset,$$

$$\forall j \in \{1..N_c\}: C_j = Ab_k, Ab_k \in Ab_{\{n\}}, \text{ где}$$

$$k = round\left\{\frac{j}{round(\beta \cdot N)}\right\},$$

$$N_c = \prod_{i=1}^n round(\beta \cdot N),$$

$round(x)$ - оператор взятия целой части числа.

Шаг 5. Гипермутация. Получить популяцию измененных клонов $C_{\{N_c\}}^*$ из $C_{\{N_c\}}$.

$$C_{\{N_C\}}^* = \emptyset;$$

$$\forall j \in \{1..N_C\}: C_j^* = \begin{cases} mutate(C_j), rnd(p_m) = 1; \\ C_j, rnd(p_m) = 0; \end{cases}$$

$$C_j \in C_{\{N_C\}},$$

где $rnd(p_m)$ - функция моделирования наступления случайного события с заданной вероятностью p_m , $mutate(C_j)$ - оператор мутации, который случайно изменяет один или несколько генов антитела.

Шаг 6. Определение аффинности популяции измененных клонов. Для каждого антитела $C_j^* \in C_{\{N_C\}}^*$ вычислить значение целевой функции $y_j = f(C_j^*)$ и определить аффинность $g_j = affinity(y_j)$, $j \in \{1..N_C\}$.

Шаг 7. Селекция. Выбрать подмножество $C_{\{n\}}^*$ из n антител с самой высокой аффинностью из популяции измененных клонов $C_{\{N_C\}}^*$, подобно шагу 3.

Шаг 8. Замещение. Заменить подмножество $Ab_{\{n\}}$ на $C_{\{n\}}^*$.

$$\forall j \in \{1..n\}: Ab_j \in C_j^*, Ab_j \in Ab_{\{n\}}, C_j^* \in C_{\{n\}}^*.$$

Шаг 9. Клональное удаление. Заменить подмножество антител $Ab_{\{d\}}$ с самой низкой аффинностью новыми индивидуумами.

$$Ab_{\{d\}} = \{Ab_j \in Ab^t | negselect(Ab_j, Ab^t, d) = 1\},$$

где

$$negselect(Ab_j, Ab^t, d) = \begin{cases} 1, rank(Ab_j) < N - d; \\ 0, rank(Ab_j) \geq N - d; \end{cases}$$

$$rank(Ab_j) = j, \text{если}$$

$$\forall j \in \{1..N-1\}: affinity(f(Ab_j)) \geq affinity(f(Ab_{j+1})),$$

$$k \in \{1..d\}.$$

Случайно выбрать $i_k \in I$.

$$Ab_k = i_k, Ab_k \in Ab_{\{d\}}.$$

Шаг 10. Проверка условия останова. По выбранному критерию ϵ проверить выполнение условия останова алгоритма.

$$Ab^{t+1} = Ab_{\{n\}} \cup Ab_{\{d\}} \cup Ab_{\{N-(n+d)\}}, t = t + 1,$$

Вывод $Ab_{\{n\}}$, если $stop(\epsilon) = true$;

возврат к шагу 2, если $stop(\epsilon) = false$.

Замечания. Оператор мутации подробно не рассматривается в данной работе, потому что на сегодняшний день существует большое число различных вариантов его реализации. Параметры n и d не связаны друг с другом

напрямую, а только через размер популяции N , т.е. на них накладываются ограничения - $n + d \leq N$. Это означает, что в популяции по прошествии одного поколения могут оставаться неизменные индивидуумы. Перед применением оператора селекции все индивидуумы должны быть проранжированы, т.е. необходимо осуществить сортировку популяции по убыванию аффинности. Следовательно, наименьший ранг получит индивидуум с наибольшим значением аффинности.

4 ОБЪЕКТНО-ОРИЕНТИРОВАННЫЙ ПОДХОД

В настоящее время уже существует ряд пакетов программ, предназначенных для решения различных математических задач. Большинство пакетов представляет собой набор стандартизированных автономных программных модулей, организованных в виде библиотек, написанных на процедурно-ориентированных языках, таких, например, как Фортран или Си. Процедурные модули пакетов могут быть непосредственно включены в любую авторскую программу. При этом предполагается, что пользователь библиотеки имеет квалификацию математика-программиста, необходимую для корректного использования модулей при программной реализации средств постановки и решения задачи. На практике подобные допущения не всегда оправданы, и пользователь сталкивается с серьезными трудностями, связанными с необходимостью адаптации модулей к конкретным прикладным задачам, которые могут отличаться от стандартных постановкой и алгоритмами решения. Процедурные модули плохо приспособлены для модификаций и пользователю нередко приходится перепрограммировать их заново. В связи с этим, важное значение, приобретает применение объектно-ориентированного подхода (ООП) [11], обеспечивающего широкие возможности для развития и адаптации уже созданного программного обеспечения. Однако преимущества использования ООП могут быть достигнуты только в результате глубокого объектного анализа предметной области и тщательного объектного проектирования. Объектно-ориентированные библиотеки, по существу, представляют собой наборы классов, организованные в виде деревьев или иерархий. Такие иерархии могут включать в себя производные классы (классы - наследники) и классы, содержащие в себе другие классы (классы-контейнеры). Преимущество ООП по сравнению с процедурным программированием состоит в том, что набор классов библиотеки уже, как правило, обеспечивает базовую функциональность того или иного математического метода и пользователю остается провести лишь "косметическую" адаптацию одного или нескольких классов под конкретную задачу. Для этого создается потомок нужного класса, добавляются необходимые свойства и переписываются необходимые виртуальные методы (рис.1).

```

//////// стандартная библиотека //////////
class LibraryClass
{
    char m_property1;
    virtual int Method1(char*);
    virtual char* Method2(void);
};

//////// аффсаая рффаа //////////

class MyClass : public LibraryClass
{
    int m_property2;
    int Method1(char*);
};
int MyClass :: Method1(char* ch)
{
    // метод
}
    
```

Рисунок 1 - Общий вид стандартной библиотеки

В данной работе представлены результаты реализации объектно-ориентированной библиотеки, написанной на языке Си++ и предназначенной для решения одномерных и многомерных задач мультимодальной нелинейной оптимизации с ограничениями и без ограничений при помощи алгоритма клональной селекции [9,10].

5 ОПИСАНИЕ БИБЛИОТЕКИ

Прежде чем приступить к описанию структуры библиотеки, остановимся на условных обозначениях, используемых в схемах [12] (табл.1).

Основываясь на данных модели, рассмотренной в разделе 2, можно выделить следующие объекты предметной области.

1. Проблема (Класс CTask). Класс, описывающий задачу, решаемую при помощи алгоритма клональной селекции и содержащий в себе методы, необходимые для представления данных задачи в виде иммунных структур, обратного представления, а, также, оценки результатов.

2. Ген (Класс CGene). Класс, представляющий минимальный неделимый сегмент рецептора антитела.

3. Антитело (Класс CAntiBody). Класс, содержащий в себе полную информацию о данных задачи (один вариант решения задачи).

4. Популяция (Класс CImmunePopulation). Популяция антител, реализующая основные механизмы клонального отбора.

5. Алгоритм (Класс CImmuneAlgorithm). Класс, включающий в себя основные и вспомогательные объекты модели и обеспечивающий функционирование алгоритма клональной селекции.

Пример взаимного включения основных объектов для построения работоспособного алгоритма показан на рисунке 2.

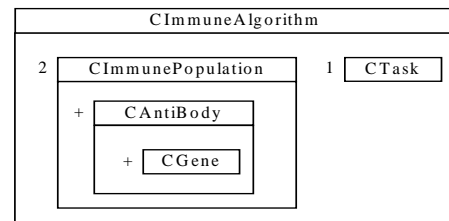


Рисунок 2 - Пример включения основных объектов

Таблица 1 - Используемые условные обозначения

| Обозначение | Описание |
|-------------|---|
| | Показывает, что Class1 является контейнером для объектов Class2 и Class3. Цифра в левой части прямоугольника показывает, какое количество экземпляров объектов соответствующего класса может содержать контейнер. Знак "+" говорит о том, что должен содержаться как минимум один экземпляр |
| | Показывает отношение наследования между классами. Class2 является производным от Class1 |
| | Показывает отношение между классами. Например, передача сообщения объекту Class2, представленная ниже BOOL Class1:: Create(void) { Class2->Init(void); } реализует отношение инициализации |
| | Черный квадрат показывает, что реализация данного отношения зависит от конкретной задачи и в стандартной библиотеке может быть не определена. При этом вполне возможно наличие соответствующих интерфейсов, адаптация и доработка которых оставлена на усмотрение пользователя |

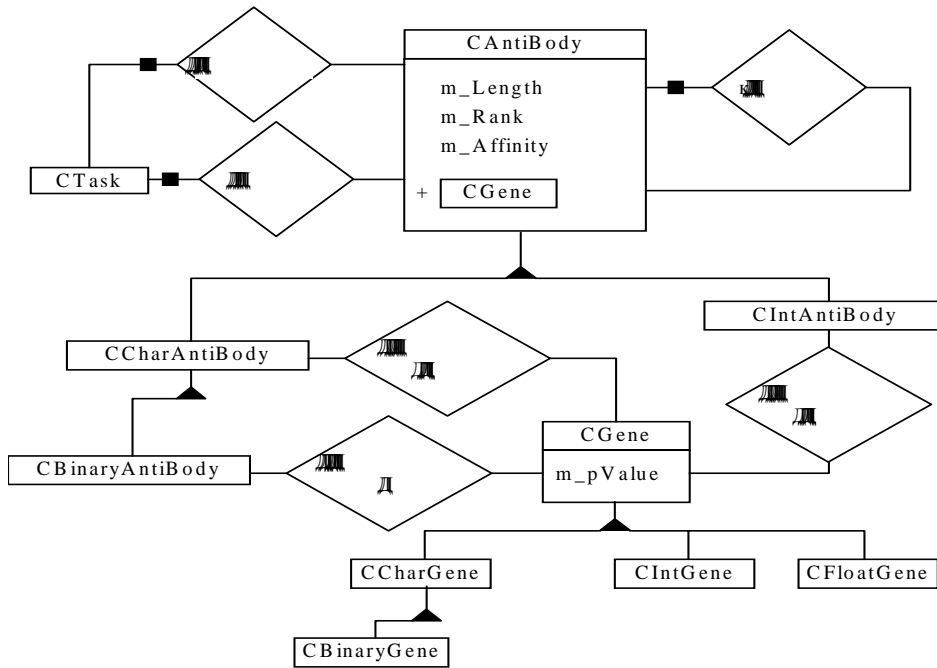


Рисунок 3 - Пример описания иерархии классов

На рисунке 3 показана иерархия классов, производных от CGene и CAntiBody. Класс CGene содержит единственный атрибут m_pValue, являющийся указателем на структуру данных, кодирующую сегмент рецептора антитела.

Атрибут m_pValue описан как указатель на void, благодаря чему имеется возможность использования в качестве данных, структуры любого стандартного или определяемого пользователем типа, например типа double (рис. 4).

```

//// стандартная структура
class CGene {
    void* m_pValue;
    ... };

//// альтернативная структура
class CDoubleGene : public CGene {
public:
    CDoubleGene() {
        ...
        m_pValue = new double;
        ... };
}
    
```

Рисунок 4 - Пример описания атрибута m_pValue как указателя на void

В рассматриваемой здесь библиотеке реализованы классы, производные от CGene: CCharGene, CIntGene и CFloatGene, работающие с данными соответствующих типов. Кроме того, от класса CCharGene создан производный CBinaryGene, в котором *m_pValue может содержать только два значения "0" и "1".

Класс CAntiBody является контейнером для объектов класса CGene и содержит в себе массив данных, являю-

щихся рецептором антитела. Рецептор кодирует набор входных переменных решаемой задачи. Кроме того, класс CAntiBody содержит атрибуты: m_Length - длина рецептора, m_Rank - значение ранга антитела в популяции и m_Affinity - аффинность антитела, определяемая на шаге 2 и 6 алгоритма (раздел 2). Возможны три вида отношений инициализации между классами, производными от CAntiBody и классом CGene.

1. Инициализация случайным значением в интервале, определяемом максимальным и минимальным значением стандартного типа данных.
2. Инициализация случайным значением в интервале, определяемом пользователем.
3. Инициализация заданным значением.

Для класса CBinaryAntiBody возможно, также, отношение инверсии гена. Отношения инициализации являются базой для работы оператора гипермутации.

Отношения между классами CTask и CAntiBody зарезервированы в библиотеке, но не реализованы, потому что заранее не известно какую именно задачу будет решать пользователь. Конструирование рецептора антитела, а, также, его декодирование и оценка зависят от конкретной задачи и должны быть реализованы самим пользователем. Класс CTask содержит лишь интерфейсы соответствующих методов (рис. 5).

```

class CTask {
    ...
    virtual void Construct(CImmParams* Params){}
    virtual double AffinityFunc(CAntiBody* Ab){}
    ...
};
    
```

Рисунок 5 - Описание класса класса CTask

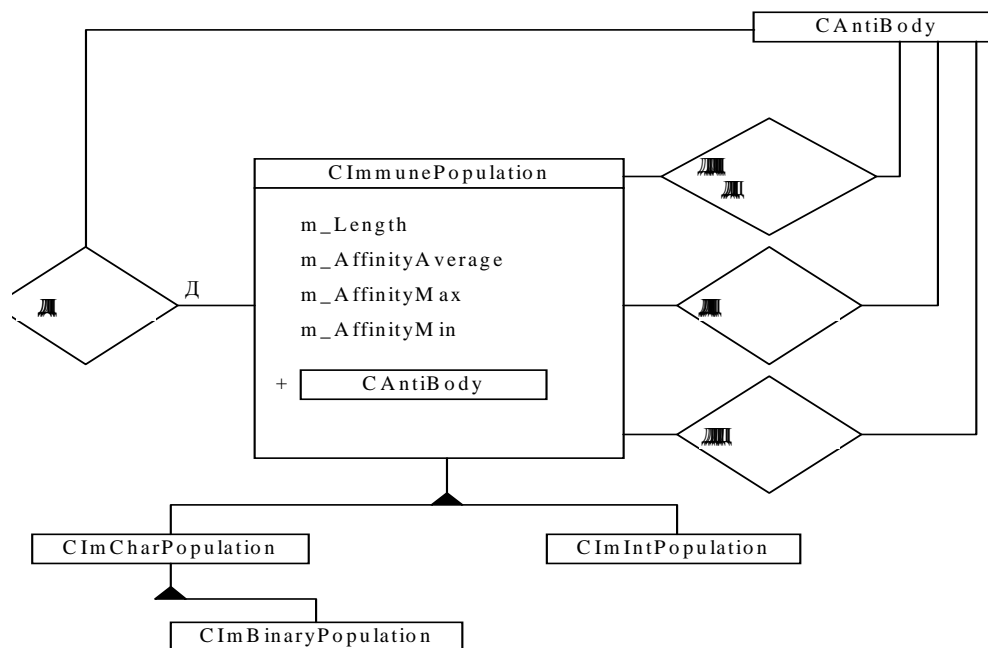


Рисунок 6 - Основные характеристики класса *CImmunePopulation* и его отношения с классом *CAntiBody*

На рисунке 6 показаны основные характеристики класса *CImmunePopulation* и его отношения с классом *CAntiBody*. Этот класс содержит в своем составе несколько атрибутов таких как `m_Length` (размер популяции), `m_AffinityAverage` (средняя аффинность популяции), `m_AffinityMax` (аффинность лучшего антитела в популяции), `m_AffinityMin` (аффинность худшего антитела в популяции), а, также, массив антител, в котором должен находиться как минимум один элемент. Класс *CImmunePopulation* содержит методы обработки массива антител, а именно: инициализация антител, назначение ранга антителу в зависимости от его аффинности, сортировка массива, поиск лучшего антитела в популяции и метод, реализующий оператор селекции (шаги 3 и 7 алгоритма). Как и классы *CGene* и *CAntiBody* класс *CImmunePopulation* является абстрактным и для того, чтобы работать с конкретными структурами данных созданы производные классы: *CImCharPopulation* (популяция антител, рецепторы которых представляют собой строки символов типа `char`), *CImBinaryPopulation* (рецепторы антител представлены в виде бинарных строк), *CImIntPopulation* (используется кодировка рецепторов в виде строк целых чисел). Здесь, как и в других классах библиотеки, имеется возможность создания пользовательского класса, производного от *CImmunePopulation*, который будет обрабатывать структуры, созданные пользователем для решения его задачи.

Класс алгоритма (*CImmuneAlgorithm*) (рис. 7) на самом высоком уровне реализует эволюционный процесс в иммунной системе - последовательность действий, представленную в разделе 2 настоящей работы. Он содержит в

себе два экземпляра объектов класса *CImmunePopulation*, один из которых представляет основную популяцию антител (Ab), а другой - популяцию клонов (C). Для этих двух популяций в классе *CImmuneAlgorithm* представлены методы, осуществляющие перенос антител из популяции Ab в популяцию C и их дальнейшее клонирование. Используя данные класса *CImmParams*, а именно `MutationProbability` (вероятность мутации) класс алгоритма осуществляет мутацию антител популяции клонов C. Класс *CImmParams* содержит в себе основные параметры алгоритма: `PopSize` - размер популяции, `IndSize` - размер индивидуума, `n_param` - количество антител, отбираемых для клонирования, `d_param` - количество антител, удаляемых из популяции и заменяемых новыми, `Beta` - множительный коэффициент, определяющий интенсивность клонирования (шаг 4 алгоритма) и т.д. Некоторые из этих параметров устанавливаются до начала работы системы, пользователем, другие же рассчитываются с использованием класса *CTask*, в зависимости от типа решаемой задачи. Условие останова (`StopCondition`), также реализовано в виде объекта, потому как для разных задач эти условия могут быть различны (например, останов по прошествии определенного числа поколений или останов по достижении определенного качества решения и т.д.).

Сборка алгоритма может осуществляться либо в отдельном модуле, либо, непосредственно, в тексте пользовательской программы. Библиотека может использоваться как статически подключаемая на этапе компиляции (`.lib`), так и как динамически подгружаемая (`.dll`) на этапе выполнения пользовательской программы.

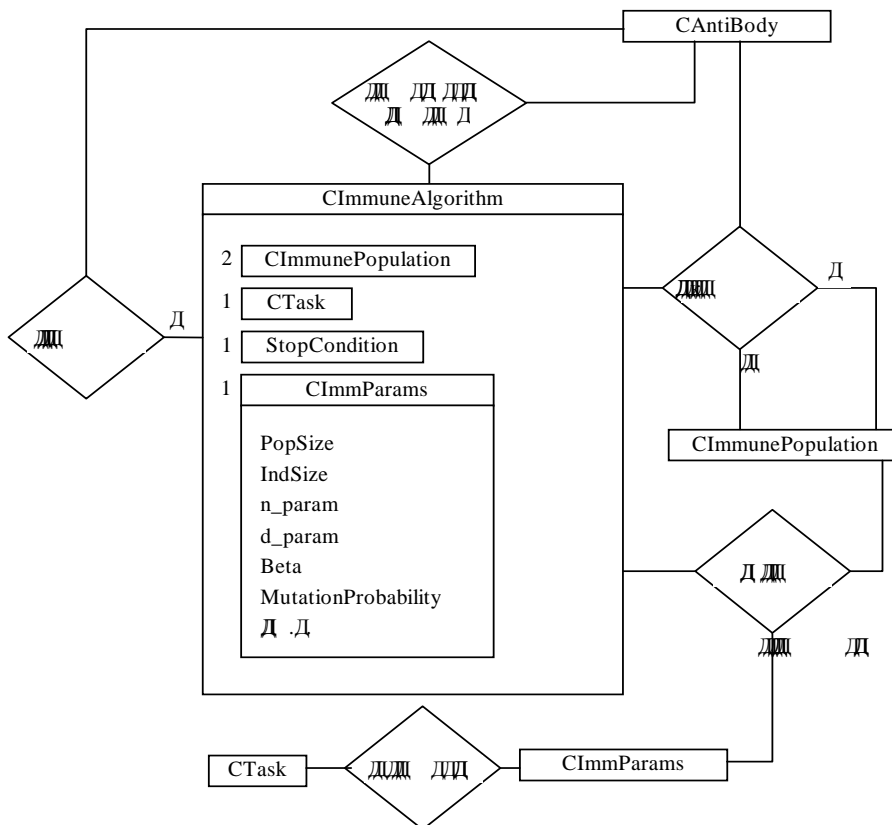


Рисунок 7 - Схема реализации эволюционного алгоритма при помощи класса CImmuneAlgorithm

ЗАКЛЮЧЕНИЕ

Иммунная система позвоночных является большой темой для из-за ее мощных способностей по обработке информации. В ней осуществляется много сложных параллельных и распределенных вычислений. Так же как и нервная система, иммунная система может самостоятельно обучаться и осуществлять операцию распознавания образов децентрализованным образом. При ее обучении имеют место эволюционные процессы.

В данной статье рассмотрена объектно-ориентированная реализация лишь клональной теории объясняющей некоторые механизмы функционирования иммунной системы. Нами разработана библиотека классов для решения задач мультимодальной, нелинейной оптимизации с ограничениями и без ограничений при помощи алгоритма клональной селекции. Спроектированная классовая иерархия носит открытый характер, т.е. является расширяемой и адаптируемой под конкретную пользовательскую задачу. Программный код системы реализован на языке программирования C++ и включен в состав более обширной библиотеки классов по вычислительному интеллекту, которая разрабатывается в данный момент. В нее уже вошли такие составляющие как генетические алгоритмы, генетическое программирование и системы программирования генных выражений. Дальнейшее усовершенствование разработанной структуры предполагает включение в ее состав классов и методов для осуществления

многокритериальной оптимизации с получением на выходе множества Парето оптимальных решений.

ПЕРЕЧЕНЬ ССЫЛОК

1. Fogel, D. V. Evolutionary Computation: Toward a New Philosophy of Machine Intelligence, 2nd ed., IEEE Press (2000).-328 p.
2. Белых Л.Н. Анализ математических моделей в иммунологии / Под ред. Г.И. Марчука. - М.: Наука, Гл. ред. физ.-мат. лит., 1988. - 192 с.
3. Молер Р.Р., Бруни К., Гандолфи А. Системный подход в иммунологии. - ТИИЭР, 1980, т. 68, № 8, с.25-56
4. Пол У., Сильверстайн А, Купер М. и др. Иммунология: в 3-х томах: Пер. с англ. / Под ред.У.Пола. - М.: Мир, 1987-1988.
5. Альбертс Б., Брей Д. Льюис Дж. Рэфф М., Робертс К. Уотсон Дж. Молекулярная биология клетки: Пер. с англ. - М.: Мир, 1987. - Т.5. - 231 с.
6. Burnet F.M. The Clonal Selection Theory of Acquired Immunity. The University Press, Cambridge, 1959.
7. Вершигора А.Е. Общая иммунология: Учеб. пособие. - К.: Вища школа, 1989. -736 с.
8. Ройт А. Основы иммунологии: Пер. с англ. - М.: Мир, 1991.-328 с.
9. De Castro, L. N. & Von Zuben, F. J. (1999), "Artificial Immune Systems: Part I - Basic Theory and Applications", Technical Report - RT DCA 01/99, FEEC/UNICAMP, Brazil, 95 p.
10. Leandro Nunes de Castro, Fernando J. Von Zuben. The clonal selection algorithm with engineering applications. Proc. of GECCO'00, pp. 36-37.
11. Пол А. Объектно-ориентированное программирование на C++.- 2-е изд. / Пер. с англ. -М.: "Изд. Бинум", СПб: "Невский диалект", 1999. - 560 с.
12. Partice Roger Calegari. Parallelization of population-based evolutionary algorithms for combinatorial optimization problems. Ecole polytechnique federale de Lausanne. These # 2046. Lausanne, EPFL, 1999, 145 p.

ИЕРАРХИЧЕСКАЯ СТРУКТУРА МОДЕЛИ АНАЛИЗА И ПРОГНОЗИРОВАНИЯ ПОДТОПЛЕНИЯ: АНАЛИЗ МОДЕЛЕЙ-ОСНОВАНИЙ

С.А.Митрохин

Представлено ієрархічну структуру проблемо-розв'язуючої моделі аналізу і прогнозування підтоплення. Розглянуто моделі природних і техногенних джерел підтоплення, знакові моделі факторів розвитку процесу, особливості рішення задач зонуювання і візуалізації нечітких границь об'єктів.

Представлена иерархическая структура проблеморазрешающей модели анализа и прогнозирования подтопления. Рассмотрены модели естественных и техногенных источников подтопления, знаковые модели факторов развития процесса, особенности решения задач зонирования и визуализации нечетких границ объектов.

It presents a hierarchical structure of problem-solving model of estimation and forecasting of underflooding. The models of natural and man-made sources of underflooding, signary models of factors of evolution of process, peculiaritys of a decision of zoning tasks and visualization of fuzzy boundaries of objects were considered.

ВВЕДЕНИЕ

Для территорий большинства современных промышленно-городских агломераций характерным является процесс подъема уровня грунтовых вод, который ведет к подтоплению жилых зданий и промышленных объектов, а также к активизации других опасных экзогенных процессов (оползневых, обвально-осыпных, карстово-суффозионных и т.д.). В связи с обострением кризиса, связанного с интенсификацией процесса подтопления территорий, сегодня все более важное значение принимает системная деятельность по управлению динамикой уровня грунтовых вод с целью защиты территорий и сооружений от опасных экзогенных процессов и их последствий.

Важнейшим, организующим элементом такой деятельности является цель, представляющая собой образ желаемого будущего, т.е. модель экологического состояния, на достижение которого и направлена деятельность. Подобная модель может быть представлена совокупностью нормативных медикосанитарных условий жизни населения и нормативных санитарногигиенических, социальных и рекреационных условий защищаемых территорий. Однако роль моделирования этим не ограничивается. Системность деятельности по управлению динамикой уровня грунтовых вод заключается в том, что она должна осуществляться по определенному алгоритму, разработанному на основании всесторонне взвешенного и достаточно полного исследования проблемы оценки и прогнозирования подъема уровня грунтовых вод в пределах промышленно-городских агломераций (ПГА).

ИЕРАРХИЧЕСКАЯ СТРУКТУРА МОДЕЛИ АНАЛИЗА И ПРОГНОЗИРОВАНИЯ ПОДТОПЛЕНИЯ

Привлечение современных методов системного анализа [6] к исследованию данной проблемы позволило определить структуру многоуровневой системы оценки и прогнозирования процесса опасного подъема уровня грунтовых вод [3] и сформировать в виде системы взаимосвязанных уровней единую многоуровневую иерархическую структуру проблеморазрешающей модели анализа и прогнозирования подтопления территорий ПГА [3], которая приводится на рис. 1.

В качестве ключевых элементов предлагаемой модели выделяются пространственные и временные критерии подтопления (уровень 2), физические и математические модели воздействия источников подтопления и модели комплексного воздействия факторов развития процесса подтопления (уровень 3), модели инфильтрации атмосферных осадков, модели конденсации и концентрации влаги в грунтах, модели инфильтрации утечек из водосодержащих сооружений и водонесущих коммуникаций, модели фильтрационных потерь вод, задействованных в производственном цикле предприятий с "мокрым" технологическим процессом, модели геоморфологических, геологическо-геологических, гидрологических и гидрографических характеристик территории, модели комплекса функционального освоения территории, модели техногенных изменений естественных условий (уровни 4 и 5).

Иерархический способ построения модели (разноразноуровневость составляющих ее элементов, включенность одних в другие) позволяет успешно применять ее при различной подробности, детализации, формализованности составляющих ее частей (моделей-оснований). При такой организации модели оказывается возможным в единой структуре направить на достижение общей цели самые разнообразные модели (отличающиеся не только степенью формализованности, но и самими способами представления знаний об объекте исследования) - от условных (знаковых), логических, до формальных физико-математических моделей [6].

Основополагающим этапом построения проблеморазрешающей модели анализа и прогнозирования подтопления является проведение всестороннего анализа моделей, лежащих в основе моделей воздействия источников подтопления и комплекса факторов развития процесса. Специфические особенности моделей-оснований обусловлены сложностью изучаемых объектов и процессов, наличием полуквантитативной и качественной информации, неоднозначностью геологических классификаций и сложностью формализации геологических представлений.

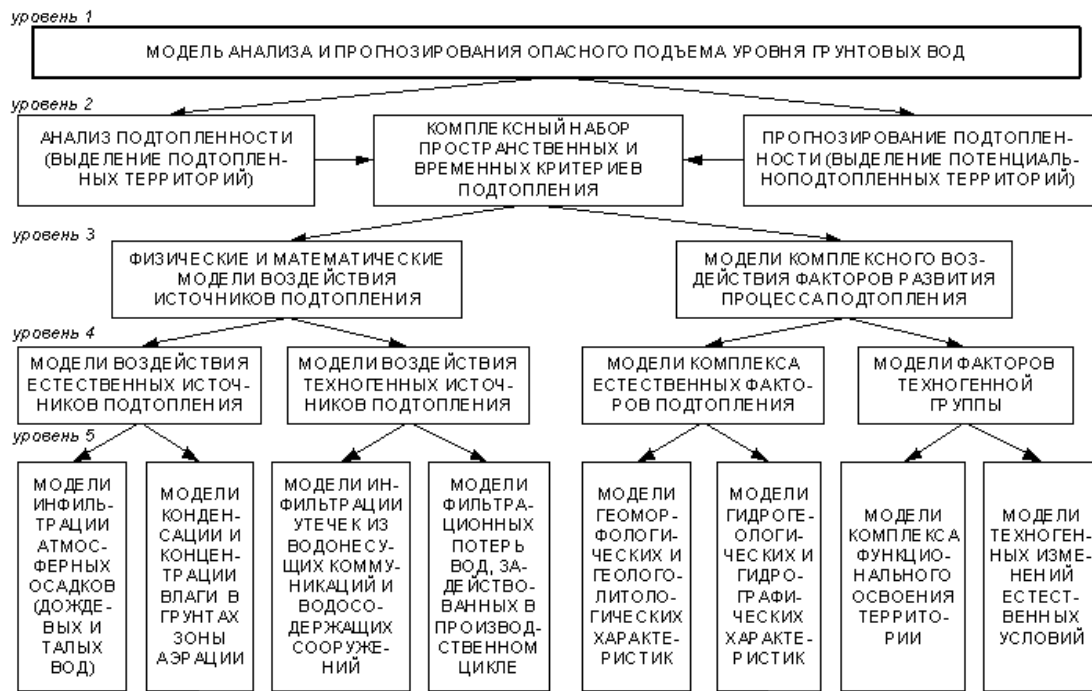


Рисунок 1 - Иерархическая структура проблеморазрешающей модели анализа и прогнозирования подтопления

МОДЕЛИ ВОЗДЕЙСТВИЯ ЕСТЕСТВЕННЫХ И ТЕХНОГЕННЫХ ИСТОЧНИКОВ ПОДТОПЛЕНИЯ

Математическая интерпретация воздействия естественных и техногенных источников подтопления представляет собой своеобразный набор модулей (моделей-оснований) из которых конструируется комплексная модель воздействия источников подтопления на экосистему (см. рис. 1). Математическое описание каждой составляющей комплексной модели основывается на общепринятых научно-обоснованных методиках [7, 9], согласно которым и выполняется фактическая и прогнозная количественная оценка составляющих участвующих в пополнении (питании) подземных вод.

Одной из важнейших составляющих модели воздействия естественных источников подтопления является модель инфильтрации атмосферных осадков (дождевых и талых вод). Атмосферные осадки являются наиболее мощным естественным источником подтопления, определяющим приходную часть водного баланса территорий, и в подавляющем большинстве случаев определяют формирование грунтовых вод. Кроме годового количества осадков, очень большое значение при формировании режима грунтовых вод принадлежит внутригодовому распределению осадков и тому виду, в котором он выпадают. В связи с этим количественная оценка атмосферных осадков подразумевает количество выпадающих осадков, продолжительность, интенсивность, вид и время выпадения [1,7]. Результаты наблюдений представляют в виде слоя осадков, рассчитанного для каждого из наблюдений, а по результатам многолетних наблюдений составляют карты среднесногодового годового распределения осадков на изучаемой территории. Графическая интерпретация подобной модели осуществляется посредством использования

кривых, графиков, номограмм. Приближенную прогностическую оценку атмосферных осадков для исследуемого района осуществляют методами интерполяции [1, 9].

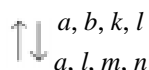
В модели воздействия техногенных источников подтопления предлагается выделить следующие элементы: модели инфильтрации утечек из водонесущих коммуникаций и водосодержащих сооружений всех типов, а также модели фильтрационных потерь вод задействованных в производственном цикле предприятий с "мокрым" технологическим процессом. Однако количественный анализ и прогноз вышеперечисленных составляющих модели воздействия техногенных источников подтопления связан с преодолением информационного барьера обусловленного неполнотой и неточностью исходной информации. Естественно, что величина утечек из водосодержащих и водонесущих коммуникаций находится в прямой зависимости от величины водозабора, однако, несмотря на то, что сведения о водозаборе обычно имеются в коммунальных органах городов и на промышленных предприятиях оценить фактические утечки с необходимой точностью практически невозможно. Невозможно также принимать в расчетах в качестве фактических утечек оговоренные соответствующими СНиПами (в виде % от водозабора) величины максимально допустимых утечек, так как вследствие крайне изношенного состояния водонесущих коммуникаций и водосодержащих сооружений фактические утечки намного больше допустимых, т.е. существующие на сегодня СНиПы вошли в противоречие с реальной действительностью. Также следует отметить, что на суммарную величину утечек кроме величины водозабора оказывают непосредственное влияние большое количество других разнотипных факторов, а существующие на сегодняшний день методики расчета не учитывают смешанный характер входных данных, основываются в

основном непосредственно на данных наблюдений и не могут быть использованы для систем поддержки принятия решений с участием ЛПР без осуществления качественной переработки.

МОДЕЛИ КОМПЛЕКСА ЕСТЕСТВЕННЫХ И ТЕХНОГЕННЫХ ФАКТОРОВ ПОДТОПЛЕНИЯ

Спецификой моделей, применяемых для представления взаимодействующих факторов развития процесса подтопления, является то, что новые сведения, полученные на модели, либо вообще не могут быть проверены на природном объекте (например, из-за невозможности воспроизведения условий на современном техническом уровне), либо требуют очень больших экономических затрат, вследствие чего особенно важен вопрос обоснования применяемых моделей (их аналогии с объектом по изучаемым признакам, их истинности). Именно неудовлетворительным обоснованием применяемых математических моделей и методов объясняется преимущественное использование для этих целей знаковых моделей (карт, схем, разрезов и колонок), подобие которых оригиналу не является ни прямым, ни косвенным, а устанавливается в результате соглашения о том, какое состояние реального объекта ставится в соответствие данному элементу абстрактной модели. Такое соглашение принимает вид совокупности правил построения моделей условного подобия и правил пользования ими [5].

Знаковые модели изоморфны, т.е. геометрически подобны оригиналу, обладая взаимно однозначным соответствием элементов и сходством отношений между ними. Важно отметить, что констатация наличия изоморфных аналогий не требует доказательств тождества даже основных существенных качеств. Так, в изоморфных аналогиях может отсутствовать материальное сходство отдельных элементов, но имеется сходство отношений между ними, т.е. в знаковых моделях сохраняются те же взаимоотношения между таксономическими единицами - "знаками", изображающими те или иные геологические образования и их свойства и их прототипами в природе. Например, на карте взаимоотношения между распространением отдельных элементов остаются теми же, что и на местности (на долю крупных промышленных зон и комплексов приходится, например, 60% площади подтопленных территорий и на местности и на карте). В то же время каждому из элементов на местности соответствует строго определенный элемент (обозначение) на карте. Это и есть взаимно однозначное соответствие, которое можно изобразить следующей схемой:



Элементам a, b, k, l в одной системе соответствуют вполне определенные элементы a, l, m, n в другой и наоборот.

Также следует отметить, что применяемые знаковые модели комплекса естественных и техногенных факторов подтопления имеют эвристическую и прогностическую направленность, применяясь для всевозможных прогно-

зных оценок современных геологических процессов. Полученные прогнозные оценки, отражают направление эволюции процесса (затухание, усиление, внезапное прекращение), стадийность, последовательность составляющих процессов (механизм общего процесса), постоянство проявления, частоту проявлений, продолжительность, интенсивность (так, например, скорость подтопления определяется как приращение площади территории с заданной глубиной уровня грунтовых вод за промежуток время Δt). Прогнозные оценки для территорий, на которых проявятся процессы, включают прогноз размеров территорий, зональных или азональных изменений интенсивности процесса, изменение свойств пород, рельефа, строения толщ и т.д.

Для моделирования комплекса естественных факторов подтопления предлагается использовать условные модели геологическо-литологических, геоморфологических, гидрогеологических и гидрографических характеристик исследуемых территорий [2, 4, 5].

В знаковых моделях геологических особенностей территории (на геологических картах) с помощью качественного фона (цветного или штрихового), буквенных, цифровых и других условных знаков отображают возраст, состав и происхождение горных пород, условия их залегания, характер границ между отдельными комплексами. Различают три основных вида условных знаков: цветовые, штриховые, буквенные и цифровые. Цветовые знаки служат для обозначения возраста осадочных, вулканогенных и метаморфических пород. Штриховыми знаками обозначается состав пород. Исключением являются интрузивные породы, при изображении которых цвет применяется для указания их состава. Буквенные и цифровые обозначения (индексы) служат для указания возраста и происхождения пород; состав интрузивных и некоторых вулканогенных пород обозначается также буквами.

Для каждой знаковой модели отражающей геологические особенности территории принято также строить стратиграфическую колонку и геологические разрезы. Стратиграфическая колонка представляет собой столбец шириной 2-4 см, в котором условной горизонтальной штриховкой в стратиграфической последовательности показывается положение и состав осадочных, вулканогенных и метаморфических пород, развитых на изображенной на карте территории, а также вскрытых буровыми скважинами. Породы в колонке расчленяются в соответствии с выделенными на карте стратиграфическими подразделениями. Слева от колонки указываются их возраст (система, отдел, ярус, свита и т.п.) и индексы; справа приводится мощность, описываются состав, встреченные окаменелости, геохимические параметры и физические свойства пород, их абсолютный возраст. Границы между стратиграфическими подразделениями с согласным залеганием на колонке указываются прямыми линиями, при стратиграфическом несогласии - волнистыми, при структурном - волнисто-угловатыми.

Геологические разрезы представляют собой вертикальные сечения земной коры от ее поверхности на ту или иную глубину. На разрезах изображаются условия залегания и форма геологических тел, изменения их

мощности, геологические структуры, различные фации, разрывы. Они составляются по геологическим картам, данным буровых скважин или геофизическим материалам. Разрезы по геологической карте строятся по прямым линиям в направлении, которое дает наиболее полное представление о залегании пород, слагающих изображенную на карте территорию. При необходимости разрезы могут строиться и по ломаным линиям. Если существуют опорные скважины, разрезы проводятся через них. Вертикальный и горизонтальный масштабы разрезов в большинстве случаев соответствуют масштабу карты. При пологом или горизонтальном залегании пород вертикальный масштаб разрезов принято увеличивать. Концы линии разреза и точки ее изгиба ограничиваются буквами. Такие же буквы ставятся у концов линии разреза на карте. Для каждой знаковой модели геологических особенностей территории обычно строятся 1-3 разреза. Все геологические границы на разрезах (согласные, несогласные и интрузивные) указываются одним знаком в виде сплошных линий. Глубина разреза обосновывается конкретными данными. На разрезах изображаются все породы, обнаженные на поверхности, вскрытые скважинами и установленные геофизическими методами, а также наиболее важные буровые скважины. Разрезы составляются в соответствии с условными обозначениями знаковой модели геологических особенностей территории.

Четвертичные образования отражаются в особых условных моделях четвертичных отложений, где коренные породы показывают одним цветом (фиолетовым) без расчленения, с указанием их возраста и типа с помощью индексов и крапа. Четвертичные отложения на картах разделяются по генезису, возрасту и составу; приводятся данные о мощностях стратиграфических подразделений, палеогеографические характеристики стратиграфических подразделений.

В условных моделях литологических особенностей территории (на литологических картах) на фоне окраски, соответствующей возрасту пород, штрихами изображается состав пород, выходящих на поверхность или скрытых под покровом четвертичных образований. На картах масштаб крупнее 1:10000 условной штриховкой обычно удается показать площади распространения каждой разновидности пород. На картах масштаба мельче 1:10000 указывается лишь состав преобладающих на данном участке пород либо при помощи смешанной штриховки отображаются отдельные разновидности пород.

Условные модели геоморфологических особенностей территории отражают основные типы рельефа и его отдельные элементы с учетом их происхождения и возраста.

Условные модели гидрогеологических особенностей территории отражают водоносные свойства горных пород: их водообильность, условия распространения, залегания, химическую характеристику и другие свойства подземных вод. В зависимости от указанных параметров горные породы объединяются в комплексы, которые изображаются на гидрогеологической карте в возрастной последовательности или по генетическому признаку.

Зеркало грунтовых вод на гидрогеологических картах показывают гидроизогипсами, линиями, соединяющими

точки одинаковой высоты поверхности грунтовых вод. Карты гидроизогипс составляются обычно как при гидрогеологических съемках, так и при целевых гидрогеологических изысканиях и являются основным элементом условной модели гидрогеологических характеристик территории, представляя собой основную площадную характеристику грунтового потока. При кратковременных исследованиях обычно ограничиваются составлением карты гидроизогипс по данным на определенную дату. В случае если велись продолжительные наблюдения по скважинам участка составляют несколько карт гидроизогипс на различные периоды года (в межень, паводок и т.п.). Такие карты позволяют выявить изменение условий питания и дренирования, связь поверхностных и подземных вод в разное время года.

При построении карты, сечение гидроизогипс (частота заложения гидроизогипс) выбирается таким образом, чтобы на карте были отражены особенности зеркала грунтовых вод. Для построения карты гидроизогипс пользуются данными замеров глубин залегания уровней грунтовых вод в скважинах, шурфах, колодцах, горных выработках, отметками источников, сведениями водомерных постов на поверхностных водоемах. Все данные, используемые при построении подобной условной модели, должны быть взяты на одну дату, т. е. получены по одновременным замерам всех точек наблюдения, поэтому карты гидроизогипс всегда датируются. В случае если имеются замеры уровня грунтовых вод в точках на различные даты, но которые можно использовать для построения карты гидроизогипс, нужно все замеры уровней привести к одной определенной дате. Это возможно, когда на участке проводились более или менее продолжительные наблюдения за колебаниями уровней грунтовых вод по нескольким скважинам, расположенным в аналогичных гидрогеологических условиях. В этом случае по данным колебаний в наблюдательных скважинах приблизительно вычисляют величину поправки уровня для каждой из скважин, с учетом которой изменяют в них уровни и строят ориентировочную карту гидроизогипс.

Глубина залегания грунтовых вод в каждой точке замера пересчитывается на абсолютные или относительные отметки

$$H_B = H_3 - h, \quad (1)$$

где H_B - абсолютная отметка уровня грунтовых вод; H_3 - абсолютная отметка поверхности земли; h - глубина залегания подземных вод.

Полученные согласно выражению (1) отметки наносятся на топографическую основу и по ним методом интерполяции строят гидроизогипсы.

Модели факторов техногенной группы представляют собой знаковые модели, отражающие характер функционального освоения территорий, где отдельные таксономические единицы выделяют в зависимости от характера освоения территорий (территории промышленных зон и комплексов, коммунально-складских зон, селитебные территории, сельскохозяйственные, рекреационного назначения и т.д.).

МНОГОКОМПОНЕНТНАЯ МОДЕЛЬ ИССЛЕДУЕМОЙ ТЕРРИТОРИИ

В результате сопоставления моделей воздействия источников подтопления и знаковых моделей естественных и техногенных факторов развития процесса подтопления (например, в качестве различных слоев электронной карты в ГИС) может быть получена многокомпонентная модель реальности, позволяющая исследовать объекты, явления и процессы с учетом динамики их развития. При анализе многокомпонентной модели ГИС существенное внимание следует уделить выработке принципов выделения таксономических единиц классификации различного уровня. Большинство рекомендаций по разработке методов геологических классификаций и формализации представлений сводится к следующему [2, 4]:

1. Таксономические единицы классификации одного уровня должны выделяться только по одному классификационному признаку.
2. Ни один из объектов классификации не должен входить одновременно в разные таксономические единицы одного уровня.
3. Не должно быть объектов, которые не входят ни в одну из таксономических единиц одного уровня.

4. Не должно быть таксономических единиц, в которых не имеется ни одного объекта. Исключение могут составлять таксоны, по которым делается прогноз.

5. Признаки построения классификации должны быть делящими, т.е. имеется по крайней мере две таксономические единицы объектов по данному признаку.

В этой связи целесообразно вернуться к опыту районирования защищаемых от подтопления и затопления территорий по совокупности естественно-геологических и техногенных факторов, накопленному УкрНИИИТИЗом в процессе инженерно-гидрогеологических работ по защите территорий промышленно-городских агломераций.

При таком способе выделения таксономических единиц (см. рис. 2) комплекс естественно-геологических факторов определяет выделение крупных таксонов: бассейнов грунтовых вод (бассейнов стока), фрагментов бассейнов и районов, причем последние два в возрастающей степени увязываются с факторами подтопления, а факторы техногенной группы (характер освоения территории) определяют выделение подрайонов и участков, что количественно определяет воздействие человека на геологическую среду. Следует отметить, что подобный подход к районированию территорий включает в себя элементы прогноза изменения состояния геологической среды.

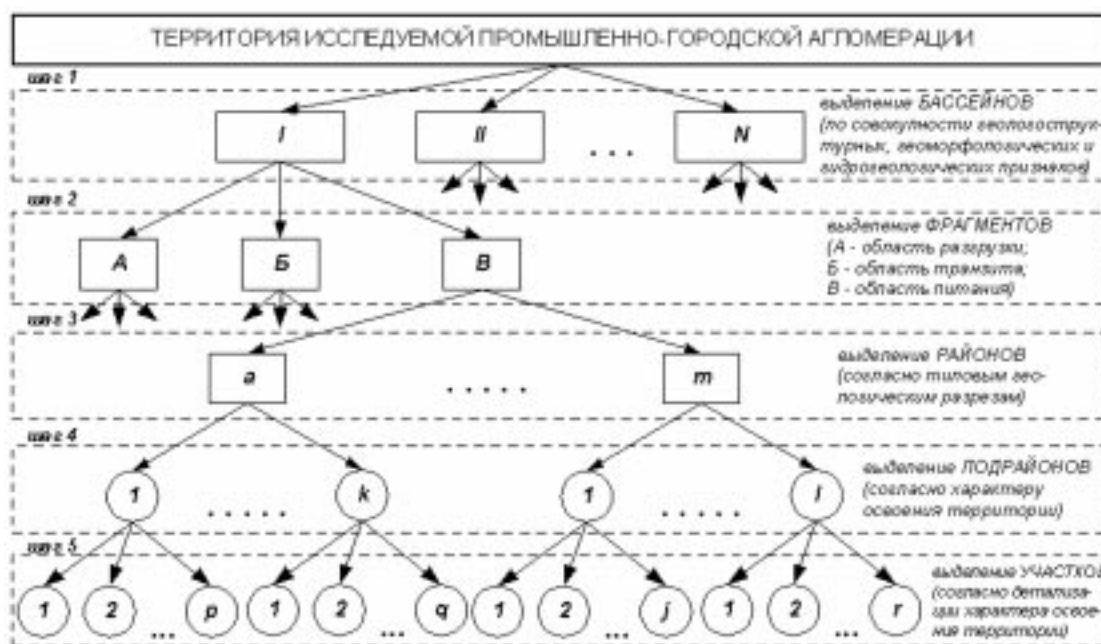


Рисунок 2 - Обобщенная схема алгоритма районирования территорий промышленно-городских агломераций

Бассейны выделяются по совокупности геолого-структурных, геоморфологических и гидрогеологических признаков и представляют собой участок (полосу) подземного потока от области питания до области разгрузки (местной дренажной сети). При этом с соседними бассейнами они граничат либо вдоль линии водораздела, либо вдоль линии тока, ориентированной нормально к местной речной сети. В пределах каждого бассейна выделяются три фрагмента, которые отвечают области разгрузки, транзита и

питания в пределах выделенного бассейна. Каждому из фрагментов соответствует свой комплекс факторов и свой механизм процесса подтопления, что в достаточной степени изучено. В пределах каждого фрагмента выделяются районы (отвечающие типизированным геологическим разрезам) в той или иной степени определяющие развитие процесса подтопления. Подрайоны, отвечающие функциональным зонам, выделяются по характеру освоения территорий (селитебная застройка, промышленная, сельско-

хозяйственная, рекреационная и т.д.). Наконец, участки, выделяемые в пределах каждого подрайона, отвечают детализации характера функционального освоения территории. Например, застройка одно- и многоэтажная, или промышленная зона с "сухим" или "мокрым" технологическим циклом.

**ОСОБЕННОСТИ РЕШЕНИЯ ЗАДАЧИ
ЗОНИРОВАНИЯ И ВИЗУАЛИЗАЦИИ ГРАНИЦ
НОВЫХ ОБЪЕКТОВ**

Одним из наиболее трудных вопросов при решении задач зонирования, т.е. задач связанных с построением новых объектов (зон до того на карте не существовавших) однородных в смысле некоторого критерия или группы критериев является вопрос проведения границ между отдельными таксономическими единицами. Границы зон могут либо совпадать с границами ранее существовавших объектов, либо строиться в результате различных видов моделирования. Дело осложняется тем, что многие классификационные признаки не формализованы и их количественная характеристика затруднительна. Вследствие чего затруднено и проведение четких границ между отдельными таксономическими единицами. В подобной ситуации более естественный способ описания таксонов всех уровней классификации заключается в ослаблении строгого разделения между ними, посредством использования формального аппарата для работы с неопределенностями неустраимого характера, предложенного Лотфи Заде и получившего название нечеткая логика (fuzzy logic). Тогда базисом для определения (описания) размытых границ отдельных таксонов будет являться функция принадлежности $\mu_i(u) \in [0;1]$, $\forall u \in U$, определяющая степень членства анализируемой территории u (из конечного множества территорий для которых выполняются мониторинговые наблюдения) в i -м нечетком классификационном подмножестве.

Подобный подход к решению задачи зонирования повышает адекватность используемых моделей и, как следствие, качество получаемых на моделях данных (естественно при наличии эффективного аппарата визуализации нечеткости, облегчающего понимание и восприятие географических объектов с размытыми границами). В качестве инструмента используемого для визуализации нечетких границ может быть использован цвет объектов, рассматриваемый как наиболее важная визуальная переменная. Тогда значения красный, зеленый, синий и т.д. могут быть интерпретированы как обозначения нечетких подмножеств универсального множества, и каждый объект может быть идентифицирован как некоторый цветовой оттенок с тем или иным цветом в центре (где $\mu_i(u) = 1$), окруженном более светлыми или более темными его тонами, определяемыми в соответствии с принципом визуального равенства [8]. При таком принципе визуализации объектов цвета изображающие одинаково различные друг от друга значения функции принадлежности должны быть восприняты одинаково различно, т.е.

$$\forall 1 \leq k, l, m, n \leq N: \text{если}$$

$$v_k - v_l = v_m - v_n \Rightarrow L(C_k, C_l) = L(C_m, C_n),$$

где k, l, m, n - четыре последовательных индекса цвета в цветном масштабе с числом цветов N ; - значение функции принадлежности, которое изображается i -м цветом; C_i - собственно значение i -го цвета; $L(C_p, C_j)$ - воспринятое различие (расстояние) между i -м и j -м цветами. Такой способ визуализации позволяет воспринимать размытые объекты, с различной светлотой цветовых тонов или с различной плотностью одного цвета, представляющими различную степень неопределенности (рис. 3).

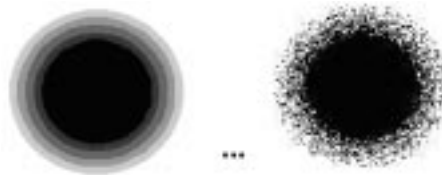


Рисунок 3 - Визуализация нечетких границ

ВЫВОДЫ

Проведенный в данной работе анализ моделей, которые предлагается использовать в качестве моделей-оснований многоуровневой иерархической модели анализа и прогнозирования подтопления, показал, что, несмотря на то, что в настоящее время они существуют разрозненно, в их основе не заложен принцип системности, и в таком виде они не позволяют решать задачи многоуровневого управления в системе экологического мониторинга, включение вышерассмотренных моделей в качестве моделей-оснований в многоуровневую иерархическую модель анализа и прогнозирования подтопления не является принципиально невозможным. Таким образом, с одной стороны, использование методов системного анализа для построения целостных систем оценки и прогнозирования опасного подъема уровня грунтовых вод в пределах ПГА и создания на их основе эффективных систем поддержки принятия управленческих решений адаптированных для работы региональных центров экологического мониторинга (РЦЭМ) является актуальной научной и практической проблемой. С другой стороны, использование таких систем в стратегическом плане перспектив социально-экономического развития региона, как элемента единой социально-экономической системы Украины и элемента жизненной сферы человека в целом, является насущной необходимостью.

ПЕРЕЧЕНЬ ССЫЛОК

1. Гальперин А.М., Зайцев В.С., Норватов Ю.А. Гидрогеология и инженерная геология: Учеб. для вузов. - М.: Недра, 1989 - 383с.
2. Инженерно-геологические прогнозы и моделирование: Учебн. пособие / Л.Б.Розовский, И.П. Зелинский. - Одесса, 1975. - 116с.
3. Митрохин С.А. Системный подход к исследованию проблемы оценки и прогнозирования опасного подъема

- уровня грунтовых вод. // Вісник СНУ ім. В.Даля. - 2002. - №3(49). - с.145-153.
4. Михайлов А.Е. Структурная геология и геологическое картирование. Учеб. пособие для вузов. - 4-е изд., перераб. и доп. - М.: Недра, 1984. - 464с.
 5. Основные требования к содержанию и оформлению обязательных геологических карт масштаба 1:50 000 (1:25 000).- Л.: Недра, 1977. - 119с.
 6. Перегудов Ф.И., Тарасенко Ф.П. Введение в системный анализ. - М.: Высшая школа, 1989.- 368 с.
 7. Рекомендации по выбору исходных данных для модели прогноза подтопления городских территорий / ПНИИИС Госстроя СССР. - М.: Стройиздат, 1986 - 136с.
 8. Jiang B., Brown A., Ormeling F.J. Some Perceptual Aspects of Colouring Uncertainty. // Advances in GIS II. London: Taylor&Francis, 1996. - p.477 - 490.
 9. Рубан С.А., Шинкаревський М.А. Гідрогеологічні оцінки і прогнози режиму підземних вод України за результатами спостережень: Довідково-методичний посібник. - Дніпропетровськ: ДВ УкрДІМР, 2000. - 144с.

УДК 658.14/.17:65.011.56

АВТОМАТИЗИРОВАННАЯ СИСТЕМА АНАЛИЗА ФИНАНСОВО-ЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ ДЕЯТЕЛЬНОСТИ ПРЕДПРИЯТИЯ

А.В.Пархоменко, Н.А.Кузовенкова, О.В.Шинкарук

В статті розглянута концепція розробки сучасної автоматизованої системи аналізу фінансово-економічних показників діяльності підприємства з використанням OLAP-технології.

В статті розглянута концепція розробки сучасної автоматизованої системи аналізу фінансово-економічних показників діяльності підприємства з використанням OLAP-технології.

The conception of investigation of modern computer-aided system of an enterprise activities financial and economical analysis using OLAP-technology is considered in this paper.

ВВЕДЕНИЕ

Эффективное управление любым предприятием, как правило, требует двусторонней направленности, которая заключается в отслеживании и анализе данных, отражающих реальное состояние этого предприятия и ситуацию, сложившуюся вокруг него. Постоянное наличие актуальной информации позволяет с точностью оценить текущее состояние этого предприятия, а отслеживание изменений некоторых его характеристик - выявить тенденции развития и прогнозировать их изменения в будущем. Одним из путей осуществления многогранного анализа информации является применение технологии OLAP (On-Line Analytical Processing) [1].

OLAP (оперативный анализ данных) - информационная технология, которая обеспечивает аналитикам, управленцам и руководителям возможность изучать информацию при помощи быстрого, последовательного, интерактивного отображения данных с различных возможных точек зрения, достигаемую за счёт преобразования исходных данных в соответствии с реальными представлениями конечного пользователя о пространстве предпринимательской деятельности. Технология OLAP позволяет творчески работать с данными и концентрировать внимание на их содержательной стороне: прибыли, продажах, продукции или поставщиках.

OLAP-технология основана на построении многомерных наборов данных - кубов. Оси куба (измерения) содержат параметры анализируемого процесса (например,

для кредиторской задолженности это могут быть товар, регион, поставщик, покупатель, время). Вдоль каждой оси данные могут быть организованы в виде иерархии, представляющей различные уровни их детализации. Благодаря такой модели данных пользователь может формулировать сложные запросы, генерировать отчеты, получать подмножества данных. Измерения образуют некое виртуальное пространство, в котором хранятся переменные, т.е. измерения являются базисом для организации переменных в виде многомерных кубов. Переменная - это объект, хранящий фактические данные (например, сумма по салдо кредитовое/дебетовое). Каждая переменная состоит из одного или более измерений. Для каждой допустимой комбинации значений может храниться только одно значение переменной [2,3].

Сегодня на рынке программного обеспечения представлено достаточное количество систем оперативной аналитической обработки данных (Hyperion Essbase OLAP, ORACLE EXPRESS, Microsoft OLAP Services и др.), сочетающих в себе простоту использования и аналитическую мощь. Проблема заключается в том, что стоимость таких систем пока еще остается очень высокой. Поскольку эффективное управление предприятием невозможно без их применения, актуальной является разработка собственных аналитических систем, ориентированных на потребности конкретного предприятия.

1 ПОСТАНОВКА ЗАДАЧИ

На любом современном предприятии возникает проблема такой организации управления данными, которая обеспечила бы наиболее эффективную работу. Для этого используют системы управления базами данных (СУБД), позволяющие хранить, извлекать информацию и управлять большими объемами данных. Так, в отделе автоматизированных систем управления предприятием "Запорожсталь" в качестве центра обработки данных используется СУБД Oracle [4,5]. Комплекс задач главной бухгалтерии комбината реализован с помощью СУБД Clipper. В качестве средства анализа данных широко используется Microsoft Excel [6-8].

Для решения задач главной бухгалтерии разработана специализированная автоматизированная система, которая обеспечивает операционную обработку данных, выполняет учёт всех финансово-хозяйственных операций, а также генерирует всю необходимую бухгалтерскую и налоговую отчётность. Существенным недостатком системы является то, что она не ориентирована на аналитическую обработку имеющейся информации. Однако на сегодняшний день проведение такого анализа является необходимостью, поскольку в процессе своей деятельности предприятие накопило значительные объёмы данных, которые хранят в себе большие потенциальные возможности по извлечению полезной информации, позволяющей оценивать текущее состояние этого предприятия, ситуацию, сложившуюся вокруг него, выявлять скрытые тенденции, строить стратегии развития и находить новые решения.

Целью данной работы является разработка современной автоматизированной системы анализа финансово-экономических показателей деятельности предприятия, позволяющей генерировать необходимую бухгалтерскую отчётность и осуществлять многогранный анализ и обработку данных.

Для достижения поставленной цели необходимо решить следующие задачи:

- разработать архитектуру проектируемой системы;
- разработать модели передачи данных между Clipper, Oracle, Excel;
- разработать интерфейс пользователя;
- выполнить программную реализацию подсистемы формирования данных и подсистемы анализа данных;

**2 КОНЦЕПЦИЯ РАЗРАБОТКИ
АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ
АНАЛИЗА ФИНАНСОВО-ЭКОНОМИЧЕСКИХ
ПОКАЗАТЕЛЕЙ ДЕЯТЕЛЬНОСТИ
ПРЕДПРИЯТИЯ**

**2.1 Разработка архитектуры системы и
моделей передачи данных**

Архитектура проектируемой системы показана на рис.1. В данной архитектуре Oracle и Clipper используются в качестве средств хранения данных, Excel - в качестве средства анализа и обработки данных, а Delphi выступает средством разработки интерфейса.

Проектируемая система состоит из двух подсистем - подсистемы формирования и подсистемы анализа данных. Подсистема формирования данных предназначена для организации доступа к таблицам Oracle и Clipper, установления связи между этими таблицами и построения многомерного куба, используемого в подсистеме анализа.

Для организации доступа к Oracle из приложения Delphi существует библиотека, содержащая компоненты прямого доступа к Oracle (Direct Oracle Access - DOA). Эти компоненты обеспечивают повышенную производительность и позволяют использовать возможности сервера Oracle наиболее простым способом. Прямой доступ к Oracle в проектируемой системе реализован следующими компонентами:

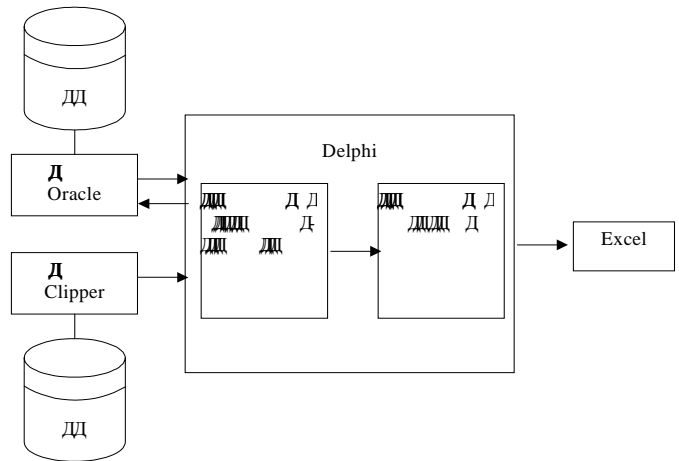


Рисунок 1 - Архитектура проектируемой системы

- TOracleSession - компонент, предназначенный для соединения с базой данных Oracle и контроля транзакций;
- TOraclePackage - компонент, обеспечивающий удобный интерфейс доступа к функциям, процедурам, константам, сохранённым в пакете;
- TOracleDataSet - компонент, с помощью которого можно управлять источником данных.

Для организации прямого доступа к СУБД Clipper из приложения Delphi существует специальная библиотека компонентов Halcyon. При разработке системы были использованы следующие компоненты:

- THalcyonDataSet - компонент, обеспечивающий интерфейс между dbf-файлами и компонентами Delphi;
- TCreateHalcyonDataSet - компонент, определяющий, dbf-файлы какой СУБД будут использоваться.

Таким образом, средства Delphi позволяют организовать доступ к таблицам Oracle и Clipper, а написание программного кода позволяет осуществить связь между этими таблицами.

Подсистема анализа данных предназначена для предоставления пользователю возможности проведения многогранного анализа данных и их обработки. Средством конечного анализа выступает MS Excel, так как именно этот программный продукт позволяет решать задачи аналитической обработки информации и широко используется на данном предприятии. Поскольку куб сформирован в Oracle, а анализ необходимо выполнять в Excel, для доступа к набору данных в Oracle из среды Excel разработаны макросы, позволяющие создавать сводные таблицы, содержащие данные OLAP-куба [9].

Сводная таблица является мощным аналитическим инструментом. При создании сводной таблицы определяется, какие из полей будут участвовать в формировании строк, столбцов и страниц, задаётся способ организации таблицы и тип выполняемых вычислений. При этом пользователи, анализирующие информацию, могут "разрезать" куб по разным направлениям, получать сводные или, наоборот, детальные сведения и осуществлять прочие манипуляции. Различные пользователи рассматривают кубы в разных разрезах и на разных уровнях детализации. Так, бухгалтер из отдела сбыта

может рассматривать кредиторскую задолженность в разрезе некоторого промежутка времени для различных рынков сбыта и различных видов продукции или изучать текущий и предыдущий временные периоды для всех рынков сбыта всех видов товаров. В то же время, бухгалтер из финансового отдела изучает кредиторскую задолженность в разрезе продукции, поставляемой комбинату внешними предприятиями по всем имеющимся балансовым счетам. Схематический взгляд на куб приведен на рис.2.

После того, как в сводной таблице Excel отобразили содержимое OLAP-куба, этим отображением можно манипулировать. Бухгалтер может изменить полученное расположение данных для просмотра под другим углом зрения.

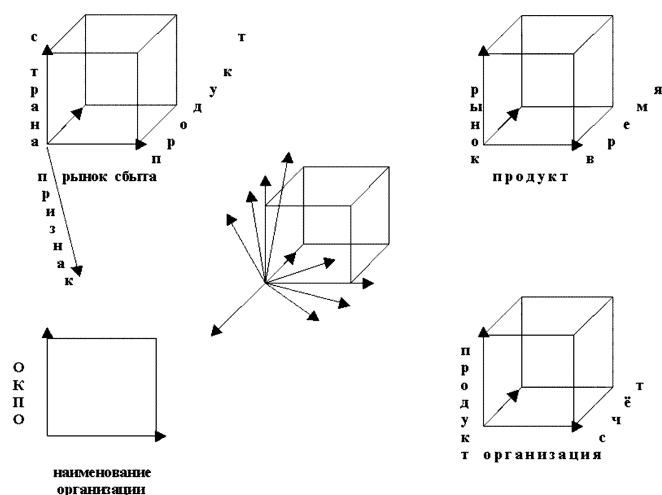


Рисунок 2 - Схематический взгляд на куб

Кроме того, что подсистема анализа данных позволяет осуществлять анализ и обработку данных, она также позволяет генерировать необходимую бухгалтерскую отчетность. Отчетность дополняет, расшифровывает те данные, которые содержатся в балансе. Она включает в себя, например, сведения о движении основных средств и фондов, о структуре доходов, о себестоимости и др., т.е. в различных отчетных формах отражается та или иная сторона деятельности предприятия.

Поскольку бухгалтерский учет строго документален, то каждый отчетный документ должен быть оформлен надлежащим образом, т.е. содержать необходимые реквизиты (наименование предприятия, наименование документа, номер, дата, содержание хозяйственной операции, её денежное выражение др.). Подсистема анализа данных в зависимости от выбранного пользователем вида документа формирует соответствующую отчетную форму. Отчетная форма представляет собой документ строго заданного образца. Пользователь определяет критерии, по которым должна отбираться информация для заполнения формы.

Подсистема анализа данных с помощью специальных макросов, разработанных в среде MS Excel, в зависимости от потребностей пользователя вытягивает из сформированного в Oracle куба необходимые пользователю данные. Такой документ может состоять из различных частей - колонтитулов, повторяющихся на каждой странице, заголовка, находящегося в начале отчета, областей, в которых располагаются фактические данные, областей, в которых расположены агрегатные данные (суммы, средние и т.д.), и др.

Кроме того, подсистема анализа данных предоставляет пользователю возможность формировать собственные уникальные отчеты, которые ему необходимы. При этом пользователям достаточно понимать структуру хранения данных и уметь пользоваться Excel. Преимуществом является то, что при этом не нагружается сервер баз данных и нет необходимости прибегать к помощи программистов, что является немаловажным достоинством системы.

2.2 Разработка интерфейса пользователя

Главная экранная форма системы представлена на рис.3. При помощи манипулятора мышь или клавиш табуляции и управления курсором бухгалтер выбирает: вид задолженности, критерий отбора, вид отчетного документа, период, за который необходимо сформировать отчетный документ. При выборе вида задолженности и вида документа бухгалтер может выбрать из предложенного набора только один вариант, а при выборе критерия отбора - один или несколько вариантов. Отчетный период пользователь может выбрать из выпадающего календаря или ввести вручную.

После ввода всех необходимых данных пользователю становится доступна кнопка "Просмотр". При нажатии данной кнопки система выполняет проверку входной информации на наличие ошибок и отслеживает сбойные ситуации.

Если при проверке на наличие ошибочных ситуаций таковых не были выявлены системой, то выполняется загрузка Excel и перед пользователем в зависимости от выбранного вида документа открывается соответствующая форма.

При нажатии на кнопку "Просмотр документа" система формирует отчетный документ установленного образца. Используя встроенные средства Excel, пользователь может редактировать полученные в отчете данные и выполнять над ними различные операции (суммирование, нахождение минимального и максимального значений и др.). При нажатии на кнопку "Сводная таблица" пользователь получает возможность работы со сводной таблицей, которая является мощным аналитическим инструментом и позволяет просматривать данные под разными углами зрения. При нажатии на кнопку "Сводная диаграмма", по сводной таблице выполняется построение сводной диаграммы.



Рисунок 3 - Главная экранная форма системы

РЕЗУЛЬТАТЫ И ВЫВОДЫ

Практическая ценность данной разработки заключается в том, что предложенная автоматизированная система позволяет повысить производительность работы бухгалтера, на 40% сократить временные затраты на решение задач по данной проблематике. Система прошла тестирование, внедрена в опытную эксплуатацию, функционирует надёжно.

УДК 681.3.06

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ ПРИМЕНЕНИЯ ПАРАЛЛЕЛЬНЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ ДЛЯ МОДЕЛИРОВАНИЯ ДВИЖЕНИЯ КОСМИЧЕСКОГО АППАРАТА

Н.Д.Пиза

В статті досліджується коефіцієнт прискорення та ефективність паралельних блочних однокрокових методів інтегрування для моделювання руху космічного апарата на паралельних обчислювальних системах з розподіленою пам'яттю і реалізацією обміну даними між процесорами за допомогою функцій бібліотеки MPI.

В статье исследуется коэффициент ускорения и эффективность паралельных блочных одношаговых методов интегрирования для моделирования движения космического аппарата на паралельных вычислительных системах с распределенной памятью и реализацией обмена данными между процессорами с помощью функций библиотеки MPI.

In the article the acceleration coefficient and the efficiency of the parallel block one-step integration methods are investigated on the parallel computing systems with distributed memory and implementation of the data exchange among the processors with the help of MPI library functions.

ВВЕДЕНИЕ

При разработке систем управления космических аппаратов (КА), для отработки и испытаний бортовых алгоритмов, программ и вычислительных комплексов используются комплексные и полунатурные моделирующие стенды. При этом возникает необходимость моделировать движение КА, работу его подсистем и процессы окружающей среды в реальном, а иногда и в ускоренном масштабах времени. В настоящее время для решения сложных вычислительных задач все чаще используются параллельные методы и средства вычислений [1]. Представляет интерес оценить возможность и эффективность использования параллельных вычислительных систем для моделирования движения КА.

Моделирование движения КА сводится к интегрированию обыкновенных дифференциальных уравнений (ОДУ). Как отмечают авторы работы [2], в последнее

Использование OLAP технологии является весьма перспективным, поскольку позволяет значительно облегчить процесс формирования аналитической информации, предоставляет возможность обращаться к данным, характеризующим предприятие и его рынки, анализировать эти данные для выявления сильных и слабых сторон в работе предприятия и в результате предпринимать соответствующие меры, а также прогнозировать возможности будущего развития.

ПЕРЕЧЕНЬ ССЫЛОК

1. Сахаров А.А. Концепция построения и реализации информационных систем, ориентированных на анализ данных.- Системы управления базами данных, 1996, №4. - С.55-70.
2. Сахаров А.А. Принципы проектирования и использования многомерных баз данных.- Системы управления базами данных, 1996, №3. - С.44-59.
3. Львов В. Создание систем поддержки принятия решений на основе хранилищ данных.- Системы управления базами данных, 1997, №3. - С.30-40.
4. Баженова И. Ю. Oracle 8/8i - М.: Диалог - МИФИ, 2000. - 304с.
5. Пейдж, Вильям Дж. Использование Oracle 8/8i / Пер. с англ.-М.: Издательский дом Вильямс, 2000.-1024 с.
6. Гарнаев А.Ю. Использование MS Excel и VBA в экономике и финансах. - СПб.: БХВ - Санкт-Петербург, 2000. - 336 с.
7. Долженков В.А., Колесников Ю.В. Microsoft Excel 2000. - СПб.: БХВ -Петербург, 2000. - 1088 с.
8. Карлберг К. Бизнес-анализ с помощью Excel: Пер. с англ. - К.: Диалектика, 1997. - 439 с.
9. Карпов Б. Visual Basic 6: Специальный справочник. - СПб.: Питер, 2000. - 416 с.

время появилось большое количество численных методов интегрирования ОДУ, которые ориентированы на вычислительные системы с параллельной архитектурой, однако практическое использование большинства методов не всегда оправдано, поскольку многие из них либо обладают численной неустойчивостью, либо имеют очень сложную структуру, приводящую к потере эффективности. Блочные методы интегрирования обладают "внутренним" параллелизмом, поскольку допускают одновременное вычисление интегрируемых функций в нескольких точках временной оси моделирования. В работе [3] предложена методика экспериментальной оценки точности параллельных блочных методов для интегрирования дифференциальных уравнений, описывающих модель движения КА.

ПОСТАНОВКА ЗАДАЧИ

Модель движения КА состоит из уравнений, описывающих движение центра масс КА по орбите и движение КА относительно центра масс. Модель движения КА относительно центра масс представляется совокупностью динамических и кинематических дифференциальных уравнений [4, 5]:

$$J\dot{\bar{\omega}} + \bar{\omega} \times J\bar{\omega} = \bar{M}, \quad (1)$$

$$Q = \frac{1}{2}(Q \circ \bar{\omega} - \bar{\omega}_0 \circ Q), \quad (2)$$

где $J = \|J_{nm}\|$ - матрица моментов инерции ($n, m = x, y, z$);

$\bar{\omega}$ - вектор угловой скорости КА в связанной системе координат;

\bar{M} - вектор главного момента сил, действующих на КА;

$Q = q_0 + q_1\bar{i}_1 + q_2\bar{i}_2 + q_3\bar{i}_3$ - кватернион, составленный из действительных элементов и трех мнимых единиц, формально совпадающих с осями базовой (орбитальной) системы координат; данный кватернион определяет взаимное положение орбитального и связанного базисов;

$\bar{\omega}_0$ - кватернионное отображение вектора орбитальной скорости КА, заданного проекциями на оси орбитальной системы координат;

\circ - знак умножения кватернионов.

Математическая модель движения центра масс КА по орбите имеет вид [6]:

$$\begin{aligned} \ddot{x} - \omega_3^2 x - 2\omega_3 \dot{y} &= \mu \frac{\partial U}{\partial x}, \\ \ddot{y} - \omega_3^2 y + 2\omega_3 \dot{x} &= \mu \frac{\partial U}{\partial y}, \\ \ddot{z} &= \mu \frac{\partial U}{\partial z}, \end{aligned} \quad (3)$$

где $r = (x \ y \ z)^T$ - радиус-вектор, задающий положение КА в гринвичской системе координат;

μ - гравитационная постоянная Земли, ω_3 - угловая скорость вращения Земли;

U - возмущающая часть гравитационного потенциала Земли.

Структурная схема моделирования движения КА представлена на рисунке 1. Здесь $\Omega, \omega_{II}, f_{orb}, i_{orb}, e, a$ - Кеплеровы элементы орбиты; $\varphi(0), \vartheta(0), \psi(0)$ - начальные значения углов Эйлера, определяющие положение КА относительно орбитальной системы координат; r и V - положение и скорость центра КА в гринвичской системе координат; U - гравитационный потенциал Земли; B^B - вектор геомагнитной индукции Земли в проекциях на оси связанной системы координат; L - управляющий магнитный момент; M_B - суммарный возмущающий момент.

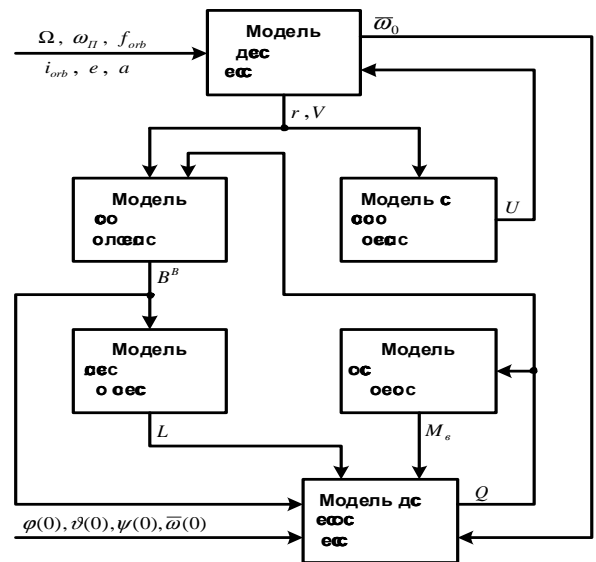


Рисунок 1 - Структурная схема моделирования движения КА

Представляет интерес исследовать эффективность и коэффициент ускорения, получаемые при использовании параллельных блочных методов и параллельных вычислительных систем для моделирования движения КА.

ПАРАЛЛЕЛЬНЫЕ БЛОЧНЫЕ ОДНОШАГОВЫЕ МЕТОДЫ

Будем рассматривать одношаговые двухточечный и четырехточечный параллельные блочные методы для решения задачи Коши: $\dot{y} = f(x, y)$, $y(0) = y_0$. Блочным называется метод, в котором для блока из k точек новые k значений функции вычисляются одновременно. Эта особенность методов согласуется с архитектурой парал-

льных вычислительных систем. Метод является одношаговым, поскольку для расчета значений в новом блоке используется только последняя точка предшествующего блока. Пусть y_n - приближенное значение решения задачи Коши с начальным условием в точке x_n - начальной точке обрабатываемого блока. Предполагается, что в пределах одного блока точки сетки значений независимой переменной x находятся на равных расстояниях: $x_{n+r} = x_n + rh$, где $r = \overline{1, k}$, h - шаг сетки.

Параллельный блочный одношаговый метод четвертого порядка точности $O(h^4)$ с двухточечным блоком реализуется формулами [7]:

$$y_{n+r,0} = y_n + rhf_n, \quad r = 1, 2,$$

$$\left. \begin{aligned} y_{n+1,s+1} &= y_n + \frac{h}{12}(5f_n + 8f_{n+1,s} - f_{n+2,s}) \\ y_{n+2,s+1} &= y_n + \frac{h}{3}(f_n + 4f_{n+1,s} + f_{n+2,s}) \end{aligned} \right\} s = 0, 1, 2. \quad (4)$$

Как видно из структуры приведенных рекуррентных формул, вычисления значений функций $f_{n+1,s}$, $f_{n+2,s}$, и получаемых решений $y_{n+1,s+1}$, $y_{n+2,s+1}$, можно выполнять параллельно. Таким образом, решение в n -м блоке получается сразу в двух точках: n_1, n_2 - это, соответственно $y_{1,3}$, $y_{2,3}$, причем, значение $y_{2,3}$ является начальным для следующей итерации, т.е. выступает в качестве y в $n+1$ -м блоке в приведенных выше формулах. На рисунке 2 показана условная схема процесса итераций.

Параллельный блочный одношаговый метод шестого порядка точности $O(h^6)$ с четырехточечным блоком реализуется формулами [7]:

$$y_{n+r,0} = y_n + rhf_n, \quad r = \overline{1, 4},$$

$$\left. \begin{aligned} y_{n+1,s+1} &= y_n + \frac{h}{720}(251f_n + 646f_{n+1,s} - \\ &\quad - 264f_{n+2,s} + 106f_{n+3,s} - 19f_{n+4,s}) \\ y_{n+2,s+1} &= y_n + \frac{h}{90}(29f_n + 124f_{n+1,s} + \\ &\quad + 2f_{n+2,s} + 4f_{n+3,s} - f_{n+4,s}) \\ y_{n+3,s+1} &= y_n + \frac{3h}{80}(9f_n + 34f_{n+1,s} + \\ &\quad + 24f_{n+2,s} + 14f_{n+3,s} - f_{n+4,s}) \\ y_{n+4,s+1} &= y_n + \frac{2h}{45}(7f_n + 32f_{n+1,s} + \\ &\quad + 12f_{n+2,s} + 32f_{n+3,s} + 7f_{n+4,s}) \end{aligned} \right\} s = \overline{0, 4}. \quad (5)$$

Здесь также вычисления значений функций $f_{n+1,s}$, $f_{n+2,s}$, $f_{n+3,s}$, $f_{n+4,s}$ и получаемых решений $y_{n+1,s+1}$, $y_{n+2,s+1}$, $y_{n+3,s+1}$, $y_{n+4,s+1}$ можно вы-

полнять параллельно. В данном методе решение в n -м блоке получается сразу в четырех точках: n_1, n_2, n_3, n_4 - это, соответственно, $y_{1,5}$, $y_{2,5}$, $y_{3,5}$, $y_{4,5}$ причем, значение $y_{4,5}$ является начальным для следующей итерации, т.е. выступает в качестве y в $n+1$ -м блоке в приведенных выше формулах. Условная схема процесса итераций показана на рисунке 3.]

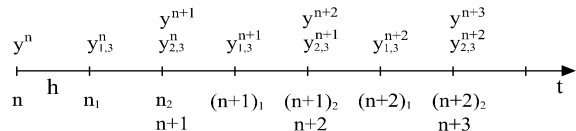


Рисунок 2 - Схема процесса итераций двухточечного блочного метода

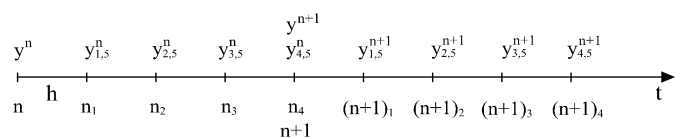


Рисунок 3 - Схема процесса итераций четырехточечного блочного метода

ИССЛЕДОВАНИЕ БЫСТРОДЕЙСТВИЯ ПАРАЛЛЕЛЬНЫХ БЛОЧНЫХ МЕТОДОВ

Для сравнения и оценки качества алгоритмов параллельных вычислений обычно используют такие характеристики, как коэффициент ускорения и эффективность [8]. Ускорение определяется следующим соотношением:

$$S = \frac{T_1}{T_n}, \quad (6)$$

где T_1 - время выполнения программы одним процессором;

T_n - время выполнения программы процессорами.

Эффективность параллельного алгоритма определяется как:

$$E = \frac{S}{n}. \quad (7)$$

Безусловно, эффективная реализация параллельных вычислений может быть достигнута лишь при совместном рассмотрении используемых параллельных численных методов, архитектуры параллельной вычислительной системы и технологии параллельного программирования. Будем ориентироваться на архитектуру многопроцессорных систем с распределенной памятью, поскольку они являются оптимальным вариантом создания параллельных вычислительных систем с точки зрения соотношения производительность/стоимость. Как правило, в таких системах отсутствует единое адресное

пространство и для обмена данными между параллельными процессорами используется модель передачи сообщений через коммуникационную среду. В качестве технологии параллельного программирования будем использовать известный интерфейс для организации обмена сообщениями в параллельных вычислительных системах - MPI [1]. Таким образом, будем исследовать ускорение и эффективность параллельных блочных одношаговых методов для моделирования движения КА на параллельных вычислительных системах с распределенной памятью и реализацией обмена данными между процессорами с помощью функций технологии MPI.

Как отмечалось ранее, в блочных методах для блока из k точек новые k значений функции вычисляются одновременно. Так, согласно формулам (4) и (5) для реализации двухточечного метода можно использовать два процессора, а для четырехточечного - четыре процессора, каждый из которых будет производить обработку своей точки блока на каждой итерации. Алгоритмы соответствующих методов с учетом их реализации в рамках модели передачи сообщений представлены на рисунках 4 и 5.

Из рисунков 4 и 5 видно, что на каждой итерации для двухточечного параллельного метода необходимо выпол-

нить 7 обменов данными, а для четырехточечного - 63 обмена. Обмены вызывают дополнительные накладные расходы в параллельной вычислительной системе, отсутствующие при последовательном выполнении программы.

По нашей оценке, полученной экспериментально, вычислительная сложность правых частей приведенных выше уравнений модели движения КА эквивалентна 1×10^4 вычислительных операций умножения чисел типа double. Для построения более полных моделей движения КА (см. рисунок 1), необходимо учитывать и другие модели, в частности, динамическую модель атмосферы, модель движения Солнца, звездное время и т.д. Это приводит к усложнению правых частей системы ОДУ. В результате, даже на параллельных системах с относительно низкой скоростью обмена между процессорами время вычисления правых частей уравнений может оказаться большим по сравнению со временем, необходимым на обмен данными. Таким образом, при достижении некоторого порогового значения вычислительной сложности правых частей уравнений становится эффективным использование параллельных блочных методов.

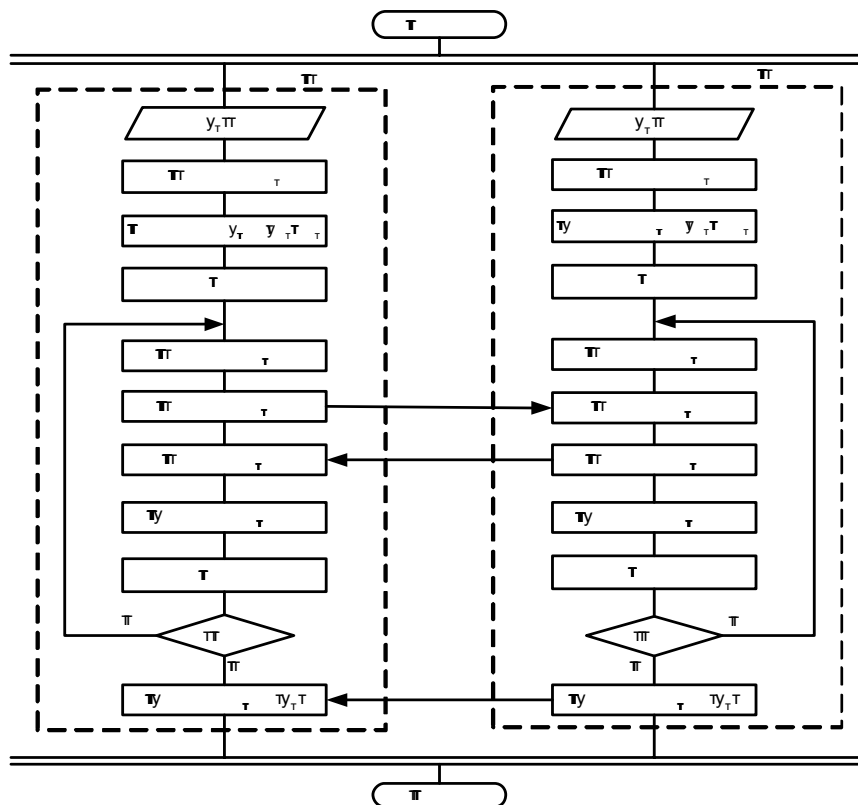


Рисунок 4 - Алгоритм итерации параллельного блочного двухточечного метода

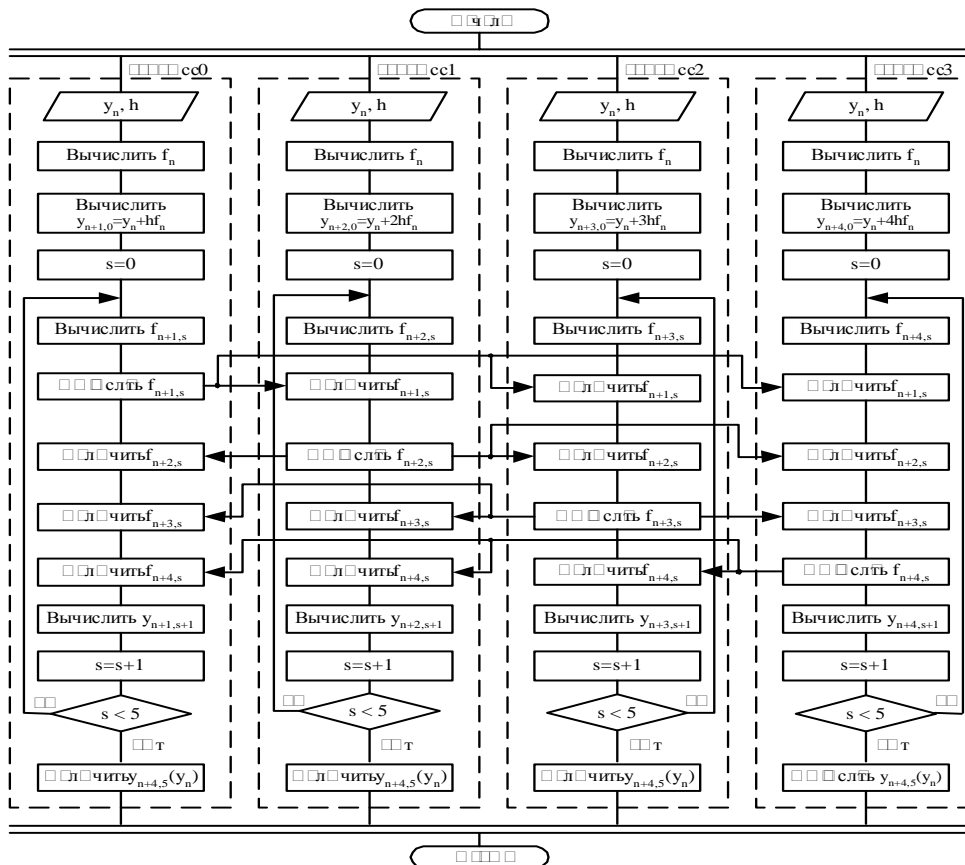


Рисунок 5 - Алгоритм итерации параллельного блочного четырехточечного метода

Для моделирования движения КА и экспериментального исследования ускорения и эффективности параллельных блочных методов интегрирования разработана программа на языке C++ с использованием библиотеки MPI реализации MPICH (MPI CHameleon) 1.2.4.

Эксперименты проводились на двух параллельных системах. В качестве первой параллельной системы использовалась компьютерная сеть типа Fast Ethernet с рабочими станциями Intel Pentium IV (1.76 ГГц) одной из компьютерных лабораторий Запорожского национального технического университета. В качестве второй параллельной системы был использован SCI-кластер Петродворцового телекоммуникационного центра¹. SCI-кластер имеет следующие характеристики: число узлов - 20, число процессоров на узле - 2, тип процессоров - Intel Pentium III (933 МГц), объем ОЗУ на узле 512 Мб; способ соединения узлов - однонаправленные кольца SCI для передачи данных в системе MPI и сеть Fast Ethernet для управления машинами.

Эксперименты заключались в измерении затрат времени на моделирование 10000 секунд полета КА с шагом

интегрирования 1 секунда при различной вычислительной сложности правых частей уравнений.

С целью сравнения результатов, получаемых при использовании SCI-связей и связей посредством общей памяти, на SCI-кластере были проведены два варианта экспериментов: single и double. В варианте single на каждом узле кластера использовался только один процессор, т.е. программа запускалась на двух и четырех узлах, по одному процессу на каждом узле (в предположении, что используются однопроцессорные машины). Соответственно, двухпроцессная задача использовала одну связь SCI, а четырехпроцессная - три связи SCI. В варианте double на каждом узле кластера использовались оба процессора, т.е. двухпроцессная задача использовала только один узел и обмен данными реализовывался внутри узла, а четырехпроцессная задача использовала два узла с одной SCI-связью между ними.

Результаты вычислительных экспериментов приведены в таблице 1 и позволяют определить условия, при которых моделирование движения КА эффективно на параллельных вычислительных системах таких классов.

1. Автор выражает свою признательность доцентам кафедры вычислительной физики Санкт-Петербургского государственного университета Комолкину Андрею Владимировичу и Немнюгину Сергею Андреевичу за содействие в организации экспериментов на SCI-кластере Петродворцового телекоммуникационного центра

Таблица 1 - Результаты вычислительных экспериментов

| Количество вычислительных операций $\times 10^4$ | Время вычислений при применении блочного метода, секунды | | | | | | | | | |
|--|--|---------------------------|--------------|------------|------------|------------------|---------------------------|--------------|------------|------------|
| | двухточечный | | | | | четырёхточечный | | | | |
| | последовательный | | параллельный | | | последовательный | | параллельный | | |
| | Net | SCI-single/ SCI-double | Net | SCI-single | SCI-double | Net | SCI-single/ SCI-double | Net | SCI-single | SCI-double |
| 1 | 3,23 | 3,80 | 5,11 | 3,46 | 3,11 | 4,86 | 7,84 | 8,83 | 2,93 | 2,83 |
| 2 | 6,34 | 4,54 | 6,93 | 3,87 | 3,55 | 9,48 | 8,97 | 10,14 | 3,26 | 3,10 |
| 2,5 | 8,02 | 4,93 | 7,92 | 4,11 | 3,77 | 11,82 | 9,54 | 10,80 | 3,43 | 3,29 |
| 3 | 9,48 | 5,30 | 8,69 | 4,32 | 3,98 | 14,13 | 10,09 | 11,50 | 3,59 | 3,44 |
| 5 | 15,62 | 6,81 | 12,25 | 5,20 | 4,85 | 23,46 | 12,36 | 14,18 | 4,24 | 4,16 |
| 7 | 21,84 | 8,30 | 15,80 | 6,09 | 5,71 | 32,73 | 14,31 | 16,89 | 4,92 | 4,82 |
| 10 | 31,16 | 10,55 | 21,15 | 7,42 | 7,03 | 46,70 | 17,72 | 20,88 | 5,90 | 5,83 |
| 26 | 80,98 | 22,62 | 49,59 | 14,47 | 13,92 | 121,37 | 35,82 | 42,06 | 11,22 | 11,06 |
| 51 | 158,27 | 41,44 | 93,65 | 25,54 | 24,75 | 237,40 | 64,05 | 75,35 | 19,55 | 19,56 |
| 76 | 236,19 | 60,25 | 137,75 | 36,55 | 35,56 | 354,50 | 92,32 | 108,67 | 27,87 | 27,90 |
| 101 | 313,90 | 79,10 | 181,71 | 47,62 | 46,35 | 471,18 | 120,59 | 141,76 | 36,15 | 36,26 |
| 151 | 469,18 | 116,80 | 269,64 | 69,21 | 68,06 | 704,40 | 177,04 | 208,23 | 52,79 | 52,87 |

Для удобства ссылок на результаты экспериментов введены следующие обозначения: Net - результаты, полученные на первой параллельной системе, SCI-single и SCI-double - результаты, полученные на второй параллельной системе в вариантах single и double соответственно.

Для измерения затрат на обмены были применены тесты скорости обменов, разработанные в НИВЦ МГУ [9]. С помощью этих тестов определены общие затраты времени на выполнение обменов в двухточечном и четырёхточечном методах для рассматриваемых вычислительных систем, результаты представлены в таблице 2.

Таблица 2 - Затраты времени на обмен данными

| Блочный метод | Время, секунды | | |
|-----------------|----------------|------------|------------|
| | Net | SCI-single | SCI-double |
| двухточечный | 2,113 | 0,208 | 0,043 |
| четырёхточечный | 8,192 | 1,027 | 0,926 |

Как видно из представленных результатов, применение параллельных блочных методов становится эффективным при решении задачи моделирования движения КА на первой параллельной системе, когда вычислительная сложность правых частей уравнений превышает $2,5 \times 10^4$ вычислительных операций. Применение параллельных блочных методов для решения данной задачи на SCI-кластере эффективно уже в начальной постановке задачи, без усложнения правых частей уравнений. На рисунках 6 и 7 показаны зависимости коэффициента ускорения и

эффективности параллельного решения задачи моделирования движения КА от количества вычислительных операций. На рисунках 6 и 7 используются принятые ранее обозначения Net, SCI-single и SCI-double, а цифры 2 и 4 указывают блочный метод - двухточечный и четырёхточечный соответственно.

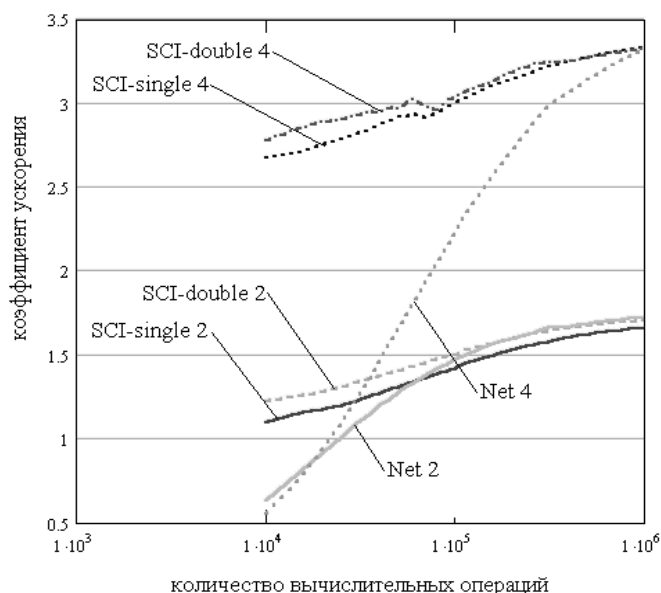


Рисунок 6

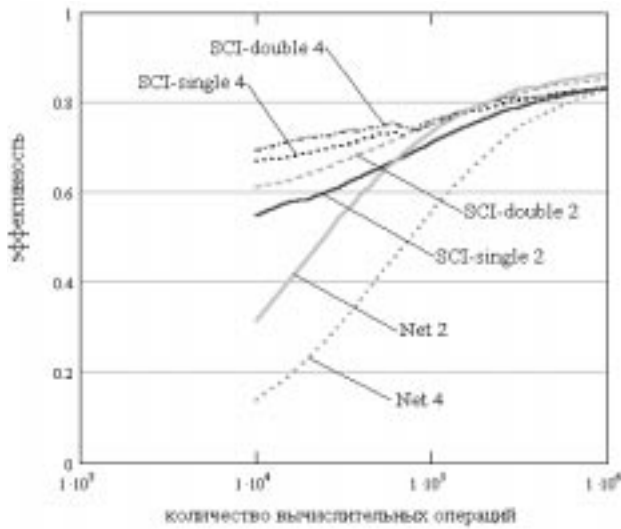


Рисунок 7

ВЫВОДЫ

Таким образом, моделирование движения КА на параллельных вычислительных системах с помощью параллельных блочных методов интегрирования может быть эффективным по точности [3] и эффективным по быстродействию при определенных значениях произво-

дительности процессоров, коммуникационных средств и сложности модели. Результаты исследований предполагается использовать при создании программно-аппаратных средств моделирования для наземной отработки и испытаний системы управления КА.

ПЕРЕЧЕНЬ ССЫЛОК

1. Воеводин В.В., Воеводин Вл.В. Параллельные вычисления.- СПб.: БХВ-Петербург, 2002. - 608 с.
2. Фельдман Л.П., Дмитриева О.А. Разработка и обоснование параллельных блочных методов решения обыкновенных дифференциальных уравнений на SIMD-структурах // Наукові праці Донецького національного технічного університету. Вип. 29. - Донецьк, 2001. С. 70-79.
3. Пиза Н.Д., Кудерметов Р.К. Применение параллельных блочных методов для моделирования движения космического аппарата // Вісник технологічного університету Поділля. - 2003. - № 3. т. 1. - С. 93-96.
4. Алексеев К.Б., Бебенин Г.Г. Управление космическими летательными аппаратами. - М.: Машиностроение, 1974. - 300 с.
5. Бранец В.А., Шмыглевский И.П. Применение кватернионов в задачах ориентации твердого тела. - М.: Наука, 1973. - 320 с.
6. Справочное руководство по небесной механике и астродинамике / Абалкин В.К., Аксенов Е.П., Гребенников Е.А. и др. Изд. 2-е.-М.: Наука, 1971.-584 с.
7. Системы параллельной обработки / Под ред. Д. Ивенса. - М.: Мир, 1985. -416 с.
8. Молчанов И.Н. Введение в алгоритмы параллельных вычислений / АН УССР, Институт кибернетики им. В.М. Глушкова. - К.: Наукова думка, 1990. -127 с.
9. Андреев А.Н., Воеводин В.В. Методика измерения основных характеристик программно-аппаратной среды (www.dvo.ru/bbc/benchmarks.html).

УДК 514.1/8:519.8/71

КОНФИГУРАЦИЯ ОБЪЕКТА КАК ИНТЕРФЕЙС МЕЖДУ МАТЕМАТИЧЕСКИМ И ФИЗИЧЕСКИМ СОДЕРЖАНИЕМ МОДЕЛИ

В.П.Сизиков

Установлено, что конфигурация объекта или поля является физически значимой характеристикой, а для адекватного учета конфигурации объекта необходимо держать ориентир на ее 3-мерную пространственную протяженность. Картина развернута на примере движения объекта в физическом поле, в том числе, космического аппарата с тросой системой.

It is established that the configuration of an object or a field is a physically meaningful characteristic and to adequately consider the configuration of an object it is necessary to be orientated on it's 3-dimensional spatial extent. The picture is developed on the example of movement of an object in the physical field including the space apparatus with cable system.

ВВЕДЕНИЕ

В [1] М.М. Лаврентьев справедливо указывает, что в математической физике XX в. значительное место занимают модели с недостаточно осмысленным содержанием, а процедуры моделирования в основном сводятся к подгонке теоретических положений под эмпирический материал. Это же относится и к другим предметным

областям, использующим математические методы. Перенос математических методов не сопровождается должной онтологической проработкой и адаптацией моделей систем, с которыми работает конкретная тематика. По сути, специалисты, работая с предметными областями, не имеющими онтологической проработки, применяют математические формализмы в ранге "серых ящиков". Сложилась характерная для естествознания ситуация, когда между теоретической и прикладной математикой отсутствует особый интерфейс, задачи которого - в онтологической проработке математических моделей.

Потребности к формированию указанного интерфейса у автора возникли после получения и проработки им серии результатов по тематике управления [2-4]. Вначале традиционно главное внимание уделялось учету понятий полных управляемости и наблюдаемости для систем управления, вырабатывались соответствующие критерии в терминах полноты ранга подходящих матриц. Но когда такие критерии удалось обобщить на разные варианты систем и форм управления, выяснилось, что выпо-

лнимость условий полной управляемости и наблюдаемости в системе никак не связана ни с моделью описания системы на языке уравнений, ни с формой управления. Другими словами, эти условия выступают характеристиками системы управления как единого физического объекта. Потенциально возможным оказалось и управление движением механического объекта без знания самого уравнения движения [3], проблема оставалась за преодолением стохастичности, возникающей при реализации управления [4]. Предстояло определиться, где и как в системах управления должна учитываться физическая сущность объектов, а где остается полная свобода за математическим аппаратом.

Первые шаги к формированию указанного интерфейса в ранге теории динамических информационных систем (ТДИС) [5] были сделаны на базе онтолого-метафизического подхода, использующего аппарат категориально-системной методологии [6]. К настоящему времени определились также логико-топологический и аналитико-реалистичный подходы к становлению ТДИС. Сам аппарат ТДИС продолжает расширяться и совершенствоваться, затрагивая проработку и физических понятий [7-10].

ТДИС не раз уже "призывала" к необходимости учета конфигурации объекта при изучении физических явлений, и то, насколько этот "призыв" оказался актуальным, демонстрируется на примере движения объекта в физическом поле. Рассуждения и комментарии ведутся в постановке, аналогичной задаче управления движением космического аппарата (КА). Показывается, какие качественно новые моменты и горизонты выявляет учет конфигурации объекта при его движении в физическом поле.

1 МЫСЛЕННЫЙ ЭКСПЕРИМЕНТ

По традиции все расчеты и теории, связанные с движением объекта в физическом поле, изначально предполагают рассмотрение этого объекта в ранге материальной точки, не имеющей никаких размеров и, тем более, конфигурации. Даже в задачах, изучающих изменение ориентации объекта и управление ей [2], траектория движения объекта считается изначально определенной на условиях отождествления объекта с точкой. Но оказывается, что с учетом конфигурации объекта, его протяженности в пространстве может существенно измениться и картина его движения в физическом поле.

Чтобы четче представить последний момент, проведем мысленный эксперимент, в котором происходит мгновенное превращение объекта из точки в однородный стержень. Пусть для определенности в роли физического поля выступает ньютоново поле тяготения. Тогда, пока объект предстает точкой, его движение происходит, как известно, по кеплеровой траектории. Но вот объект превратился в стержень. Что произойдет с его движением далее?

Если стержень ориентирован перпендикулярно радиусу-вектору, проведенному из центра поля тяготения в центр стержня, являющийся также центром его масс, то, очевидно, все части стержня, кроме его центра, будут удалены от центра поля тяготения несколько больше, чем сам центр стержня. Последнее означает, что сила притяжения стержня в целом окажется несколько меньшей, чем

это было бы при отождествлении объекта с точкой. А тогда объект, имея скорость, согласованную изначально с кеплеровой траекторией, должен начать удаление от центра поля тяготения, переход на более высокую траекторию. Если же стержень будет ориентирован вдоль вышеуказанного радиуса-вектора, то, наоборот, сила притяжения стержня в целом окажется несколько большей, чем это было бы при отождествлении объекта с точкой, и объект должен начать приближение к центру поля тяготения, переход на более низкую траекторию.

В общем случае объект, конечно, имеет вращение, не идущее в такт с его обращением вокруг центра поля тяготения. И при наличии асимметрии в конфигурации объекта его центр масс будет двигаться не по кеплеровой траектории, а по некоей волнообразной линии, чередуя акты приближения и удаления по отношению к центру поля тяготения. При относительно мощном физическом поле, малых расстояниях и высоких скоростях картина движения объекта и вовсе будет напоминать облако в форме кольца или сферического слоя вокруг центра поля тяготения, что и наблюдается на примере движения электрона вокруг ядра атома.

Обычно, ввиду малости размеров КА по сравнению с его расстоянием от центра поля тяготения, вышеуказанные отклонения в движении оказываются незначительными. Не существенны они и при решении задач управления ориентацией КА, так как управляющие воздействия формируются, исходя непосредственно от показателей текущей ориентации КА, и влияние параметров орбиты КА здесь оказывается довольно грубым. Однако, при наличии, например, тросовой системы (ТС) пространственная протяженность всего комплекса КА становится на порядки больше и, как следствие, возрастают отклонения в движении и их влияние на ориентацию КА. Так что задачи с ТС требуют более тонкого изучения движения комплекса КА, берущего на учет и его конфигурацию. Но тогда важно учитывать и фактор влияния частей комплекса КА друг на друга на уровне гравитации. Особенно это относится к ТС, не являющейся абсолютно твердым телом, причем, здесь важно учитывать и самогравитацию ТС. Однако тут есть свои парадоксы. Адекватного учета не состоится, если, например, ТС принимать за нить, обязательно необходимо рассмотрение ТС в ракурсе 3-мерного тела.

Проанализируем ситуации с учетом конфигурации объекта и с парадоксом самогравитации конкретнее и подробнее.

2 АВТОНОМНОСТЬ В ПОВЕДЕНИИ ВИХРЕВОГО КОМПОНЕНТА

Серия опытов, в которых наблюдалось поведение ТС относительно корпуса КА, как правило, сопровождалась накручиванием ТС на корпус КА. Сложилось мнение, что виной всему этому является непонятная динамика ТС. Но верно ли такое заключение?

Обратимся к рассмотрению пока плоской модели движения объекта в ньютоновом поле тяготения. Здесь используем общеизвестный в механике метод Лагранжа для составления уравнений движения системы в целом. Пусть имеем некоторый фиксированный базис из двух

векторов e_1, e_2 , жестко связанных с полем тяготения и неподвижный относительно звезд. Аналогично, пара векторов n_1, n_2 жестко связана с объектом как абсолютно твердым телом. Тогда движение объекта представится в терминах векторной переменной $R(t)$ как радиуса-вектора, проведённого из центра поля тяготения в центр масс объекта, и угловой переменной $\varphi(t)$, указывающей поворот второго базиса относительно первого. При этом

$$n_1 = \cos\varphi e_1 + \sin\varphi e_2, \quad n_2 = -\sin\varphi e_1 + \cos\varphi e_2.$$

В терминах этих переменных кинетическая и потенциальная энергия объекта будут равны соответственно

$$T = 0,5(M\dot{R}^2 + J\dot{\varphi}^2), \quad \Pi = -G \int_{\Omega} \frac{dm}{|R + \rho_{1m}n_1 + \rho_{2m}n_2|}.$$

Здесь дополнительно обозначены: M - масса объекта, J - момент его инерции относительно центра масс, G - напряженность поля тяготения, dm - элемент массы объекта, ρ_{1m} и ρ_{2m} - его координаты в собственном базисе объекта, Ω - область распределения массы по объекту. Применение формализма Лагранжа приводит в итоге к системе уравнений:

$$\begin{aligned} M\ddot{R} + G \int_{\Omega} \frac{(R + \rho_{1m}n_1 + \rho_{2m}n_2)dm}{|R + \rho_{1m}n_1 + \rho_{2m}n_2|^3} &= 0, \\ J\dot{\varphi} - G \int_{\Omega} \frac{\langle R, -\rho_{1m}n_1 + \rho_{2m}n_2 \rangle dm}{|R + \rho_{1m}n_1 + \rho_{2m}n_2|^3} &= Q, \end{aligned} \quad (1)$$

где \langle, \rangle - знак скалярного произведения векторов, а Q выражает возможное управляющее воздействие на вращение объекта.

Далее преобразуем систему (1), представив радиус-вектор R как в полярной системе координат, то есть через его длину $r = |R|$ и угол поворота ν в исходном неподвижном базисе из векторов e_1, e_2 . При этом координаты радиуса-вектора равны соответственно $R_1 = r \cos \nu, R_2 = r \sin \nu$. А сама система уравнений для переменных r, ν, φ примет вид:

$$\begin{aligned} \ddot{r} - r\nu^2 + \frac{G}{M} \int_{\Omega} \{ [r + \rho_{1m} \cos(\nu - \varphi) + \rho_{2m} \sin(\nu - \varphi)] / [r^2 + \rho_{1m}^2 + \rho_{2m}^2 + 2r(\rho_{1m} \cos(\nu - \varphi) + \rho_{2m} \sin(\nu - \varphi))]^{1,5} \} dm &= 0, \\ 2i\dot{\nu} + r\ddot{\nu} + \frac{G}{M} \int_{\Omega} \{ [-\rho_{1m} \sin(\nu - \varphi) + \rho_{2m} \cos(\nu - \varphi)] / \{ r^2 + \rho_{1m}^2 + \rho_{2m}^2 + 2r(\rho_{1m} \cos(\nu - \varphi) + \rho_{2m} \sin(\nu - \varphi)) \}^{1,5} \} dm &= 0, \\ \dot{\varphi} - \frac{Gr}{J} \int_{\Omega} \{ [-\rho_{1m} \sin(\nu - \varphi) + \rho_{2m} \cos(\nu - \varphi)] / \{ r^2 + \rho_{1m}^2 + \rho_{2m}^2 + 2r(\rho_{1m} \cos(\nu - \varphi) + \rho_{2m} \sin(\nu - \varphi)) \}^{1,5} \} dm &= \frac{Q}{J}. \end{aligned} \quad (2)$$

Исключая в последних двух уравнениях системы одинаковый интеграл, приходим к соотношению:

$$(Mr^2\dot{\nu} + J\dot{\varphi})' = Q \quad \text{или} \quad \dot{\varphi} = J^{-1}(\int Q dt - Mr^2\dot{\nu}), \quad (3)$$

где через $\int Q dt$ обозначена первообразная по времени функции воздействия Q , которую можно считать заранее известной, либо как-то выражающейся через переменные r, ν, φ и их производные по времени.

Соотношение (3) указывает на фундаментальную связь между скоростями вращения объекта вокруг собственного центра масс и его обращения вокруг центра поля тяготения. Изменение одной из этих величин должно компенсироваться соответствующим изменением другой. В этом плане можно предположить, что природа поля тяготения не исчерпывается его проявлениями на уровне потенциала, но на учет должен браться еще и некий вихревой компонент поля, проявляющий автономность на уровне закона сохранения суммарного вращательного момента у объекта и поля. Тем самым имеем подтверждение адекватности онтологической проработки любого физического поля [5]. Заодно получает определенное подкрепление сформированная О.Г. Ефименко [1] обобщенная теория гравитации.

Особенно важно учитывать соотношение (3) в масштабах микромира, где относительно мощное физическое поле, малые расстояния и высокие скорости движения объекта. По сути, соотношение (3) является обобщением закона сохранения момента импульса, учитывающим управляющие воздействия. Однако, такую связь невозможно было учесть, когда объект отождествлялся с материальной точкой, и собственное вращение объекта считалось, фактически, нулевым. В свою очередь, расширение практики в микромире скоро должно было натолкнуться на нарушения фундаментальных законов при недоучете такого вращения. Однако, вместо того, чтобы обратиться к учету конфигурации объекта, предпочтение было отдано просто постулированию у объекта какого-то момента вращения в ранге спина, проявляющего себя на уровне квантования с признаками случайной величины.

Недоучет указанного момента с вращением имеет место и при работе с космическими объектами, в частности, с КА. Так, перемены в динамике вращения КА могут быть весьма существенными даже при малых смещениях КА по направлению радиуса-вектора его орбиты. В частности, такие перемены неизбежны при выпуске ТС с КА, а, как следствие этого, скоро происходит наматывание ТС на корпус КА. Так что винить в таких явлениях исключительно динамику ТС не правомерно.

3 ПРИМЕРЫ И ОЦЕНКИ ВЛИЯНИЯ КОНФИГУРАЦИИ ОБЪЕКТА НА ПАРАМЕТРЫ ЕГО ОРБИТЫ

В рамках приведенной выше простейшей модели уместно любое воздействие на объект рассматривать как задание определенного режима изменения во времени у функции Q (1)-(3). Наиболее интересным представляется

случай, когда за счет такого режима достигается стабилизация в ориентации объекта по отношению к силовым линиям поля тяготения, а именно, обеспечивается условие $v - \varphi = C_0 - \text{const}$. Тогда

$Q = [(Mr^2 + J)\dot{v}]^*$. Поскольку, однако, воздействие осуществляется, согласно (2), через переменную φ , то представляется возможным регулировать и выбор постоянной C_0 . В связи с этим рассмотрим подробнее случай, когда $C_0 = 0$, т.е., фактически, вектор n_1 собственного базиса объекта оказывается сонаправленным радиусу-вектору R траектории центра масс объекта.

После фиксации режима воздействия и значения C_0 третье уравнение в (2) становится излишним и задача движения объекта сводится к первым двум уравнениям для переменных r и v , которые принимают вид:

$$\begin{aligned} \ddot{r} - r\dot{v}^2 + \frac{G}{M} \int_{\Omega} \frac{(r + \rho_{1m})dm}{[(r + \rho_{1m})^2 + \rho_{2m}^2]^{1,5}} &= 0, \\ 2r\dot{v} + r\ddot{v} + \frac{G}{M} \int_{\Omega} \frac{\rho_{2m}dm}{[(r + \rho_{1m})^2 + \rho_{2m}^2]^{1,5}} &= 0. \end{aligned} \quad (4)$$

Предположим далее, что интеграл во втором уравнении из (4) равен нулю. Такое автоматически сбывается, когда, например, масса объекта распределяется симметрично относительно оси, задаваемой вектором n_1 и проходящей через центр масс объекта. После этого второе уравнение в (4) принимает вид $2r\dot{v} + r\ddot{v}$, откуда получаем $(r^2\dot{v})^*$ или соотношение

$$r^2\dot{v} = C_1 - \text{const}, \quad (5)$$

известное как один из законов Кеплера для планет. Но а с учетом (5) первое уравнение в (4) становится определяющим для переменной r , принимая вид

$$\ddot{r} - \frac{C_1^2}{r^3} + \frac{G}{M} \int_{\Omega} \frac{(r + \rho_{1m})dm}{[(r + \rho_{1m})^2 + \rho_{2m}^2]^{1,5}} = 0. \quad (6)$$

Если в (6) полностью пренебрегать размерами объекта, т.е. считать $\rho_{1m} = \rho_{2m} = 0$, то в результате получится обычная картина движения объекта по кеплеровой траектории. Мы же пренебрегать размерами не будем, но для простоты и определенности рассмотрим два крайних случая вырождения конфигурации объекта в стержень: а) когда считаем всюду $\rho_{1m} = 0$; б) когда считаем всюду $\rho_{2m} = 0$. При этом в случае а) можно сказать, что объект ориентирован перпендикулярно силовым линиям поля тяготения, а в случае б) - параллельно силовым линиям поля тяготения.

Итак, в случае а) для переменной r получается уравнение

$$\ddot{r} - \frac{C_1^2}{r^3} + \frac{G}{Mr^2} \int_{\Omega} \frac{dm}{(1 + r^{-2}\rho_{2m}^2)^{1,5}} = 0, \quad (7)$$

которое легко интегрируется, приводя к уравнению

$$\dot{r}^2 = C_2 - \frac{C_1^2}{r^2} + \frac{2G}{M} \int_{\Omega} \frac{dm}{(r^2 + \rho_{2m}^2)^{0,5}}, \quad (8)$$

с подходящей постоянной C_2 . Учитывая, что C_1/Γ выражает тангенциальную скорость движения объекта, можно также записать

$$\dot{R}^2 = C_2 + \frac{2G}{M} \int_{\Omega} \frac{dm}{(|R|^2 + \rho_{2m}^2)^{0,5}}. \quad (9)$$

Главное отличие от случая точечного объекта, когда имеем уравнение

$$\dot{R}^2 = \tilde{C}_2 + \frac{2G}{|R|} \quad (10)$$

с несколько иной постоянной \tilde{C}_2 , состоит в том, что изменение полной скорости объекта при изменении его расстояния от центра поля тяготения оказывается в варианте (9) несколько меньшим, чем в варианте (10). Как следствие, движение объекта при учете его конфигурации в варианте (9) приводит к уменьшению кривизны траектории по сравнению с кеплеровой траекторией, определяемой вариантом (10). В частности, если начать отсчет движения объекта от наиболее удаленной точки $r = a$ эллиптической траектории из варианта (10), то в варианте (9) объект всегда будет находиться дальше от центра поля тяготения, чем ближайшая к этому центру точка $r = b$ эллиптической траектории. Если же начать отсчет движения объекта от наиболее близкой к центру поля тяготения точки $r = b$ эллиптической траектории из варианта (10), то в варианте (9) объект сможет уйти несколько дальше от центра поля тяготения, чем наиболее удаленная от этого центра точка $r = a$ эллиптической траектории.

Далее, примем во внимание, что для наиболее дальней и близкой от центра поля тяготения точки траектории объекта характерно отсутствие в ней радиальной составляющей скорости, т.е. выполняется условие $\dot{r} = 0$. Это, применительно к случаю точечного объекта, выраженного в (10), дает $\tilde{C}_2 - C_1^2 r^{-2} + 2Gr^{-1} = 0$ при $r = a$ и $r = b$. Исключая здесь постоянную \tilde{C}_2 , получим $C_1^2 G^{-1} = 2ab(a+b)^{-1}$. Обозначим далее через a_1 и b_1 соответственно наибольшее и наименьшее расстояние от центра поля тяготения для точек траектории объекта, отвечающей уравнениям (8) и (9). Так для них имеем аналогичное соотношение

$$\frac{C_1^2}{G} = 2 \int_{\Omega} \frac{a_1^2 b_1^2}{\sqrt{a_1^2 + \rho_{2m}^2} \sqrt{b_1^2 + \rho_{2m}^2} (\sqrt{a_1^2 + \rho_{2m}^2} \sqrt{b_1^2 + \rho_{2m}^2})} \frac{dm}{M}.$$

Принимая за начало отсчета движения положение в точке $r = a$, так что при этом и $a_1 = a$, мы получим

соотношение, из которого можно будет сравнить значения b и b_1 :

$$\frac{b}{a+b} = 2 \int \frac{ab_1^2}{\Omega \sqrt{a^2 + \rho_{2m}^2} \sqrt{b_1^2 + \rho_{2m}^2} (\sqrt{a^2 + \rho_{2m}^2} \sqrt{b_1^2 + \rho_{2m}^2})} dm \cdot M. \quad (11)$$

Очевидно, при $b_1 = b$ подынтегральная функция, а тогда и вся правая часть, в соотношении (11) оказывается меньше его левой части. Это указывает на то, что обязано быть $b_1 > b$. В свою очередь, если через $l > 0$ обозначить наибольшее значение для ρ_{2m} , отражающее габариты объекта, то легко проверить, что при $b_1 = b + 0,5l^2b^2 \times (a^{-1} + b^{-1})(a^{-2} + a^{-1}b^{-1} + b^{-2})$ подынтегральная функция, а тогда и вся правая часть, в соотношении (11) оказывается, наоборот, больше его левой части. И это указывает на то, что обязано быть

$$b < b_1 < b + 0,5l^2b^2(a^{-1} + b^{-1})(a^{-2} + a^{-1}b^{-1} + b^{-2}) \quad (12)$$

или просто $b < b_1 < b + 3l^2b^{-1}$. Аналогично (12) получается оценка

$$a < a_1 = a + 0,5l^2a^2(a^{-1} + b^{-1})(a^{-2} + a^{-1}b^{-1} + b^{-2}), \quad (13)$$

когда отсчет движения ведется от точки $r = b$.

Наконец, в случае б) для переменной r получается уравнение, отличающееся от (7) лишь заменой подынтегральной функции на $(1 + r^{-1}\rho_{1m})^{-2}$. А после интегрирования получится уравнение, отличающееся от (8) заменой подынтегральной функции на $(r + \rho_{1m})^{-1}$. Не теряя сути, упростим несколько ситуацию, предположив, что масса в объекте распределена симметрично также относительно оси, задаваемой вектором n_2 и проходящей через центр масс объекта. С учетом этого на смену уравнениям (8) и (9) придут соответственно

$$\begin{aligned} \dot{r}^2 &= C_2 - \frac{C_1^2}{r^2} + \frac{2Gr}{M} \int \frac{dm}{\Omega r^2 - \rho_{1m}^2}, \\ \dot{R}^2 &= C_2 + \frac{2G|R|}{M} \int \frac{dm}{\Omega |R|^2 - \rho_{1m}^2}. \end{aligned} \quad (14)$$

Теперь изменение полной скорости объекта при перемене его расстояния от центра поля тяготения оказывается в варианте (14), наоборот, несколько большим, чем в варианте (10), т.е. имеет место увеличение кривизны траектории в варианте (14) по сравнению с (10). И если начать отсчет движения объекта от наиболее удаленной точки $r = a$ эллиптической траектории из (10), то в (14) объект достигнет большего приближения $r = b_2$ к центру поля тяготения, чем ближайшая к этому центру точка $r = b$ эллиптической траектории, т.е. происходит сокращение размеров у траектории. По аналогии с вариантом (8-9) получаем оценку для b_2 :

$$b - l^2b^2(a^{-1} + b^{-1})(a^{-2} + a^{-1}b^{-1} + b^{-2}) < b_2 < b, \quad (15)$$

или просто $b - 6l^2b^{-1} < b_2 < b$, где $l > 0$ обозначает наибольшее значение для ρ_{1m} , отражающее габариты объекта. Аналогично (15) получается оценка

$$a - l^2a^2(a^{-1} + b^{-1})(a^{-2} + a^{-1}b^{-1} + b^{-2}) < a_2 < a, \quad (16)$$

когда отсчет движения ведется от точки $r = b$.

Как показывают соотношения (12-13) и (15-16), наличие размеров у объекта сказывается на значениях его расстояния от центра поля тяготения довольно слабо, внося в них изменения, не превышающие, как правило, размеров самого объекта. Однако, пренебрегать этим уместно, когда объект в целом не велик по размерам. Такое, например, допустимо при работе с КА в отсутствии ТС. А когда КА наделен ТС с характерным размером в 5 километров, изменение параметров орбиты КА может составлять десятки метров, что, в свою очередь, может приводить, согласно (3), к изменениям во вращении КА до нескольких оборотов в минуту, накручивая в итоге ТС на корпус КА.

А вот характерной причиной того, что картина движения может предстать в форме облака, установленный момент вряд ли является. На деле здесь существенен другой момент, известный в астрономии как факт смещения перигелия орбиты планеты.

4 ФАКТ СМЕЩЕНИЯ ПЕРИГЕЛИЯ ОРБИТЫ

Как известно, в варианте (10) радиус-вектор траектории оказывается периодической с периодом 2π функцией угла своего поворота ν . Это обеспечивается за счет квадратичной зависимости от r^{-1} в поведении \dot{r}^2 . Но это свойство, как правило, теряется при переходе к общему варианту (8-9). Так, рассматривая r как функцию от угла ν и учитывая (5), из (8) получаем

$$\frac{dr}{\sqrt{r^2 C_2 + \frac{2Gr}{M} \int \frac{dm}{\Omega \sqrt{1 + r^{-2} \rho_{2m}^2}} - 1}} = d\nu, \quad (17)$$

откуда сразу понятно, что радиус-вектор траектории оказывается периодической функцией от ν с периодом, равным удвоенному значению определенного интеграла от левой части (17) на отрезке $b < r < a_1$. Причем, концы отрезка интегрирования характеризуются тем, что в них обращается в нуль знаменатель левой части (17). При точном объекте под радикалом в (17) оказывается квадратный трехчлен $f(r) = -r^2 a_1^{-1} b_1^{-1} + r(a_1^{-1} + b_1^{-1}) - 1$, при котором значение определенного интеграла равно π , несмотря на произвольность a_1, b_1 . А в общем случае функция под радикалом в (17) оказывается, как нетрудно

установить, всюду на отрезке $b_1 < r < a_1$ несколько больше, чем $f(r)$, так что искомый определенный интеграл при варианте (8-9) оказывается несколько меньше π , а искомый период - несколько меньше 2π . Грубая относительная оценка дополнительных слагаемых позволяет говорить о разнице в периоде на величину около $3JM^{-1}a_1^{-2}$ или порядка $l^2a_1^{-2}$. Другими словами, в варианте (8-9) ситуация больше ассоциирует с феноменом медленного поворота в целом траектории объекта вокруг центра поля тяготения, причем, поворот происходит в направлении, обратном направлению движения объекта по этой траектории.

Аналогично, феномен медленного поворота в целом траектории объекта имеет место и в варианте (14). Правда, величина периода здесь оказывается, наоборот, больше 2π на величину около $3JM^{-1}a_2^{-2}$ или порядка $l^2a_2^{-2}$, т.е. направление поворота траектории совпадает с направлением движения по ней объекта.

Следует отметить, что факт смещения перигелия орбиты вовсе не предполагает влияния каких-либо других объектов, но вполне сбывается и для одиночного объекта в поле тяготения. Более того, именно этот факт, когда мощность поля и скорость движения в нем объекта достаточно велики, как правило, приводит к вырождению картины движения объекта в некое кольцевидное облако.

Указанные моменты, несмотря на малость изменений в параметрах орбиты и времени обращения по ней, важно учитывать при работе с КА, особенно при наличии ТС. Так, при характерном размере ТС в 5 километров годовой полет КА с такой ТС должен приводить к сдвигу перигелия орбиты на несколько градусов дуги даже при идеально сферической Земле. Если, значит, не делать достаточно частых сеансов связи с таким КА, то велика вероятность его скоро и вовсе потерять среди аналогичных собратьев.

5 К УЧЕТУ КОНФИГУРАЦИИ ФИЗИЧЕСКОГО ПОЛЯ

Как следствие феномена сокращения размеров эллиптической траектории при варианте (14), важно обратить внимание на следующий момент. Поскольку источником поля тяготения выступает тоже некий базовый объект с характерными для него размерами, то при варианте (14) оказывается более вероятным, чем при варианте (10), факт выпадения движущегося объекта на поверхность базового объекта. Кроме того, вслед за сокращением размеров траектории в варианте (14), по сравнению с (10), более быстрым окажется приближение объекта к центру поля тяготения, и, значит, более скорым сам факт выпадения этого объекта на поверхность базового объекта. Правда, разница в вероятности выпадения и продолжительности падения между вариантами (14) и (10) ничтожно мала, причем, ничуть не меньшее значение в этом могут сыграть неровности на поверхности базового объекта. Однако, этот момент делает актуальным сказать несколько слов относительно

модели, в которой на смену формальному физическому полю приходит обычный базовый объект, имеющий свои относительно большие массу и момент вращения.

Разумеется, ничего нового не добавится, если у базового объекта будет иметь место сферически симметричное распределение массы. Разве что вместо обычной массы исходного движущегося объекта придется взять, так называемую, приведенную массу двух тел. А вот отсутствие симметрии у конфигурации базового объекта будет уже существенно сказываться на параметрах орбиты исходного движущегося объекта. Размеры орбиты при этом, как правило, сравнимы с размерами базового объекта, и пренебрегать ими нет смысла. Изложенный выше материал показывает, как такие влияния могут быть учтены в теории и на практике.

Сказанное относится и к работе с КА в поле тяготения Земли. Нарушения сферичности в распределении массы в Земле автоматически приводят к описанным эффектам флуктуации движения КА по орбите и смещения перигелия орбиты КА, даже когда сам КА имеет идеально сферическую конфигурацию. Вместе с тем, можно поставить и обратную цель - формирование карты распределения массы в базовом объекте по результатам наблюдений за движением пробных объектов. В частности, это применимо к Земле с использованием КА.

6 ПАРАДОКСЫ САМОГРАВИТАЦИИ

Принятие постулата о всеобщности закона тяготения Ньютона позволяет обратиться к мысленному эксперименту, в котором некий объект, расположенный в вакууме, полностью отдан влиянию самогравитации. Если объект представляет формальное объединение нескольких изолированных друг от друга материальных точек, то для каждой из них можно без проблем выписать вполне осмысленную формулу результирующей силы притяжения к остальным точкам. Проблемы начинаются, когда объект предстает сплошным образованием без пустот. Приходится сталкиваться с ситуациями, когда фиксированный бесконечно малый элемент массы имеет бесконечно близко расположенных к нему соседей, так что при вычислении результирующей силы притяжения будут встречаться слагаемые, представляющие результат деления произведения двух бесконечно малых масс на квадрат бесконечно малого расстояния между ними. И эти слагаемые могут в ряде случаев приводить к расходимости гравитационного потенциала.

Когда элемент массы расположен внутри объекта, он подвержен притяжению со всех концов, и тогда действующие на этот элемент силы могут в сумме компенсировать друг друга. Поэтому для выявления расходимости будем заранее предполагать, что исследуемый элемент массы расположен на границе объекта так, что все остальные элементы массы находятся по одну сторону от некоторой фиксированной плоскости, проходящей через исследуемый элемент. Для простоты вычислений рассмотрим и сопоставим между собой случаи, когда объект имеет форму k -мерного ($k = 3, 2, 1$) полушара радиуса r , причем, масса в объекте

распределена равномерно с плотностью p , а исследуемый элемент единичной массы располагается в центре основания полушара. Примем во внимание, что при $k = 2$ объект имеет форму полукруга, а при $k = 1$ - отрезка, на одном из концов которого находится исследуемый единичный элемент массы.

В случае $k = 3$ координаты произвольного элемента массы dm полушара зададим тремя переменными: широтой θ , долготой ψ и расстоянием s от центра основания полушара. Тогда элемент массы $dm = ps^2 \cos\theta d\theta d\psi ds$, а его вклад в вектор ускорения, порождаемый силами притяжения, для исследуемого единичного элемента массы составит величину $dg = \gamma p \cos\theta \sin\theta d\theta d\psi ds$, где γ - гравитационная постоянная. После интегрирования по области полушара получим значение $g = \gamma \pi p r$ или $g = 1,5\gamma M r^{-2}$, где M - масса объекта. В случае $k = 2$ не будет переменной ψ и расчеты дадут соответственно $dm = ps \cos\theta d\theta ds$, $dg = \gamma p \cos\theta \sin\theta d\theta s^{-1} ds$, $g = 0,5\gamma p \ln s \Big|_{s=0}^r = \infty$. А при $k = 1$ не будет еще и переменной θ , так что расчеты дадут соответственно $dm = p ds$, $dg = \gamma p s^{-2} ds$, $g = -\gamma p s^{-1} \Big|_{s=0}^r = \infty$.

Как видно, напряженность поля тяготения в эффекте самогравитации объекта оказывается конечной величиной лишь в случае $k = 3$, т.е. когда объект учитывается как 3-мерное образование. Часто приходится иметь дело с объектами, представляющими по форме тонкие пластины или даже стержни. Однако вычисление напряженности, привносимой их самогравитацией, исходя из таких форм, приводит к расходимостям. Обязательно необходимо восприятие объектов как 3-мерных образований. В этом плане, изучение плоских моделей движения объектов в физическом поле не всегда оправдано. В частности, нельзя признать адекватной плоскую модель движения ТС, а вслед за этим, и плоскую модель движения КА с ТС.

Далее, в общем случае при каждой точке на границе объекта можно выделить некоторую достаточно малую область в форме соответствующего полушара, почти целиком расположенную в объекте. Так что напряжение поля самогравитации объекта в такой точке будет в главном не меньше, чем у выделенного полушара. При фиксированном значении плотности напряженность поля растет вслед за радиусом такого полушара. Однако на внешнем уровне объект часто характеризуется величинами своей массы и пространственной протяженности, а плотность оказывается лишь производной от них, так что более определяющей для напряжения поля самогравитации может оказаться формула $g = 1,5\gamma M r^{-2}$, где зависимость от радиуса является обратной. В качестве примера рассмотрим случаи цилиндрической ТС с фиксированными массой и длиной, где варьируется величина толщины ТС. Эта величина, очевидно, выступит в роли радиуса r полушара для центральных точек концов ТС, значение плотности в ТС будет изменяться как r^{-2} , а

масса полушара - как r . Так что напряжение поля самогравитации оказывается при этом изменяющейся как r^{-1} , т.е., чем тоньше ТС, тем больше напряжение поля ее самогравитации. При вырождении ТС в бесконечно тонкую нить приходим к бесконечно большим значениям напряжения поля ее самогравитации. Нечто подобное наблюдается и в задачах электродинамики, когда объекты имеют заостренные концы.

ЗАКЛЮЧЕНИЕ

Следует, прежде всего, добавить, что проведенные исследования для случая поля тяготения могут быть повторены и для полей отталкивания. Для них тоже справедливы факты, аналогичные только что установленным для полей тяготения.

Фактически, установлено, что конфигурация объекта или поля является физически значимой характеристикой, при недоучете которой велик риск получения неверных результатов и выводов. В частности, не бывает только потенциальных или только вихревых физических полей, но оба таких компонента есть у любого физического поля с признаками перераспределения баланса между ними. А для адекватного учета конфигурации необходимо держать ориентир на ее 3-мерную пространственную протяженность, понижение размерности пространства в модели должно иметь обоснования.

ПЕРЕЧЕНЬ ССЫЛОК

1. Лаврентьев М.М. Физические теории (математические модели), адекватные реальности - необходимое условие прогресса естествознания XXI века // Поиск математических закономерностей Мироздания: физические идеи, подходы, концепции: Избр. тр. Третьей Сибир. конф. ФПВ-2000. - Новосибирск: ИМ СО РАН, 2001. - С.5-28.
2. Сизиков В.П. Стабилизация положения космического аппарата с помощью двух гироскопов стабилизаторов // Космические исследования, 1996. - Т.54, № 1. - С.66-72.
3. Сизиков В.П. Модель дискретного управления движением космического аппарата // Спутниковая связь: Докл. 2-й Междун. конф. - М.: МЦНТИ, 1996. - Т.1. - С.109-114.
4. Сизиков В.П. О реализуемости модели адаптивного управления движением космического аппарата // Управление движением и навигация летательных аппаратов: Сб. тр. VIII Всерос. науч.-техн. сем. - Самара: Сам. филиал Академии космонавтики, 1998. - С.169-171.
5. Разумов В.И., Сизиков В.П. Математические и философские основы теории динамических информационных систем. Уч. пособие. - Поддержка ФЦП "Интеграция", ИМ-4. - <http://newasp.omskreg.ru/tdis>.
6. Разумов В.И. Категориально-системные методы в подготовке научных кадров / Интерактивная интернет-монография. <http://www.ic.omskreg.ru/~cognitiv/>.
7. Сизиков В.П. Моделирование распределения температур в дискретной среде на базе теории динамических информационных систем / Вычислительные технологии. - Новосибирск: ИВТ СО РАН, 2001. - Т.6, Спец. вып., Ч.2. - С.549-553.
8. Разумов В.И., Сизиков В.П. Универсализация языка управления и моделирования на базе ТДИС // Конф., посвящ. 90-летию со дня рожд. А.А. Ляпунова: Сб. докл. - Новосибирск: ИВТ СО РАН, 2001. - С.528-538.
9. Разумов В.И., Сизиков В.П. Согласование структурных и функциональных особенностей моделей в аспекте управления системами // Параллельные вычисления и задачи управления: Тр. Междун. конф. РАСО'2001. - М.: ИПУ, 2001. - С.245-272.
10. Разумов В.И., Сизиков В.П. Ритмическая природа организации субъекта // Актуальные проблемы электронного приборостроения: Матер. VI Междун. конф. - Новосибирск: НГТУ, 2002. - Т.5. - С.102-106.

ПРЕДВАРИТЕЛЬНАЯ НЕЙРООБРАБОТКА ИЗОБРАЖЕНИЙ В УСЛОВИЯХ ШУМОВ

О.В.Серпенинов, Е.А.Самойлин

В работе представлены результаты предварительной обработки зашумленных изображений однослойной нейронной сетью для последующего распознавания. Нейросетевая обработка, заключающаяся в выделении контуров объектов и уменьшении уровня шумов, сопоставлена с аналогичной цифровой. Моделирование сети проведено для случаев линейных и сигмоидальных активационных функций нейронов. Показано, что обработка зашумленных изображений предлагаемой нейронной сетью является более предпочтительной по сравнению с аналогичной цифровой для ее использования в оптико-электронных системах.

The results of preliminary processing of the noised images by the single noisy a one-layer neural network for the subsequent recognition are presented. A neural network processing consisting in allocation of objects contours and reduction of a noise level, is compared with a similar digital processing. The modeling of a network is carried out for cases of linear and sigmoid activation functions of neurons. It is shown, that the processing of the noised images by the offered neural network is more preferable in comparison with a similar digital processing for its use in optical-electronic systems.

ВВЕДЕНИЕ

В последнее время большое внимание уделяется работам по созданию "интеллектуальных" оптических систем распознавания и обработки различных изображений. Это объясняется широкими возможностями систем искусственного видения, вплоть до передачи им некоторых функций, которые ранее были посильны только человеку. Как правило, подобные распознающие системы осуществляют оптико-электронное преобразование изображения и последующую его обработку на основе различных алгоритмов обучения. Причем доказано, что для эффективного распознавания во многих случаях достаточно анализировать только его контурное изображение [1]. При этом внутренние элементы контуров объектов и фона можно исключить, так как они являются малоинформативными. Кроме того, в данном случае происходит уменьшение объемов памяти, необходимых для обработки изображений [1].

В настоящее время наиболее распространенными подходами, применяемыми для выделения границ объектов, являются статистический метод [2], выделение экстремумов кривизны изолиний [3] и использование теории нейронных сетей [4]. Однако в первых двух случаях для описания изображения требуется определить энергетическую функцию по достаточно большой окрестности точек, что ведет к резкому возрастанию вычислительных затрат.

В тоже время в последний период наметился повышенный интерес к теории нейронных сетей, которая

находит все большее применение для решения ряда инженерных задач [4-6]. Это объясняется возможностью решения нейронной сетью трудно формализуемых задач, именно таких как распознавание изображений, и относительной простотой ее реализации на практике.

В работе [4] приведены результаты моделирования нейронных сетей для решения задач определения границы раздела случайных полей на изображениях в условиях шумов. Полученные результаты свидетельствуют о преимуществах данных нейронных сетей перед оптимальной обработкой.

Цель настоящей работы - обоснование применимости нейронной сети для выделения границ объектов на изображениях в условиях шумов и сравнение результатов нейросетевой и цифровой обработки.

1 ПОСТАНОВКА ЗАДАЧИ

Рассматривается задача выделения границ объектов на изображениях однослойной нейронной сетью. Входной информацией для нейронов сети служит множество элементов полутоновой матрицы изображения, заданных непосредственно в виде набора значений зачерненности точек раstra. Для определения границ объектов на выходе сети, необходим синтез такой ее архитектуры, при которой будет оцениваться градиентная близость соседних элементов матриц. То есть необходимо выбрать такую функцию, при которой ее локальные экстремумы соответствовали бы характерным фрагментам изображения. Подобная функция на участках, где изображение изменяется мало, является почти постоянной, а на участках, содержащих геометрические особенности, резко колеблется. В результате на выходе сети должны быть элементы матриц, которые относятся исключительно к контурным линиям объектов.

2 СТРУКТУРА НЕЙРОННОЙ СЕТИ

Оптимальное решение задачи сегментации полутоновой матрицы может быть достижимо с помощью следующей архитектуры сети.

Сигнал от каждого пикселя $X_{i,j}$ исходного изображения поступает на соответствующий ему нейрон $N_{i,j}$, число которых равно числу пикселей. Параллельно с сигналом от пикселя $X_{i,j}$ на нейрон $N_{i,j}$ подаются значения соседних пикселей $X_{i\pm 1, j\pm 1}$, как показано на рис.1.

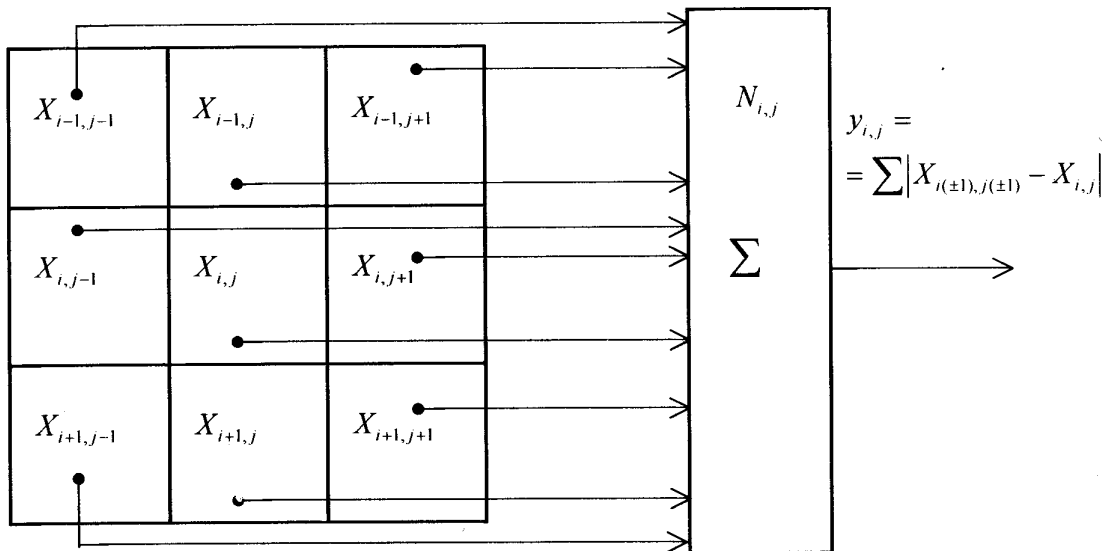


Рисунок 1 - Фрагмент полутоновой матрицы с пиксельными элементами $X_{i,j}$ и реализация его обработки нейроном

Для определения принадлежности элементов раstra к контурным линиям, каждым нейроном сети в слое проводится сравнение его яркости с яркостями смежных элементов полутоновой матрицы. То есть производится нейронное дифференцирование яркостных элементов соседних пикселей. Принадлежность элемента полутоновой матрицы к контурной линии определяется на выходе слоя сети исходя из следующего условия [7,8]:

$$y_{i,j} = \begin{cases} N \cdot M \\ i = 1, i = 1 \end{cases} |X_{i\pm 1, j\pm 1} - X_{i,j}| \geq P, \quad (1)$$

где $y_{i,j}$ - сигнал на выходе нейрона $N_{i,j}$, N - количество строк матрицы изображения, M - количество столбцов матрицы изображения, $X_{i,j}$ - вычисляемый в данный момент элемент изображения, $X_{i\pm 1, j\pm 1}$ - соседние элементы с вычисляемым, P - активационный порог, задаваемый эмпирически.

На рис.1. для упрощения показан только один нейрон $N_{i,j}$, осуществляющий суммирование дифференциалов яркостей соседних элементов $X_{i\pm 1, j\pm 1}$ с центральным элементом $X_{i,j}$. Нейроны, соседние с $N_{i,j}$ смещены друг относительно друга на один элемент матрицы. Таким образом, центральный для одного нейрона элемент матрицы $X_{i,j}$, для другого нейрона будет уже $X_{i+1,j}$ (в случае смещения "вниз", на следующую строку). При обработке пикселей первых и последних строк и столбцов $i, j = 1; j = N; j = M$ нейроны имеют неполное число входов, равных количеству окружающих элементов данного пикселя. В случае обработки четырех угловых пикселей $X_{1,1}, X_{1,M}, X_{N,1}, X_{N,M}$ у нейрона имеются всего три входа.

Сигналы с выходов нейронов $N_{i\pm 1, j\pm 1}$ подаются на блок пороговой обработки y_p (рис.2), где для каждого элемента $y_{i,j}$ вычисляется условие (1). Активационный порог P устанавливается эмпирическим путем, и может находиться в широких пределах.

В случае "срабатывания" пороговой функции передачи, сигналы подаются на блок их умножения на входной сигнал, после чего на выходе нейросети формируются сигналы границ раздела $X_{i,j}y_{i,j}^1$ элементов матриц. В остальных случаях, когда пороговая функция не "срабатывает", на выход сети поступают нулевые сигналы $Y_{i,j}^0$ элементов матриц, не являющихся граничными.

После блока умножения сигналы границ раздела $X_{i,j}y_{i,j}^1$ поступают на вход активационного блока F , где преобразуются сигмоидальной функцией передачи, которая математически выражается [5,6]:

$$F(Xy) = X_{i,j}y_{i,j}^1 + \frac{a}{e^{X_{i,j}y_{i,j}^1}}. \quad (2)$$

Общая схема нейронной сети, приведенная на рис.2, для удобства представления, также упрощена. Здесь показаны только соседние два элемента входа с центральным вместо восьми.

В отличие от известных нейронных сетей для решения подобных задач [4-6], предлагаемая обладает сравнительной простотой структуры, так как имеет всего один слой; у нее отсутствуют какие-либо алгоритмы обучения и настройки весовых коэффициентов. В данном случае сама сеть за счет своей архитектуры решает задачу с оптимальным выбором выходных векторов.

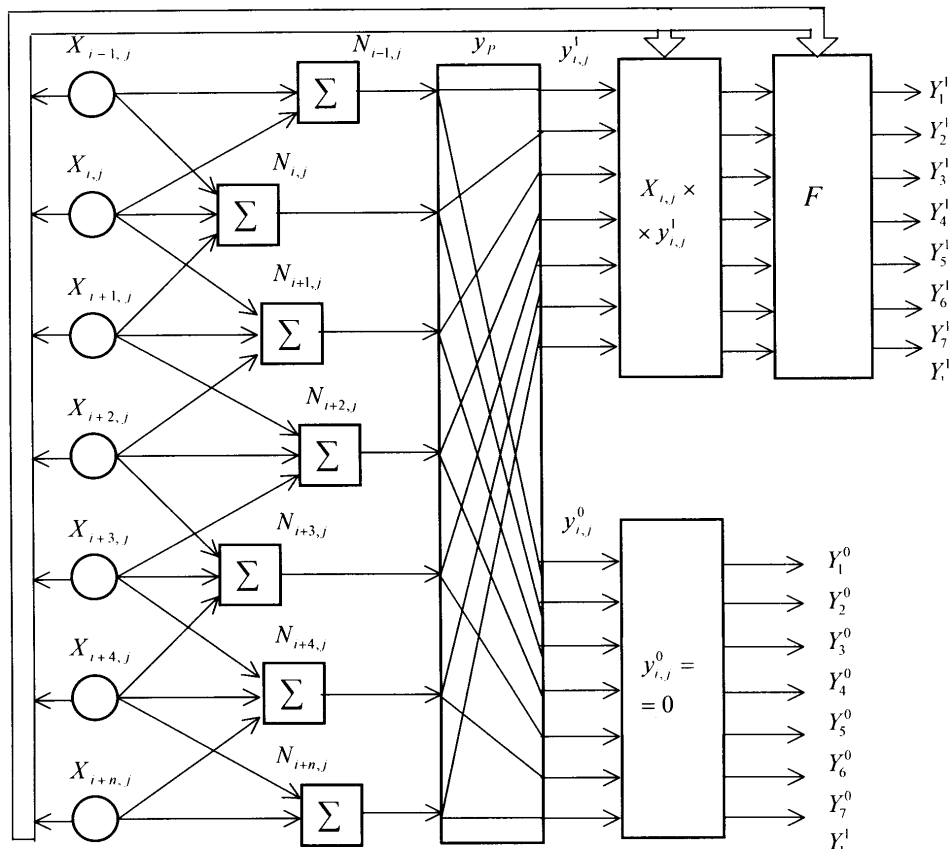


Рисунок 2 - Однослойная нейронная сеть с сигмоидальными активационными функциями

3 РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Моделирование рассмотренной выше нейронной сети проводилось в среде MathCAD версии 7.0 для операционной системы Windows 98. На вход сети подавалась полутоновая матрица исходного изображения размерностью 200x150 с яркостной градацией, равной 256. Данная матрица содержала изображение некоторых основных тестовых объектов (квадрат, треугольник, окружность, линии с различными углами наклона), с некоторым уровнем и интенсивностью зашумленности (рис.3а). На рис.3б представлен результат выделения границ объектов данной полутоновой матрицы по широко известному цифровому алгоритму [7,8].

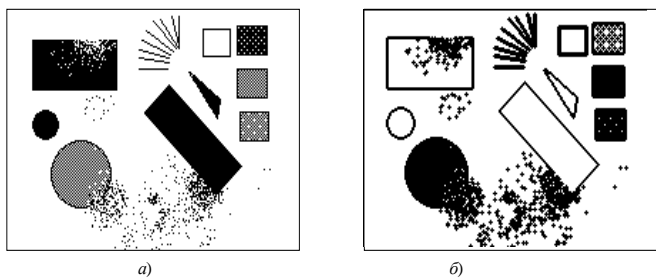


Рисунок 3 - Изображение исходной полутоновой матрицы (а), и результат ее обработки цифровым способом (б)

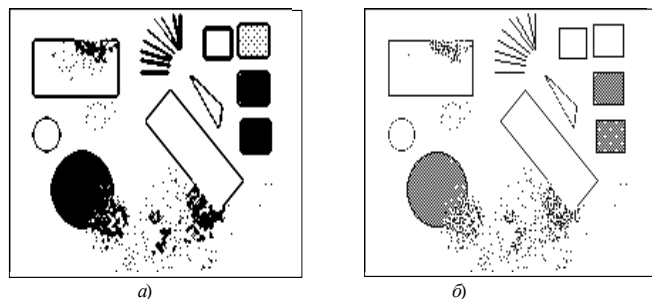


Рисунок 4 - Изображение полутоновой матрицы, полученное нейронной сетью с линейной (а), и сигмоидальной (б) функциями активации

Из рис.3а и 3б видно, что все точки фона, находящиеся внутри односвязных контуров, удалены. Оставлены лишь точки, принадлежащие границам раздела объектов и фона. Однако вместе с этим имеют место такие недостатки, как искусственное повышение уровня шумов (на 10-15%), заметное увеличение толщины замкнутых и одиночных линий (на 100%), т.е. происходит искажение исходной полутоновой матрицы.

На рис.4а приведены результаты определения границ исходной полутоновой матрицы, показанной на рис.3а, однослойной нейронной сетью, аналогичной показанной на рис.2, без активационного блока F.

Из рис.4а видно, что по сравнению с цифровым алгоритмом определения границ, нейросетевой увеличивает уровень зашумленности всего на 3-5%.

При этом толщина контурных линий также увеличивается (до 100%).

На рис.4б представлены результаты обработки исходной полутоновой матрицы нейронной сетью с сигмоидальной активационной функцией передачи, показанной на рис.2.

Сравнивая исходную полутоновую матрицу (рис.3а) с полученной нелинейной нейронной сетью (рис.4б), можно сделать вывод о том, что в данном случае имеет место уменьшение общего уровня зашумленности изображения (на 5-10%), искажений в толщине контурных линий не происходит. Это объясняется нелинейным характером нейронных активационных функций.

Таким образом, решение задачи определения границ объектов на изображениях с использованием нелинейной нейронной сети является наиболее предпочтительным по сравнению с линейной сетью и цифровой обработкой.

ЗАКЛЮЧЕНИЕ

Представленные в работе результаты моделирования алгоритма выделения границ объектов на изображениях в присутствии шумов свидетельствуют об обоснованности применения нейросетевого подхода к подобным задачам и принципиальной возможности замены традиционных алгоритмов обработки существенно более качественными -

нейросетевыми. Нейронная сеть для оптимального решения задачи выделения границ является достаточно простой, имеет один слой нейронов, у нее отсутствуют алгоритмы обучения и настройки весовых коэффициентов. Однако даже такая нейронная сеть обеспечивает намного более качественное решение задачи выделения границ объектов с уменьшением уровня шумов.

ПЕРЕЧЕНЬ ССЫЛОК

1. Кривецкий А.В. Распознавание трехмерных объектов по форме пространственных контуров // Автометрия. -2001. №2. -С.21-30.
2. Злобин В.К., Еремеев В.В., Васильев В.М. Стохастическая модель спутниковых изображений и ее использование для сегментации природных объектов // Автометрия. -2001. №2. -С.41-50.
3. Бобков В.А., Казанский А.В., Морозов М.А. Выделение размытых контуров на примере определения скорости течений по спутниковым изображениям // Автометрия. - 2001. №2.- С.3-12.
4. Попело В.Д., Сирота А.А., Маслов О.В. Сравнительный анализ оптимального и нейросетевых алгоритмов определения границ раздела случайных полей при обработке изображений // Радиотехника (журнал в журнале). 2001. №10. - С. 81-85.
5. Уоссермен Ф. Нейрокомпьютерная техника: Теория и практика /Пер. с англ. Зуев Ю.А., Точенов В.А. - М.: Мир, - 1992. - 178 с.
6. Горбань А.Н., Дунин-Барковский В.А., Миркес Е.М. и др. Нейроинформатика. - Новосибирск: Наука, 1998. - 224 с.
7. Бакут П.А., Колмогоров Г.С., Ворновицкий И.Э. Сегментация изображений: методы пороговой обработки // Зарубежная радиоэлектроника. 1987. №10. -С.6-24.
8. Бакут П.А., Колмогоров Г.С. Сегментация изображений: методы выделения границ областей // Зарубежная радиоэлектроника. -1987. № 10. -С.25-47.

УДК 007.52

АЛГОРИТМИ КЛАСТЕР-РЕГРЕСІЙНОЇ АПРОКСИМАЦІЇ ТА ЇХ НЕЙРОМЕРЕЖЕВІ ІНТЕРПРЕТАЦІЇ

С.О.Субботін

Запропоновано алгоритми, які дозволяють будувати моделі кількісних залежностей за точковими даними, що враховують топологічні та статистичні властивості даних. Розроблено нейромережеві інтерпретації запропонованих алгоритмів.

Предложены алгоритмы, позволяющие строить модели количественных зависимостей по точечным данным, учитывающие топологические и статистические свойства данных. Разработаны нейросетевые интерпретации предложенных алгоритмов.

The algorithms for construction of models of quantitative relations on a dot data are proposed. They includes topology and statistic properties of a data. The neural network interpretations of proposed algorithms are developed.

ВСТУП

Нехай задана навчальна вибірка екземплярів $x = \{x_j^s\}$, де x_j^s - значення j -ої ознаки s -го

екземпляру, $s = 1, 2, \dots, S$; S - кількість екземплярів у навчальній вибірці. Кожному екземпляру навчальної вибірки співставлене значення прогнозованого параметру y^s . Тоді залежність y від x можна записати у вигляді: $y = f(w, x) + \text{Err}$, де $f(w, x)$ - деяка функція, вид якої задається алгоритмом апроксимації або користувачем, w - множина параметрів, що дозволяють налагодити (навчити) функцію на вирішення певної апроксимаційної задачі, Err - деяка помилка, що виникає завдяки неповній відповідності виду апроксимуючої функції виду реально існуючої залежності та похибкам у визначенні значень параметрів апроксимуючої функції.

Для знаходження параметрів апроксимуючої функції можна використовувати засоби регресійного аналізу [1]. При використанні одновимірних лінійних регресійних моделей для апроксимації багатовимірних нелінійних залежностей, як правило, не вдається забезпечити бажану

точність прогнозування, але параметри апроксимуючої функції розраховуються за один прохід у неітеративному режимі. З іншого боку, у певних областях у просторі ознак значення прогнозованого параметру може визначатися в основному якоюсь однією ознакою, що є найбільш сильно корельованою з прогнозованим параметром у цій області.

Тому, узагальнюючи сказане вище, можна заключити, що кусочно-лінійна регресія, яка представляє собою систему лінійних одновимірних регресійних моделей, побудованих для кожної області окремо за найбільш інформативною для даної області ознакою, може забезпечити кращу апроксимацію та більшу точність, ніж лінійна регресійна модель для якоїсь однієї ознаки на усій навчальній вибірці. Побудова кусочно-регресійної моделі вимагає виділення областей компактно згрупованих екземплярів у просторі ознак з близькими значеннями прогнозованого параметру - кластерів.

1 АЛГОРИТМ КЛАСТЕР-РЕГРЕСІЙНОЇ АПРОКСИМАЦІЇ

Алгоритм побудови кластер-регресійної апроксимації запишемо у наступному виді.

Крок 1. Ініціалізація: задати навчальну вибірку екземплярів x та набір співставлених їм значень прогнозованого параметру y .

Крок 2. На основі екземплярів навчальної вибірки сформувані кластери C^q , $q = 1, 2, \dots, Q$, де Q - кількість сформованих кластерів. Для виділення кластерів пропонується використовувати описаний нижче алгоритм формування кластерів.

Крок 3. Розділити екземпляри навчальної вибірки на групи, що відповідають кожному кластеру за мінімумом відстані від екземпляру до центру кластера у просторі ознак.

Крок 4. Для екземплярів кожної групи знайти значення коефіцієнтів кореляції їхніх ознак та значень прогнозованого параметру:

$$r^q(x_j, y) = \frac{\int_{s=1}^S \left\{ x_j^s - \frac{1}{S^q} \int_{s=1}^S x_j^s \right\} \left\{ y^s - \frac{1}{S^q} \int_{s=1}^S y^s \right\}}{\sqrt{\int_{s=1}^S \left\{ x_j^s - \frac{1}{S^q} \int_{s=1}^S x_j^s \right\}^2 \int_{s=1}^S \left\{ y^s - \frac{1}{S^q} \int_{s=1}^S y^s \right\}^2}}, x^s \in C^q,$$

$$j = 1, 2, \dots, N,$$

де N - кількість ознак, S^q - кількість екземплярів, що відносяться до q -го кластеру.

Крок 5. Для кожної групи екземплярів знайти номер ознаки, модуль коефіцієнту кореляції котрого з прогнозованим параметром є найбільшим:

$$g^q = \arg \max_j |r^q(x_j, y)|, q = 1, 2, \dots, Q; j = 1, 2, \dots, N.$$

Крок 6. Для кожної групи екземплярів оцінити значення параметрів рівняння лінійної регресії для прогнозованого параметру за ознакою g^q .

Якщо рівняння лінійної регресії прогнозованого параметру за ознакою x_{g^q} для екземплярів q -го кластеру має вид: $y_q = b_{q0} x_{g^q} + b_{q1}$, тоді значення коефіцієнтів b_{q0} та b_{q1} можуть бути знайдені за допомогою формул [2]:

$$b_{q0} = \frac{\left\{ a_1^q - a_2^q a_3^q \right\}}{\left\{ a_4^q - a_3^q a_3^q \right\}},$$

$$b_{q1} = \frac{\left\{ a_4^q a_2^q - a_3^q a_1^q \right\}}{\left\{ a_4^q - a_3^q a_3^q \right\}},$$

$$a_1^q = \frac{1}{S^q} \int_{s=1}^S y^s x_{g^q}^s, x^s \in C^q,$$

$$a_2^q = \frac{1}{S^q} \int_{s=1}^S y^s, x^s \in C^q,$$

$$a_3^q = \frac{1}{S^q} \int_{s=1}^S x_{g^q}^s, x^s \in C^q,$$

$$a_4^q = \frac{1}{S^q} \int_{s=1}^S (x_{g^q}^s)^2, x^s \in C^q.$$

Алгоритм розрахунку значення прогнозованого параметру на основі кластер-регресійної апроксимації запишемо у наступному вигляді.

Крок 1. Ініціалізація: Задати вибірку екземплярів, для котрих здійснюється прогнозування $x = \{x_j^s\}$. Задати кількість кластерів Q та координати їхніх центрів $C = \{C_j^q\}$. Для кожного кластеру задати номер ознаки g^q та значення параметрів рівняння регресії b_{q0} та b_{q1} .

Крок 2. Для кожного екземпляру визначити відстані від екземпляру до центрів усіх кластерів

$$R(x^s, C^p), s = 1, \dots, S; p = 1, \dots, Q.$$

Крок 3. Для кожного екземпляру знайти:

$$q = \arg \min_p R(x^s, C^p), s = 1, \dots, S; p = 1, \dots, Q.$$

Крок 4. Значення прогнозованого параметру для кожного s -го екземпляру визначити за формулою:

$$y^s = y_q^s = b_{q0} x_{g^q}^s + b_{q1}, s = 1, \dots, S.$$

2 АЛГОРИТМ ФОРМУВАННЯ КЛАСТЕРІВ

Для формування кластерів (компактних областей групування екземплярів у просторі ознак) пропонується використовувати наступний алгоритм [3].

Крок 1. Ініціалізація параметрів алгоритму побудови навчальної вибірки. Задати вихідну вибірку екземплярів $x_{\text{вих}}$ та співставлені їм номери класів або значення прогнозованого параметру $y_{\text{вих}}$, а також L - кількість разбиттів вихідної вибірки. Занести у змінну N кількість ознак, що характеризують екземпляри, а у змінну S - кількість екземплярів вихідної вибірки. Прийняти ширину допустимого інтервалу варіації прогнозованого параметру:

$$dy = |\max(y_{\text{вих}}) - \min(y_{\text{вих}})| / L,$$

де $\min(a)$ та $\max(a)$ - мінімальне та максимальне значення вектору a , відповідно. Установити лічильник $\text{newind} = 1$.

Крок 2. Обчислити відстані між екземплярами вихідної вибірки:

$$R(p, q) = \begin{cases} \sqrt{\sum_{j=1}^N (x_j^p - x_j^q)^2}, p \neq q; \\ \text{RealMax}, p = q, \end{cases}$$

$$s = 1, \dots, S; p = 1, \dots, Q,$$

де RealMax - максимальне число, що може бути представлене у ЕОМ, x_j^p - значення j -ої ознаки p -го екземпляру.

Крок 3. Знайти у матриці відстаней R мінімальний елемент minx та його індекси q та p , а також максимальний елемент maxx , за умови, що при знаходженні мінімуму й максимуму тут і далі ігноруються елементи, які дорівнюють RealMax .

Крок 4. Прийняти $a = |\text{maxx} - \text{minx}| / (2L)$.

Крок 5. Якщо $\text{minx} < \text{RealMax}$, тоді перейти на крок 6, у противному випадку - перейти на крок 13.

Крок 6. Прийняти:

$$C^{(\text{newind})} = x_{\text{вих}}(q), y^*(\text{newind}) = y_{\text{вих}}(q),$$

де C та y^* - масиви еталонних екземплярів кластерів та співставлених ним прогнозованих значень, відповідно. Установити: $\text{newind} = \text{newind} + 1$, значення текущего мінімального елемента у рядку $\text{uteck} = R(q, p)$. Зайти мінімальний елемент mminx та його індекси mqmin та mpmin серед елементів q -го рядку матриці R .

Крок 7. Якщо $\text{mminx} < \text{Realmax}$, тоді перейти на крок 8, у противному випадку - перейти на крок 11.

Крок 8. Установити значення покажчика видаленого екземпляру із стовпця $\text{deleted} = 0$ (у матриці R нумерація рядків та стовпців повинна начинатися з 1).

Крок 9. Якщо $|\text{uteck} = R(q, \text{mpmin})| \leq a$, тоді перейти на крок 10, інакше прийняти: $\text{deleted} = \text{mpmin}$, $R(q, \text{mpmin}) = \text{Realmax}$, $R(\text{mpmin}, q) = \text{Realmax}$, та перейти на крок 11.

Крок 10. Якщо $|y_{\text{вих}}(\text{mpmin}) - y_{\text{вих}}(q)| \leq dy$ тоді прийняти:

$R(v, \text{mpmin}) = \text{realmax}$, $R(\text{mpmin}, v) = \text{realmax}$
 $v = 1, \dots, M$, у противному випадку - прийняти:

$$C^{(\text{newind})} = x_{\text{вих}}(\text{mpmin}),$$

$$y^*(\text{newind}) = y_{\text{вих}}(\text{mpmin}),$$

$$\text{newind} = \text{newind} + 1,$$

$$R(v, \text{mpmin}) = \text{Realmax},$$

$$R(\text{mpmin}, v) = \text{Realmax},$$

$$v = 1, \dots, S.$$

Крок 11. Знайти мінімальний елемент mminx та його індекси mqmin та mpmin серед елементів q -го рядку матриці R .

Крок 12. Перейти на крок 7.

Крок 13. Прийняти: $R(v, q) = \text{Realmax}$, $R(q, v) = \text{Realmax}$, $v = 1, \dots, S$, покажчик видаленого екземпляру із рядку $\text{dstr} = q$. Знайти у матриці відстаней R мінімальний елемент minx та його індекси q та p .

Крок 14. Перейти на крок 5.

Крок 15. Якщо $(\text{deleted} \neq \text{dstr})$ та $(\text{deleted} > 0)$, тоді прийняти:

$$C^{(\text{newind})} = x_{\text{вих}}(\text{deleted}), y^*(\text{deleted}) = y_{\text{вих}}(\text{deleted}).$$

Крок 16. Зупинення.

У результаті виконання цього алгоритму для вихідної вибірки $x_{\text{вих}}$ та співставленого їй набору значень $y_{\text{вих}}$ ми отримаємо множину еталонних екземплярів кожного кластеру C та співставлений ним набір значень y^* . Після цього необхідно здійснити розділення на кластери для всіх екземплярів вихідної вибірки за M мінімумом близькості до еталонних екземплярів кластеру, а потім для кожного кластеру прийняти координати центру кластеру рівними:

$$C_j^q = \frac{1}{S^q} \int_{s=1}^S x_j^s, x^s \in C^q, q = 1, \dots, Q; j = 1, \dots, N;$$

де S^q - кількість екземплярів вихідної вибірки, що належать до q -го кластеру, Q - кількість сформованих кластерів.

Після формування кластерів алгоритм кластер-регресійної апроксимації можна використовувати для синтезу та налагоджування вагових коефіцієнтів п'ятишарової НМ.

3 НЕЙРОМЕРЕЖЕВА ІНТЕРПРЕТАЦІЯ АЛГОРИТМУ ОДНОВИМІРНОЇ ЛІНІЙНОЇ КЛАСТЕР-РЕГРЕСІЙНОЇ АПРОКСИМАЦІЇ

Якщо сформовано Q кластерів, тоді дискримінантні функції нейронів мережі будуть задаватися формулами:

$$\begin{aligned} \varphi^{(1,i)}(x,w) &= \int_{j=1}^N w_j x_j + w_0, i = 1, 2, \dots, Q; \\ \varphi^{(1,i)}(x,w) &= \int_{j=1}^N (x_j - w_j)^2, i = Q+1, Q+2, \dots, 2Q; \\ \varphi^{(2,i)}(x,w) &= \int_{j=1}^2 w_j x_j + w_0, i = 1, 2, \dots, Q^2; \\ \varphi^{(3,i)}(x,w) &= \int_{j=1}^Q w_j x_j + w_0, i = 1, 2, \dots, Q; \\ \varphi^{(4,i)}(x,w) &= \prod_{j=1}^2 w_j x_j, i = 1, 2, \dots, Q; \end{aligned}$$

$$\varphi^{(5,i)}(x,w) = \int_{j=1}^2 w_j x_j + w_0, i = 1;$$

де $\varphi^{(\mu,i)}$ - дискримінантна функція i -го нейрону μ -го шару.

Функції активації нейронів будуть визначатися виразами:

$$\psi^{(\mu)}(\alpha) = \begin{cases} 0, & \alpha \leq 0, \\ 1, & \alpha > 0; \end{cases} \mu = 2, 3;$$

$$\psi^{(\mu)}(\alpha) = \alpha, \mu = 1, 4, 5.$$

Схема НМ, синтезованої та навченої на основі запропонованого алгоритму, зображена на рис. 1. Ваговий коефіцієнт j -го входу i -го нейрону μ -го шару мережі буде розраховуватися за формулою:

$$w_j^{(\mu,i)} = \begin{cases} 0, & \mu = 5, i = 1, j = 0; \\ 1, & \mu = 5, i = 1, j > 0; \\ 1, & \mu = 4, \forall i, j = 0; \\ 1 - Q, & \mu = 3, i = 1, j = 0; \\ 1, & \mu = 3, i = 1, j > 0; \\ 0, & \mu = 2, \forall i, j = 0; \\ 1, & \mu = 2, \forall i, j = 2; \\ -1, & \mu = 2, \forall i, j = 1; \\ b_{q0}, & \mu = 1, i = q, j = g^q, q = 1, \dots, Q; \\ b_{q1}, & \mu = 1, i = q, j = 0, q = 1, \dots, Q; \\ 0, & \mu = 1, i = q, 0 > j \neq g^q, q = 1, \dots, Q; \\ C_j^q, & \mu = 1, i = Q+1, Q+2, \dots, 2Q, j = 1, \dots, N. \end{cases}$$

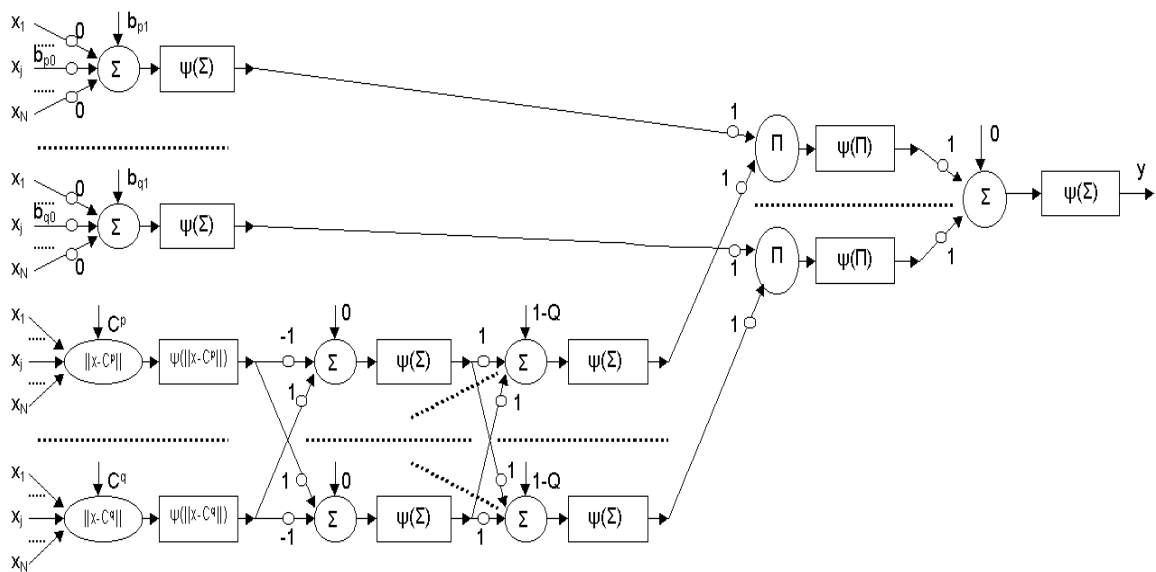


Рисунок 1 - Схема НМ, синтезованої на основі алгоритму кластер-регресійної апроксимації (випадок лінійної одновимірної регресії)

Розглянутий алгоритм дозволяє досягнути суттєво більшої точності, ніж метод лінійної одновимірної регресії, проте він має таку ж простоту та ефективність обчислювальних процедур.

НМ, що отримується у результаті синтезу й навчання на основі розглянутого алгоритму, є логічно прозорою та легко перетворюється у дерево рішачих правил "Якщо - то", що дозволяє використовувати її для видобування знань з даних. Точність класифікації та швидкість навчання НМ, сформованої на основі запропонованого алгоритму, є достатніми для багатьох прикладних задач діагностики та прогнозування.

4 АЛГОРИТМ БАГАТОВИМІРНОЇ ЛІНІЙНОЇ КЛАСТЕР-РЕГРЕСІЙНОЇ АПРОКСИМАЦІЇ

У випадку, коли запропонований алгоритм одновимірної лінійної кластер-регресійної апроксимації не дозволяє отримати модель залежності прогнозованого параметру від набору ознак із заданою точністю, тоді кроки 5 та 6 алгоритму побудови кластер-регресійної апроксимації можна замінити на наступні.

Крок 5. Для кожного q -го кластеру усі ознаки екземплярів, для яких $|r(x_j, y)| < P$, де P - заданий порог, $0 < P < 1$, вилучити із набору ознак екземплярів, що належать до q -го кластеру.

Крок 6. Для кожного q -го кластеру оцінити параметри багатовимірної лінійної регресії прогнозованого параметру:

$$y^s = \int_{j=1}^{m_q} b_{qj} x_j^s, x_j^s \in C^q,$$

$$b = (x^T x)^{-1} x^T y,$$

де m_q - кількість ознак екземплярів q -го кластеру, для яких $|r(x_j, y)| < P$.

Розглянемо багатовимірний випадок лінійної регресійної моделі [4].

Нехай $x = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \dots & \vdots \\ x_{p1} & x_{p2} & \dots & x_{pm} \end{bmatrix}$ - матриця p значень, що

спостерігаються, m змінних x_1, x_2, \dots, x_m ,

$y = (y_1, y_2, \dots, y_p)^T$ - вектор значень пояснювальної

змінної, що спостерігаються, де T - символ транспонування, тоді загальна лінійна регресійна модель може бути представлена в стандартному вигляді так:

$$y = \int_{j=1}^m b_j x_{ij} + u, i = 1, 2, \dots, \text{ або } y = xb + u,$$

де $b = (b_1, b_2, \dots, b_m)^T$ та $u = (u_1, u_2, \dots, u_p)^T$ - вектор

констант та вектор членів помилки (залишків), відповідно.

Оцінкою найменших квадратів є такий вектор параметрів b , що мінімізує суму квадратів регресійних залишків E :

$$E = u^T u = (y - xb)^T (y - xb).$$

Виходячи з умови мінімізації суми квадратів регресійних залишків:

$$\frac{\partial E}{\partial b_i} = 0,$$

де $i = 1, 2, \dots, m$, одержуємо систему нормальних рівнянь:

$$\frac{\partial E}{\partial b_i} = -2(x^T y - x^T x b) = 0,$$

звідки $b = (x^T x)^{-1} x^T y$, де $(x^T x)^{-1}$ - матриця, зворотна матриці $x^T x$, що за припущенням існує.

Для знаходження зворотної матриці рекомендується використовувати узагальнений метод виключення з вибором ведучого елемента матриці, заснований на використанні ітеративної формули Бен-Ізраеля [4].

Узагальненою зворотною матрицею A_+ для довільної матриці A розміру $(m \times n)$ називається матриця, що задовольняє наступним умовам:

$$AA_+A = A,$$

$$A_+AA_+ = A_+,$$

$$AA_+ = (AA_+)^T,$$

$$A_+A = (A_+A)^T.$$

Теорема Пенроуза говорить, що для будь-якої матриці існує матриця, що задовольняє всім цим умовам і притому єдина.

Нехай λ^* - максимальне характеристичне значення матриці AA^T . Тоді покладемо $\alpha = \frac{2}{|\lambda^*|} 0,8$. Якщо

позначити через $A_+^{(t)}$ t -е наближення, найближче до узагальненої зворотної матриці A_+ для матриці A , то ми зможемо одержати A_+ з рекурентного співвідношення:

$$A_+^{(0)} = \alpha A; A_+^{(t+1)} = A_+^{(t)}(2I - AA_+^{(t)}). \quad (9)$$

Таким чином, $\lim_{t \rightarrow \infty} A_+^{(t)} = A_+$.

Максимальне характеристичне значення λ^* можна знайти за допомогою наступного алгоритму.

Крок 1. Ініціалізація. Установити: $k = 1$, $z_i = 1$, $i = 1, 2, \dots, m$. Задати значення критерію збіжності ξ .

Крок 2. Обчислити $y_i = \int \int_{j=1}^m \int_{d=1}^n A_{ij} A_{id}$, $i = 1, 2, \dots, m$.

Крок 3. Якщо $y_k = 0$, то перейти на крок 4, інакше перейти на крок 6.

Крок 4. Якщо $k < m$, то $k = k + 1$, інакше $k = 1$.

Крок 5. Перейти на крок 3.

Крок 6. Обчислити $z_i = \frac{y_i}{y_k}$, $i = 1, 2, \dots, m$, та установити $\lambda^* = y_k$.

Крок 7. Перевірка збіжності: Якщо $\left| \frac{\lambda^*}{y_k} - 1 \right| > \xi$, то перейти на крок 2, інакше - зупинення.

На основі розрахованих b_{qj} визначити коефіцієнти для усього вихідного набору ознак екземплярів q -го кластеру:

$$\beta_{qj} = \begin{cases} 0, & \text{якщо } |r(x_j, y)| < P; \\ b_{qk}, & \text{у протилежному випадку;} \end{cases}$$

де k - номер ознаки з набору ознак, що враховувалися при визначенні параметрів багатовимірної регресії для q -го кластеру, яка відповідає j -й ознаці із вихідного набору ознак.

Алгоритм розрахунку значення прогнозованого параметру на основі кластер-регресійної апроксимації для випадку багатовимірної лінійної регресії запишемо у наступному вигляді.

Крок 1. Ініціалізація: Задати вибірку екземплярів для котрих здійснюється прогнозування $x = \{x_j^s\}$. Задати кількість кластерів Q та координати їхніх центрів $C = \{C_j^q\}$. Для кожного кластеру задати значення параметрів багатовимірної лінійної регресії β_{qj} .

Крок 2. Для кожного екземпляру визначити відстані від екземпляру до центрів усіх кластерів $R(x^s, C^p)$, $s = 1, \dots, S; p = 1, \dots, Q$.

Крок 3. Для кожного екземпляру знайти:

$$q = \arg \min_p R(x^s, C^p), \quad s = 1, \dots, S; \quad p = 1, \dots, Q.$$

Крок 4. Значення прогнозованого параметру для кожного s -го екземпляру визначити за формулою:

$$y^s = \int_{j=1}^N \beta_{qj} x_j^s, \quad x_j^s \in C^q, \quad s = 1, \dots, S.$$

Розглянутий алгоритм кластер-регресійної апроксимації для випадку багатовимірної лінійної регресії можна використовувати для синтезу та налагоджування вагових коефіцієнтів п'ятишарової НМ.

Якщо сформовано Q кластерів, тоді дискримінантні функції нейронів мережі будуть задаватися формулами:

$$\varphi^{(1,i)}(x, w) = \int_{j=1}^N w_j x_j + w_0, \quad i = 1, 2, \dots, Q;$$

$$\varphi^{(1,i)}(x, w) = \int_{j=1}^N (x_j - w_j)^2, \quad i = Q+1, Q+2, \dots, 2Q;$$

$$\varphi^{(2,i)}(x, w) = \int_{j=1}^2 w_j x_j + w_0, \quad i = 1, 2, \dots, Q^2;$$

$$\varphi^{(3,i)}(x, w) = \int_{j=1}^Q w_j x_j + w_0, \quad i = 1, 2, \dots, Q;$$

$$\varphi^{(4,i)}(x, w) = \prod_{j=1}^2 w_j x_j, \quad i = 1, 2, \dots, Q;$$

$$\varphi^{(5,i)}(x, w) = \int_{j=1}^2 w_j x_j + w_0, \quad i = 1;$$

де $\varphi^{(\mu,i)}$ - дискримінантна функція i -го нейрону μ -го шару.

Функції активації нейронів будуть визначатися виразами:

$$\psi^{(\mu)}(\alpha) = \begin{cases} 0, & \alpha \leq 0, \\ 1, & \alpha > 0; \end{cases} \quad \mu = 2, 3;$$

$$\psi^{(\mu)}(\alpha) = \alpha, \quad \mu = 1, 4, 5.$$

Ваговий коефіцієнт j -го входу i -го нейрону μ -го шару мережі буде розраховуватися за формулою:

$$w_j^{(\mu,i)} = \begin{cases} 0, & \mu = 5, i = 1, j = 0; \\ 1, & \mu = 5, i = 1, j > 0; \\ 1, & \mu = 4, \forall i, j = 0; \\ 1 - Q, & \mu = 3, i = 1, j = 0; \\ 1, & \mu = 3, i = 1, j = 0; \\ 0, & \mu = 2, \forall i, j = 0; \\ 1, & \mu = 2, \forall i, j = 2; \\ -1, & \mu = 2, \forall i, j = 1; \\ \beta_{ij}, & \mu = 1, i = 1, \dots, Q, j = 1, \dots, N; \\ 0, & \mu = 1, i = q, 0 > j \neq g^q, q = 1, \dots, Q; \\ C_j^q, & \mu = 1, i = Q+1, Q+2, \dots, 2Q, j = 1, \dots, N. \end{cases}$$

Схема НМ, синтезованої та навченої на основі запропонованого алгоритму, зображена на рис. 2.

Запропонований алгоритм дозволяє досягнути більшої точності, ніж методи лінійної регресії, проте він має достатньо високу ефективність та простоту обчислювальних процедур.

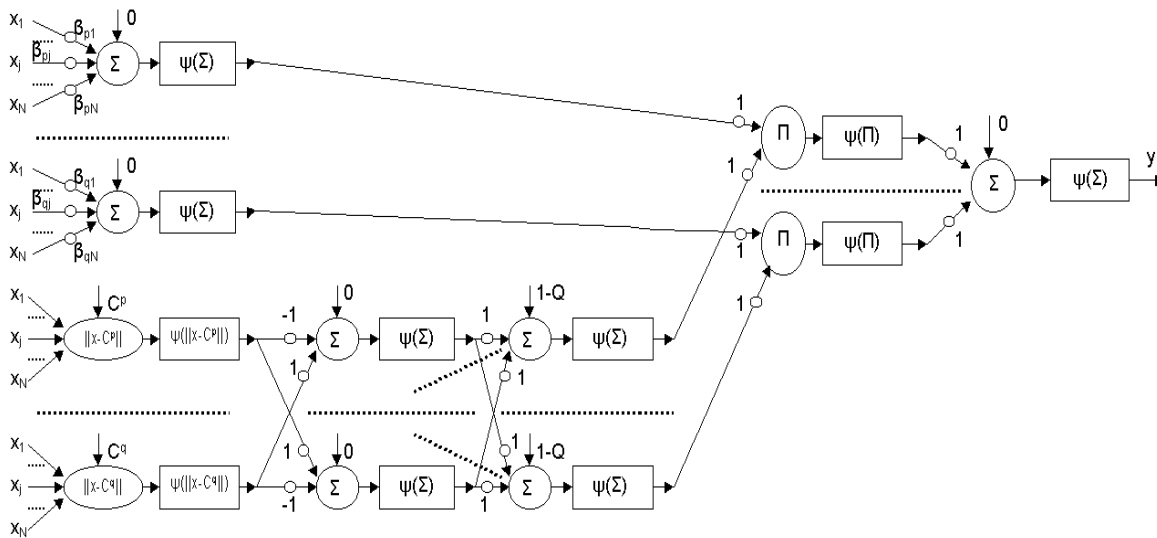


Рисунок 2 - Схема НМ, синтезованої на основі алгоритму кластер-регресійної апроксимації (випадок лінійної багатовимірної регресії)

5 АЛГОРИТМ БАГАТОВИМІРНОЇ НЕЛІНІЙНОЇ КЛАСТЕР-РЕГРЕСІЙНОЇ АПРОКСИМАЦІЇ

У випадку, коли алгоритм кластер-регресійної апроксимації для випадку багатовимірної лінійної регресії не дозволяє отримати модель залежності прогнозованого параметру від набору ознак із заданою точністю, тоді кроки 5 та 6 алгоритму побудови кластер-регресійної апроксимації можна замінити на наступний крок.

Крок 5. Для кожного q -го кластеру оцінити параметри багатовимірної нелінійної регресії прогнозованого параметру:

$$y^s = \frac{1}{1 + \exp\left\{-\int_{j=1}^N \beta_{qj} x_j^s\right\}}, x_j^s \in C^q,$$

де β_{qj} - ваговий коефіцієнт j -ої ознаки для багатовимірної нелінійної регресії прогнозованого параметру у q -му кластері.

У тих випадках, коли рівняння регресії є нелінійним щодо оцінюваних параметрів, використовується нелінійний метод найменших квадратів [4].

Нехай y - змінна, що пояснюється; y_1, y_2, \dots, y_m - набір її спостережень; x_1, x_2, \dots, x_k - змінні, що пояснюють, i -те спостереження за якими являє собою вектор: $(x_{i1}, x_{i2}, \dots, x_{ik})^T$.

Необхідно змінну, що пояснюється, у виразити через x_1, x_2, \dots, x_k за допомогою функції f , вигляд якої відомий, однак невідомі деякі її параметри w_1, w_2, \dots, w_n :

$$y_i = f_i(w_1, \dots, w_n; x_{i1}, \dots, x_{ik}) + F_i, i = 1, 2, \dots, m,$$

де F_i - відхилення. Попередній вираз можна записати більш компактно, як $y_i = f_i(w_1, \dots, w_n) + F_i, i = 1, 2, \dots, m$, залишивши лише ті параметри, що будемо шукати за методом найменших квадратів, тобто мінімізуючи

$$E = \int_{i=1}^m F_i^2.$$

Введемо вектори:

$$w = (w_1, w_2, \dots, w_n)^T, f = (f_1, f_2, \dots, f_m)^T,$$

$$F = (F_1, F_2, \dots, F_m)^T, y = (y_1, y_2, \dots, y_m)^T.$$

Тепер сформулюємо задачу нелінійного регресійного аналізу: знайти w^* , таке, що при $F = y - f$ цільова функція (сума квадратів залишків) $E = F^T F$ мінімізується.

Наближене значення w_t , що одержується на t -ому кроці ітеративного процесу і наступне наближене значення w_{t+1} зв'язані між собою вектором поправки Δw :

$w_{t+1} = w_t + \Delta w$. Формула вектора поправки Δw відповідно до умови мінімізації виводиться з рішення системи лінійних рівнянь:

$$(A^T A) \Delta w = (-A^T F),$$

звідки

$$\Delta w = -(A^T A)^{-1} A^T F.$$

Тут A - Якобіан F (матриця перших часткових похідних):

$$A = \begin{bmatrix} \frac{\partial F_1}{\partial w_1} & \dots & \frac{\partial F_1}{\partial w_n} \\ \dots & \frac{\partial F_i}{\partial w_j} & \dots \\ \frac{\partial F_m}{\partial w_1} & \dots & \frac{\partial F_m}{\partial w_n} \end{bmatrix}_{w=w_i}$$

Мінімізація цільової функції E здійснюється за допомогою градієнтних методів оптимізації.

Алгоритм розрахунку значення прогнозованого параметру на основі кластер-регресійної апроксимації для випадку багатовимірної нелінійної регресії запишемо у наступному виді.

Крок 1. Ініціалізація: Задати вибірку екземплярів, для котрих здійснюється прогнозування $x = \{x_j^s\}$. Задати кількість кластерів Q та координати їхніх центрів $C = \{C_j^q\}$. Для кожного кластеру задати значення параметрів багатовимірної нелінійної регресії $\{\beta_{qj}\}$.

Крок 2. Для кожного екземпляру визначити відстані від екземпляру до центрів усіх кластерів $R(x^s, C^p)$, $s = 1, \dots, S; p = 1, \dots, Q$.

Крок 3. Для кожного екземпляру знайти:

$$q = \arg \min_p R(x^s, C^p), s = 1, \dots, S; p = 1, \dots, Q.$$

Крок 4. Значення прогнозованого параметру для кожного s -го екземпляру визначити за формулою:

$$y^s = \frac{1}{1 + \exp \left\{ - \int_{j=1}^N \beta_{qj} x_j^s \right\}}, x_j^s \in C^q, s = 1, \dots, S.$$

Розглянутий алгоритм кластер-регресійної апроксимації для випадку багатовимірної лінійної регресії можна використовувати для синтезу та налагоджування вагових коефіцієнтів п'ятишарової НМ.

Якщо сформовано Q кластерів, тоді дискримінантні функції нейронів мережі будуть задаватися формулами:

$$\begin{aligned} \varphi^{(1,i)}(x,w) &= \int_{j=1}^N w_j x_j + w_0, i = 1, 2, \dots, Q; \\ \varphi^{(1,i)}(x,w) &= \int_{j=1}^N (x_j - w_j)^2, i = Q+1, Q+2, \dots, 2Q; \\ \varphi^{(2,i)}(x,w) &= \int_{j=1}^2 w_j x_j + w_0, i = 1, 2, \dots, Q^2; \\ \varphi^{(3,i)}(x,w) &= \int_{j=1}^Q w_j x_j + w_0, i = 1, 2, \dots, Q; \end{aligned}$$

$$\varphi^{(4,i)}(x,w) = \prod_{j=1}^2 w_j x_j, i = 1, 2, \dots, Q;$$

$$\varphi^{(5,i)}(x,w) = \int_{j=1}^2 w_j x_j + w_0, i = 1;$$

де $\varphi^{(\mu,i)}$ - дискримінантна функція i -го нейрону μ -го шару.

Функції активації нейронів будуть визначатися виразами:

$$\psi^{(\mu)}(\alpha) = \frac{1}{1 + e^{-\alpha}}, \mu = 1; i = 1, \dots, Q;$$

$$\psi^{(\mu)}(\alpha) = \alpha, \mu = 1; i = Q+1, \dots, 2Q;$$

$$\psi^{(\mu)}(\alpha) = \begin{cases} 0, & \alpha \leq 0, \\ 1, & \alpha > 0; \end{cases} \mu = 2, 3;$$

$$\psi^{(\mu)}(\alpha) = \alpha, \mu = 4, 5.$$

Ваговий коефіцієнт j -го входу i -го нейрону μ -го шару мережі буде розраховуватися за формулою:

$$w_j^{(\mu,i)} = \begin{cases} 0, & \mu = 5, i = 1, j = 0; \\ 1, & \mu = 5, i = 1, j > 0; \\ 1, & \mu = 4, \forall i, j = 0; \\ 1 - Q, & \mu = 3, i = 1, j = 0; \\ 1, & \mu = 3, i = 1, j = 0; \\ 0, & \mu = 2, \forall i, j = 0; \\ 1, & \mu = 2, \forall i, j = 2; \\ -1, & \mu = 2, \forall i, j = 1; \\ \beta_{ij}, & \mu = 1, i=1, \dots, Q, j=1, \dots, N; \\ 0, & \mu = 1, i = q, 0 > j \neq q^q, q = 1, \dots, Q; \\ C_j^q, & \mu = 1, i = Q+1, Q+2, \dots, 2Q, j = 1, \dots, N. \end{cases}$$

Схема НМ, синтезованої та навченої на основі запропонованого алгоритму, буде виглядати також, як і зображена на рис. 2.

Запропонований алгоритм дозволяє досягати ще більшої точності у порівнянні із лінійною регресією, але необхідність розв'язання задачі багатовимірної нелінійної оптимізації для кожного кластеру призведе до суттєво більших витрат часу у процесі навчання мережі.

ПЕРЕЛІК ПОСИЛАНЬ

1. Дрейпер Н., Смит Г. Прикладной регрессионный анализ: В 2-х книгах, Кн. 1.-М.: Финансы и статистика, 1986.-366с., Кн. 2.-М.: Финансы и статистика, 1987.-351 с.
2. Гринчишин Я.Т., Ефимов В.И., Ломакович А.Н. Алгоритмы и программы на бейсике.-М.: Просвещение, 1988.-160 с.
3. Субботин С.А. Алгоритм планирования он-лайнного эксперимента в нейросетевой диагностике // Нейроинформатика и ее приложения / Материалы IX Всероссийского семинара, 5-7 октября 2001 г. / Под ред. А.Н.Горбаня. Отв. за выпуск Г.М.Цибульский.- Красноярск: КГТУ, 2001.- С.180-181.
4. Математическая экономика на персональном компьютере: Пер. с яп. / М. Кубонива, М. Табата, С. Табата, Ю. Хасэбэ; Под ред. М. Кубонива.-М.: Финансы и статистика, 1991.-304 с.

III. УПРАВЛІННЯ

УДК 621.314/611

АНАЛИТИЧЕСКИЙ МЕТОД НАХОЖДЕНИЯ ПЕРИОДИЧЕСКИХ НЕГАРМОНИЧЕСКИХ РЕШЕНИЙ В НЕКОТОРЫХ ЗАДАЧАХ ТЕОРИИ РАДИОТЕХНИЧЕСКИХ И ЭЛЕКТРОННЫХ ЦЕПЕЙ

М.Н.Горбачев

Обговорюється метод знаходження періодичних рішень, що дозволяє описати у замкнутій формі сталені процеси у лінійних радіотехнічних та електричних ланцюгах з зосередженими параметрами при живленні негармонічною напругою.

Обсуждается метод нахождения периодических решений, позволяющий описать в замкнутой форме установившиеся процессы в линейных радиотехнических и электрических цепях с сосредоточенными параметрами при питании негармоническим напряжением.

A method for the finding of periodic solutions which permits to represent in closed form steady - state processes in linear radiotechnic and electric circuits with lumped parameters when they are powered by nonharmonic voltage is discussed.

ВВЕДЕНИЕ

Предложен и разработан усовершенствованный метод нахождения в замкнутом виде периодических решений, описывающий установившиеся процессы в линейных цепях с сосредоточенными параметрами при существенно негармонических входных сигналах. Такого типа цепи широко применяются в радиотехнике, импульсной технике, преобразовательной технике и электросвязи (фильтры низкой частоты, полосовые и заграждающие фильтры, корректирующие цепи каналов связи и др.). Этот метод основан на использовании однозначной связи между частотными свойствами цепей и гармоническим спектром заданного входного сигнала с учетом законов линейных электрических цепей и является усовершенствованием и развитием известного метода гармонического синтеза, разработанного А. М. Заездным [1].

УСОВЕРШЕНСТВОВАННЫЙ МЕТОД ГАРМОНИЧЕСКОГО СИНТЕЗА

Нахождение указанных периодических решений по методу А. М. Заездного связано с выполнением значительного объема громоздких и трудоемких математических преобразований, вычислением корней характеристического уравнения и составлением вспомогательной функции, необходимых для нахождения частных (периодических) решений линейных неоднородных дифференциальных уравнений специального вида (1)

$$\frac{d^n y(x)}{dx^n} + a_{n-1} \frac{d^{n-1} y(x)}{dx^{n-1}} + \dots + a_1 \frac{dy(x)}{dx} + a_0 y(x) = F(x), \quad (1)$$

у которых правая часть $F(x)$ является не усеченным рядом Фурье, а содержит бесконечный спектр гармоник, что существенно ограничивает применение этого метода для цепей выше второго-третьего порядков.

Применение усовершенствованного метода гармонического синтеза (УМГС) позволяет значительно упростить процедуру нахождения периодических решений для указанного класса задач по сравнению с методом А.М.Заездного [1]. Покажем, что это именно так, следующим образом.

Пусть необходимо найти ток $i(x)$ на входе некоторой цепи, находящейся под воздействием негармонического сигнала в виде напряжения или э. д. с. $e(x)$, представляемого в виде не усеченного ряда Фурье

$$e(x) = \frac{1}{2} \alpha_0 + \sum_{k=1}^{\infty} (\alpha_k \cos kx + \beta_k \sin kx), \quad (2)$$

где $x = \omega t$, ω - круговая частота основной гармоники, t - текущая координата времени, α_0 , α_k , β_k - коэффициенты Фурье. Будем рассматривать исследуемую цепь как некоторый эквивалентный двухполосник, используя законы линейных электрических цепей (закон Ома, принцип суперпозиции) и функции входного комплексного сопротивления $Z(k\omega)$ или входной комплексной проводимости $Y(k\omega)$ [2]. Тогда можно составить выражение для k -ой гармоники искомого тока в общем виде, а затем представить искомый ток $i(x)$ также в виде некоторого ряда Фурье

$$i(x) = i_0 + \sum_{k=1}^{\infty} [a_k \cos(kx \pm \varphi_k) + b_k \sin(kx \pm \varphi_k)], \quad (3)$$

где

$$i_0 = \frac{\alpha_0}{2|Z_0|}; \quad a_k = \frac{\alpha_k}{|Z_k(k\omega)|}; \quad b_k = \frac{\beta_k}{|Z_k(k\omega)|}; \quad (4)$$

$\varphi_k = \varphi(k\omega)$ - фазовые углы между гармониками заданного напряжения $e(x)$ и искомого тока $i(x)$ порядка k .

Следовательно, задача сводится к суммированию нового ряда Фурье (3), если его можно просуммировать. Коэффициенты ряда (3) связаны с коэффициентами ряда (2) соотношениями (4). Отсюда следует, что если выполнены условия

$$0 < |Z_k(k\omega)| < \infty, \quad 0 < |Y(k\omega)| < \infty \quad (5)$$

и порядок λ стремления к нулю коэффициентов ряда (3) больше единицы,

$$|a_k|, |b_k| < \frac{h}{k^\lambda}, \quad \text{и } \lambda > 1, \quad (6)$$

где $h = \text{const}$, то ряд (3) сходится равномерно, что и гарантирует нахождение его суммы в замкнутом виде известными математическими методами [1, 3, 4].

Таким образом, в конечном итоге процедура нахождения частного решения дифференциального уравнения типа (1) с правой частью в виде ряда Фурье (2) для исследуемой цепи заменяется более простой и доступной процедурой построения преобразованного ряда Фурье (3) и его последующего суммирования с привлечением справочных материалов, что имеет большое значение в инженерной практике. При этом искомые периодические решения могут быть найдены аналитическим путем в замкнутом виде в двух вариантах: либо точно, либо приближенно. В последнем случае решение целесообразно находить в виде суммы двух составляющих - основной гармоники, которая всегда находится точно, и суммы всех остальных (высших) гармоник, которую можно и в ряде случаев целесообразно находить приближенно, используя аппроксимирующие (упрощенные) выражения $|\tilde{Z}_k(k\omega)|$ и $|\tilde{\Phi}_k(k\omega)|$ для соответствующих точных соотношений.

В качестве иллюстрации разработанного метода рассмотрим примеры решения модельных задач указанного класса для электрических цепей третьего-четвертого порядков.

Пример № 1. Необходимо определить закон изменения мгновенных значений входного тока $i(x)$ фильтра низкой частоты (ФНЧ), выполненного по известной симметричной T-образной схеме при следующих условиях:

1) нагрузка имеет активный (резистивный) характер $Z_H = R_H$ и ее относительная величина q_H изменяется в ограниченном диапазоне

$$0 < q_H = \frac{R}{\omega L} \leq q_{\max} = 2;$$

2) входной контур ФНЧ имеет резонансную настройку на частоте первой гармоники:

$$\omega_1^2 LC = 1,$$

где $L = L_1 = L_2$ и C - соответственно индуктивности и емкость элементов ФНЧ;

3) испытательное напряжение $e(x)$ имеет форму меандра, описываемого известным рядом Фурье [1]

$$\frac{4}{\pi} E \sum_{k=1}^{\infty} \frac{\sin kx}{k}, \quad (7)$$

где номер гармоники $k = 2n - 1$; $n = 1, 2, 3, \dots, \infty$ - числа натурального ряда. Для этого случая модули z_1, z_k, \tilde{z}_k и фазовые углы $\Phi_1, \Phi_k, \tilde{\Phi}_k$ входных комплексных сопротивлений рассматриваемой схемы ФНЧ с учетом условия $\omega_1^2 LC = 1$ приведены в табл.1.

Таблица 1

| Параметры схемы | Расчетные формулы к задаче 1 |
|---|---|
| $\frac{ z_1 }{\omega L} = \frac{z_1}{\omega L}$ | $\frac{1}{q}$ |
| $\frac{z_k}{\omega L}$ | $\frac{q(1-k^2) + jk(2-k^2)}{(1-k^2) + jkq}$ |
| $\frac{ z_k }{\omega L} = \frac{z_k}{\omega L}$ | $\frac{\sqrt{q^2 + k^2(1-k^2)^2(2-q^2-k^2)^2}}{k^4 + k^2(q^2-2) + 1}$ |
| $\frac{ \tilde{z}_k }{\omega L} = \frac{\tilde{z}_k}{\omega L}$ | $k - \frac{1}{k}$ |
| Φ_1 | 0 |
| Φ_k | $\arctg \frac{k(1-k^2)(2-q^2-k^2)}{q}$ |
| $\tilde{\Phi}_k$ | $\frac{\pi}{2}$ |

Как видно из табл. 1, в данном случае условие (5) выполнено. Из сравнения коэффициентов ряда Фурье (7) с коэффициентами нового ряда Фурье (3) для искомого тока $i(x)$ с учетом соотношений (4) и данных, приведенных в табл. 2, позволяющих сравнить сходимость рядов Фурье, видно, что условие (6) тоже выполнено.

Таблица 2

| Сравнение рядов Фурье | | | | | |
|-----------------------|---------------|-----------|---------------|-----------------|-----------|
| Ряд Фурье (7) | | | Ряд Фурье (3) | | |
| α_k | β_k | λ | a_k | b_k | λ |
| 0 | $\frac{1}{k}$ | 1 | 0 | $\frac{1}{k^2}$ | 2 |

Следовательно, периодическое решение для искомого входного тока $i(x)$ существует и может быть найдено в замкнутом виде. Найдем это решение в виде суммы двух составляющих

$$i(x) = i_1(x) + \sum_{k=3,5,\dots}^{\infty} \tilde{i}_k(kx). \quad (8)$$

На основании выражения (8) и формул, приведенных в табл.1, определяем составляющие входного тока

$$i_1(x) = \frac{4Eq_n}{\pi\omega L} \sin x; \quad (9)$$

$$\begin{aligned} \sum_{k=3,5,\dots}^{\infty} i_k(kx) &= -\frac{4E}{\pi\omega L} \sum_{n=1}^{\infty} \frac{\cos(2n+1)x}{(2n+1)^2-1} = \\ &= \frac{1}{2} \sin x \left\{ \sum_{n=1}^{\infty} \frac{\cos 2nx}{2n} - \sum_{n=1}^{\infty} \frac{\cos 2nx}{2n+2} \right\} + \\ &+ \frac{1}{2} \cos x \left\{ \sum_{n=1}^{\infty} \frac{\sin 2nx}{2n+2} - \sum_{n=1}^{\infty} \frac{\sin 2nx}{2n} \right\}. \quad (10) \end{aligned}$$

Выражение (10) с учетом соответствующих формул [7] после алгебраических и тригонометрических преобразований для $0 < x < \pi$ приводится к следующему виду:

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{\cos(2n+1)x}{(2n+1)^2-1} &= \frac{1}{2} \sin x [\cos 2x \ln(2 \sin x) + \\ &+ \left(x - \frac{\pi}{2} \right) \sin 2x + \frac{1}{2}]. \quad (11) \end{aligned}$$

Отсюда следует, что сумма всех высших гармоник, найденная в замкнутом виде путем сворачивания бесконечных функциональных рядов в выражении (10), имеет достаточно простое аналитическое представление (11) в классе элементарных функций от двух аргументов: $x = \omega_1 t$, $2x = 2\omega_1 t$.

Таким образом, на основании (9) и (11) находим общее приближенное решение для искомого тока для первой половины периода $0 < x < \pi$

$$\begin{aligned} i(x) &= \frac{4Eq_n}{\pi\omega L} \sin x - \frac{2E}{\pi\omega L} \sin x \left[\left(x - \frac{\pi}{2} \right) \sin 2x + \right. \\ &\left. + \cos 2x \ln(2 \sin x) + \frac{1}{2} \right]. \quad (12) \end{aligned}$$

Найдя входной ток и зная параметры цепи, можно определить токи во всех остальных ветвях схемы и напряжения на ее элементах.

Пример № 2. Необходимо найти периодические решения, описывающие установившиеся негармонические процессы в линейной двухконтурной электрической цепи с магнитной (трансформаторной) связью, применяемой для согласования сопротивления нагрузки и повышения избирательности системы с одновременным расширением ее полосы пропускания. Эта известная схема имеет

четвертый порядок и применяется в радиоприемных и радиопередающих устройствах. Электромагнитные процессы в контурах этой схемы описываются известной системой уравнений

$$e(t) = i_1 R_1 + L_1 \frac{di_1}{dt} + M \frac{di_2}{dt} + \frac{1}{C_1} \int i_1(t) dt, \quad (13)$$

$$0 = i_2 R_2 + L_2 \frac{di_2}{dt} + M \frac{di_1}{dt} + \frac{1}{C_2} \int i_2(t) dt, \quad (14)$$

где R_1 , R_2 , L_1 , L_2 , C_1 и C_2 - электрические параметры элементов соответственно входного и выходного контуров; M - коэффициент взаимной индукции; $e(t)$ - входной негармонический сигнал в виде э.д.с. или напряжения. Будем искать периодическое решение для тока $i_1(\omega t)$ первого (входного) контура во временной области методом УМГС полагая, что входной сигнал $e(t)$ имеет, как и ранее, форму меандра и описывается рядом Фурье (7), а параметры контуров одинаковы ($R_1=R_2=R$; $L_1=L_2=L$; $C_1=C_2=C$) и, следовательно, равны их резонансные частоты $\omega_{01} = \omega_{02} = \omega_0 = \frac{1}{\sqrt{LC}}$ и $h_1 = h_2 = h = \frac{\omega_0 L}{R}$.

Запишем искомое решение в общем виде

$$\begin{aligned} i_1(nx) &= \frac{4}{\pi} E \sum_{n=1,3,\dots}^{\infty} \frac{\sin(nx - \varphi_n)}{nZ_3(n)} = \\ &= \frac{4}{\pi} E \left\{ \sum_{n=1,3,\dots}^{\infty} \frac{R_3}{Z_3^2(n)} \frac{\sin nx}{n} - \sum_{n=1,3,\dots}^{\infty} \frac{X_3(n)}{Z_3^2(n)} \frac{\cos nx}{n} \right\}, \quad (15) \end{aligned}$$

где $x = \omega t$; $R_3 = R(1 + \gamma^2)$; $n = \omega/\omega_0$; $k_c = \frac{M}{\sqrt{L_1 L_2}} = \frac{M}{L}$.

Величины $X_3(n)$; и $Z_3(n)$, входящие в общее решение (15), являются нелинейными функциями от n и γ и имеют следующий вид:

$$X_3(n) = Rh(1 - \gamma^2) \left(n - \frac{1}{n} \right); \quad (16)$$

$$Z_3(n) = R \sqrt{(1 + \gamma^2)^2 + h^2(1 - \gamma^2)^2} \left(n - \frac{1}{n} \right)^2, \quad (17)$$

где

$$\operatorname{tg} \varphi_3(n) = h \frac{1 - \gamma^2}{1 + \gamma^2} \left(n - \frac{1}{n} \right); \quad (18)$$

$$\gamma = \frac{nhk_c}{\sqrt{1+h^2}n - \frac{1}{n}} = f(n) . \quad (19)$$

Опуская промежуточные математические преобразования, можно показать, что искомое точное периодическое решение для тока входного (первого) контура имеет следующий замкнутый вид:

$$i_1(x) = \frac{E}{R} \cdot \frac{1}{h} \frac{1}{(1-\gamma^2)\sqrt{\frac{a^2}{4}-1}} \left[\frac{e^{n_1x}}{1+e^{n_1x}} - \frac{e^{n_2x}}{1+e^{n_2x}} \right], \quad (20)$$

где

$$n_1 = -\frac{a}{2} + \sqrt{\frac{a^2}{4}-1};$$

$$n_2 = -\frac{a}{2} - \sqrt{\frac{a^2}{4}-1};$$

$$a = \frac{1}{h} \cdot \frac{1+\gamma^2}{1-\gamma^2}.$$

Отметим, что эффективность применения предложенного приближенного метода возрастает с ростом порядка цепи. Это особенно заметно проявляется при исследовании реактивных двухполосников n -го порядка, называемых каноническими схемами Фостера [2]; при синтезе реактивных двухполосников, имеющих вид лестничных схем, так называемых канонических схем Кауэра [2]; при расчете корректирующих контуров, применяемых в телефонной и телеграфной связи [2]; при исследовании и расчете фильтров низкой частоты (ФНЧ) 2-9-го порядков, например, фильтров Баттерворта, Гаусса (Бесселя), Кауэра, Чебышева [5] и др.

Таким образом, преимущества усовершенствованного метода УМГС нахождения в замкнутом виде периодических решений в задачах указанного класса заключается в его доступности, простоте и исключении трудоемкой процедуры непосредственного решения неоднородных дифференциальных уравнений с правой частью специального вида [1]. Общий вид получаемых периодических решений (как точных, так и приближенных) явно не выражается через корни характеристических уравнений, соответствующих дифференциальным уравнениям исследуемых электрических цепей. Поэтому эффективность применения разработанного метода УМГС возрастает с увеличением порядка цепи, например, при расчете фильтров низкой частоты, при исследовании канонических схем Фостера и Кауэра, корректирующих фильтров в устройствах и линиях электросвязи и др.

На приведенных примерах решения модельных задач рассматриваемого класса для цепей третьего и четвертого порядков усовершенствованным методом гармонического синтеза показана возможность и целесообразность дальнейшего использования математического аппарата гар-

монического синтеза и теории функциональных рядов для аналитического исследования периодических негармонических процессов в электрических цепях и системах электросвязи в гораздо большей степени, чем это делается в настоящее время.

ОЦЕНКА ПОГРЕШНОСТИ ПРИБЛИЖЕННОГО РЕШЕНИЯ

В общем случае погрешность получаемого приближенного решения $\tilde{y}(x) = \tilde{i}(x)$ можно косвенно оценить путем раздельной численной оценки амплитудной $\sigma_z(k\omega)$ и фазовой $\sigma_\varphi(k\omega)$ составляющих общей погрешности с учетом заданных значений параметра q_n на основе формул

$$\sigma_z(k\omega) = \sqrt{\sum_{k=2}^{\infty} \delta_z^2(k\omega, q_n)} \approx \sqrt{\sum_{k=2}^l \delta_z^2(k\omega, q_n)}; \quad (21)$$

$$\sigma_\varphi(k\omega) = \sqrt{\sum_{k=2}^{\infty} \delta_\varphi^2(k\omega, q_n)} \approx \sqrt{\sum_{k=2}^l \delta_\varphi^2(k\omega, q_n)}, \quad (22)$$

где l - достаточное для практических целей число гармоник (обычно $3 \leq l \leq 15$, теоретически $l \rightarrow \infty$), а $\delta_z(k\omega, q_n)$ и $\delta_\varphi(k\omega, q_n)$ определяются по формулам

$$\delta_z(k\omega, q_n) = \left| 1 - \frac{\tilde{z}_k(k\omega, q_n)}{z_k(k\omega, q_n)} \right|; \quad (23)$$

$$\delta_\varphi(k\omega, q_n) = \left| 1 - \frac{\tilde{\Phi}_k(k\omega, q_n)}{\Phi_k(k\omega, q_n)} \right|. \quad (24)$$

ЗАКЛЮЧЕНИЕ

Предложенный и разработанный усовершенствованный метод гармонического синтеза и приведенные решения некоторых модельных задач указанного класса на основе этого нетрадиционного метода позволяют сделать следующие основные выводы.

1. Преимущество предложенного метода нахождения в замкнутом виде периодических решений для рассматриваемого класса задач по сравнению с другими методами, заключается, прежде всего, в его доступности и простоте, а также в возможности учета всех гармонических составляющих, адекватных полному спектру возмущающего воздействия, что представляет значительный интерес для исследователей этого класса задач теории линейных электрических и радиотехнических цепей. При этом решение находится обычно в классе элементарных функций.

2. Общий вид получаемого периодического решения явно не выражается через корни характеристического уравнения, соответствующего дифференциальному уравнению исследуемой цепи, поэтому эффективность

применения предложенного метода и разработанных на его основе алгоритмов и инженерной методики получения искомых решений и особенно приближенных решений возрастает с ростом порядка цепи.

3. Получение при этом замкнутого приближенного периодического решения в виде двух составляющих позволяет также легко определять основные энергетические и спектральные характеристики исследуемых электрических и радиотехнических цепей (коэффициенты гармоник и искажения, а также коэффициент мощности), что является, однако, самостоятельной задачей.

УДК 62-55:681.515

СИНТЕЗ ОПТИМАЛЬНЫХ ПО БЫСТРОДЕЙСТВИЮ ЦИФРОВЫХ РЕГУЛЯТОРОВ ДЛЯ ОБЪЕКТОВ УПРАВЛЕНИЯ С ЧИСТЫМ ЗАПАЗДЫВАНИЕМ

В.И.Гостев, В.В.Крайнев, С.Н.Скуртов

Викладені особливості синтезу оптимальних по швидкодії цифрових регуляторів для об'єктів з чистим запізнюванням. Надані результати дослідження процесів в системі автоматичного керування температурою електронечі при використанні оптимальних по швидкодії цифрових регуляторів.

Изложены особенности синтеза оптимальных по быстродействию цифровых регуляторов для объектов управления с чистым запаздыванием. Представлены результаты исследования процессов в системе автоматического управления температурой электронечи при использовании оптимальных по быстродействию цифровых регуляторов.

The features of synthesis optimal on speed of digital regulators for objects of control with a dead time are explained. The results of research of processes in the system of automatic control of temperature of the electric furnace at usage optimal on speed of digital regulators are represented.

1 ПОСТАНОВКА ЗАДАЧИ

Математические модели многих промышленных объектов, рассматриваемых как объекты управления, могут быть представлены структурными схемами, содержащими звено чистого запаздывания. В частности, к таким объектам относятся парогенераторы, водяные котлы, электрические и газовые печи. Значительное место среди мероприятий по экономии топлива и электроэнергии занимает автоматизация процесса получения и потребления тепловой и электрической энергии и актуальной проблемой является разработка автоматов и систем автоматического управления параметрами этих промышленных объектов [1]. При разработке систем управления объектами с чистым запаздыванием одной из важных задач является синтез регуляторов, учитывающий указанную особенность объектов управления [2]. В известных работах [1,2] эта задача решалась на основе применения

ПЕРЕЧЕНЬ ССЫЛОК

1. Заездный А.М. Гармонический синтез в радиотехнике и электросвязи. - М. - Л.: Госэнергоиздат, 1961. - 535 с.
2. Зевеке Г. В., Ионкин П.А., Нетушил А.В., Страхов С.В. Основы теории цепей. Изд. 2-е. - М.-Л.: Госэнергоиздат, 1963. - 440 с.
3. Толстов Г. П. Ряды Фурье.- М.: Наука, 1980. - 348 с.
4. Фихтенгольц Г. М. Курс дифференциального и интегрального исчисления. Т. 2. - М.: Гос. изд-во физ. - мат. литер., 1959. - 307 с.
5. Ханзел Г. Е. Справочник по расчету фильтров. США, Нью-Йорк, 1969. Пер. с англ. под ред. А. Е. Знаменского. - М.: Советское радио, 1974. - 288 с.

робастных регуляторов. В данной работе решение этой задачи выполнено на основе синтеза оптимальных по быстродействию цифровых регуляторов.

2 РЕШЕНИЕ ЗАДАЧИ

Синтез оптимальных по быстродействию цифровых регуляторов для объектов управления с чистым запаздыванием рассмотрен на примере конкретного объекта - электронечи большой мощности, передаточная функция линейной модели которой определена в работе [2] в виде

$$G(s) = \frac{\alpha}{(s+a)(s+b)} e^{-\tau s}, \quad (1)$$

где $a = \frac{1}{T_1}$; $b = \frac{1}{T_2}$; $\alpha = \frac{K_0}{T_1 T_2}$. Полученные в результате экспериментальных измерений значения постоянных времени и коэффициента усиления следующие: $T_2 = 122 \text{ с}$, $T_1 = 14,5 \text{ с}$, $\tau = 3,9 \text{ с}$, $K_0 = 7,2 \text{ }^\circ\text{C}/\%$ (управляющее воздействие на объект управления задается в процентах перемещения регулирующего органа). В качестве исполнительных устройств обычно используются устройства типа "электродвигатель + регулируемый вентиль". Поскольку постоянная времени электродвигателя намного меньше постоянных времени в передаточной функции (1), то исполнительное устройство можно представить интегрирующим звеном с передаточной функцией

$G_1(s) = \frac{1}{s}$. Тогда передаточная функция общего объекта управления для регулятора определяется в виде

$$G_0(s) = \frac{\alpha}{s(s+a)(s+b)} e^{-\tau s}. \quad (2)$$

При проектировании систем управления с оптимальными по быстродействию цифровыми регуляторами и объектами управления с чистым запаздыванием так, чтобы система "аппроксимировала" произвольное входное воздействие сигналом, который линейно изменяется на интервалах регулирования, необходимо определять оптимальные управляющие воздействия на объект управления. Эти воздействия для объекта с передаточной функцией $G(s)e^{-\tau s}$ можно найти, модифицируя полученные в работе [3] формулы для оптимальных управляющих воздействий на объект управления с передаточной функцией $G(s)$ [4]. Так, на вход объекта управления с передаточной функцией $G(s) = \alpha e^{-\tau s} [s(s+a)(s+b)]^{-1}$ необходимо на каждом интервале регулирования t_p , начиная с момента nt_p , $n = 0, 1, 2, \dots$, подавать воздействия $m_0 = K_0(\Delta U + S_0 \Delta \sigma) + R \sigma_{n-1}$ при $nt_p \leq t < nt_p + h$; $m_1 = K_0[q_1(\Delta U + S_0 \Delta \sigma) + h \Delta \sigma] + R \sigma_{n-1}$ при $nt_p + h \leq t < nt_p + 2h$; $m_2 = K_0[q_2(\Delta U + S_0 \Delta \sigma) + h \Delta \sigma(1 + q_1)] + R \sigma_{n-1}$ при $nt_p + 2h \leq t < nt_p + 3h$;

$$m_3 = R = \frac{ab}{\alpha} \text{ при } nt_p + 3h \leq t < nt_p + 3h + \tau; \quad t_p = 3h + \tau; \quad (3)$$

где $K_0 = \frac{ab}{\alpha h(1-A)(1-B)}$; $S_0 = 3h + \frac{a+b}{ab} - \frac{h(2+q_1)}{(1-A)(1-B)} + \tau$; $q_1 = -(A+B)$; $q_2 = AB$; $A = e^{-ah}$; $B = e^{-bh}$.

Цифровой регулятор на каждом подынтервале $nt_p \leq t < nt_p + 3h$ интервала регулирования $nt_p \leq t < (n+1)t_p$ можно описать передаточной функцией

$$W(z) = \frac{M(z)}{\theta(z)} = \frac{m_0 + m_1 z^{-1} + m_2 z^{-2}}{\Delta U(1 + z^{-1} + z^{-2})} \quad (4)$$

или разностным уравнением

$$m_i = \left\{ \sum_{k=0}^2 m_k \theta_{i-k} - \Delta U \sum_{k=1}^2 m_{i-k} \right\} / \Delta U, \quad (5)$$

где $\theta = \Delta U$ при индексе $i-k \geq 0$ и $\theta = 0$, $m = 0$ при индексе $i-k < 0$.

Если обозначить через ΔU_i ошибку в соответствующем контуре управления в моменты ih , $i = 0, 1, 2, \dots$ на интервале регулирования t_p (ΔU_0 - ошибка в момент nt_p , ΔU_1 - ошибка в момент, $nt_p + h$, ΔU_2 - ошибка в момент $nt_p + 2h$), то цифровой регулятор на каждом

подынтервале $nt_p \leq t < nt_p + 3h$ интервала регулирования $nt_p \leq t < (n+1)t_p$ можно описать передаточной функцией

$$W(z) = \frac{M(z)}{\theta(z)} = \frac{m_0 + m_1 z^{-1} + m_2 z^{-2}}{\Delta U_0 + \Delta U_1 z^{-1} + \Delta U_2 z^{-2}} \quad (6)$$

или разностным уравнением

$$m_i = \left\{ \sum_{k=0}^2 m_k \Delta U_{i-k} - \sum_{k=1}^2 \Delta U_k m_{i-k} \right\} / \Delta U_0, \quad (7)$$

где h - шаг квантования. На $(n-1)$ -м интервале регулирования $(n-1)t_p \leq t < nt_p$ входное воздействие характеризуется первой разностью (средней на интервале скоростью) σ_{n-1} , на n -м интервале регулирования $nt_p \leq t < (n+1)t_p$ - первой разностью (средней на интервале скоростью) σ_n ; приращение средней скорости на интервале регулирования $nt_p \leq t < (n+1)t_p$ можно определить как $\Delta \sigma = \sigma_n - \sigma_{n-1}$, $n = 0, 1, 2, \dots$.

Текущее значение скорости входного воздействия можно определить по формуле

$$\sigma = \{u(kh_0) - u[(k-1)h_0]\} / h_0, \quad (8)$$

где h_0 - шаг моделирования, и использовать приближенное значение первой разности

$$\hat{\sigma}_n = \sigma(nt_p). \quad (9)$$

Первая разность (средняя скорость) входного воздействия на предыдущем интервале регулирования $(n-1)t_p \leq t < nt_p$ определяется как

$$\hat{\sigma}_{n-1} = \sigma[(n-1)t_p]. \quad (10)$$

Тогда приращение скорости на интервале регулирования $nt_p \leq t < (n+1)t_p$ определяется как

$$\Delta \sigma = \{\sigma(nt_p) - \sigma[(n-1)t_p]\} / t_p. \quad (11)$$

3 ОПТИМАЛЬНЫЕ СТРУКТУРЫ РЕГУЛЯТОРОВ

Для математических моделей объектов управления с чистым запаздыванием (см. формулу (2)) можно предложить общую структурную схему оптимальных по быстродействию цифровых регуляторов, из которой можно получить несколько вариантов структурных схем, которые отличаются только блоками, реализующими различные соотношения времени запаздывания τ и шага квантования в цифровом регуляторе h . Для объекта с

передаточной функцией (2) рассмотрим три варианта:
 1. Оптимальный по быстродействию цифровой регулятор типа А, для которого примем $\tau = h = 4c$ (τ в реальном объекте равно $3,9c$);
 2. Оптимальный по быстродействию цифровой регулятор типа Б, для которого примем $\tau = 2h = 4c$;
 3. Оптимальный по быстродействию цифровой регулятор типа В, для которого примем $\tau = 4h = 4c$.
 Регуляторы, названные условно типами А, Б и В, имеют одни и те же передаточные функции, но отличаются только шагом квантования (для типа А $h = 4c$, для типа Б $h = 2c$, для типа В $h = 1c$).

Нижерассмотрена цифровая система управления температурой электропечи, в которой используются: 1 - оптимальный по быстродействию цифровой регулятор типа А; 2 - оптимальный по быстродействию цифровой регулятор

типа Б и 3 - оптимальный по быстродействию цифровой регулятор типа В. Исследование процессов в системе с указанными регуляторами при различных входных воздействиях проведено с использованием интерактивной системы MATLAB. Структурная схема системы приведена на рис.1.

Общая структурная схема оптимального по быстродействию цифрового регулятора состоит из двух субблоков (на рис.1 эти субблоки обозначены SubSystem1 и SubSystem2).

Субблок SubSystem1 представляет собой оптимальный по быстродействию цифровой регулятор, проектируемый для объекта управления без учета звена запаздывания. Структурная схема этого цифрового регулятора (приведена на рис.2) выполняется непосредственно на основании формул (3), (8)-(11).

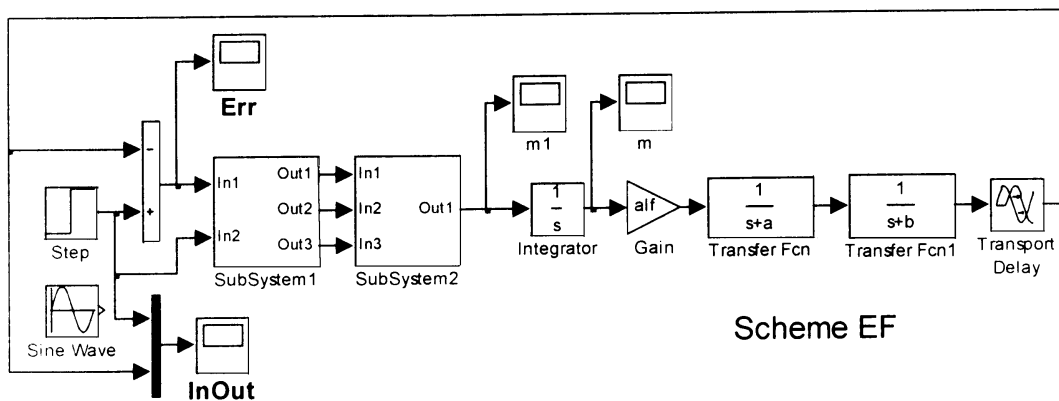


Рисунок 1

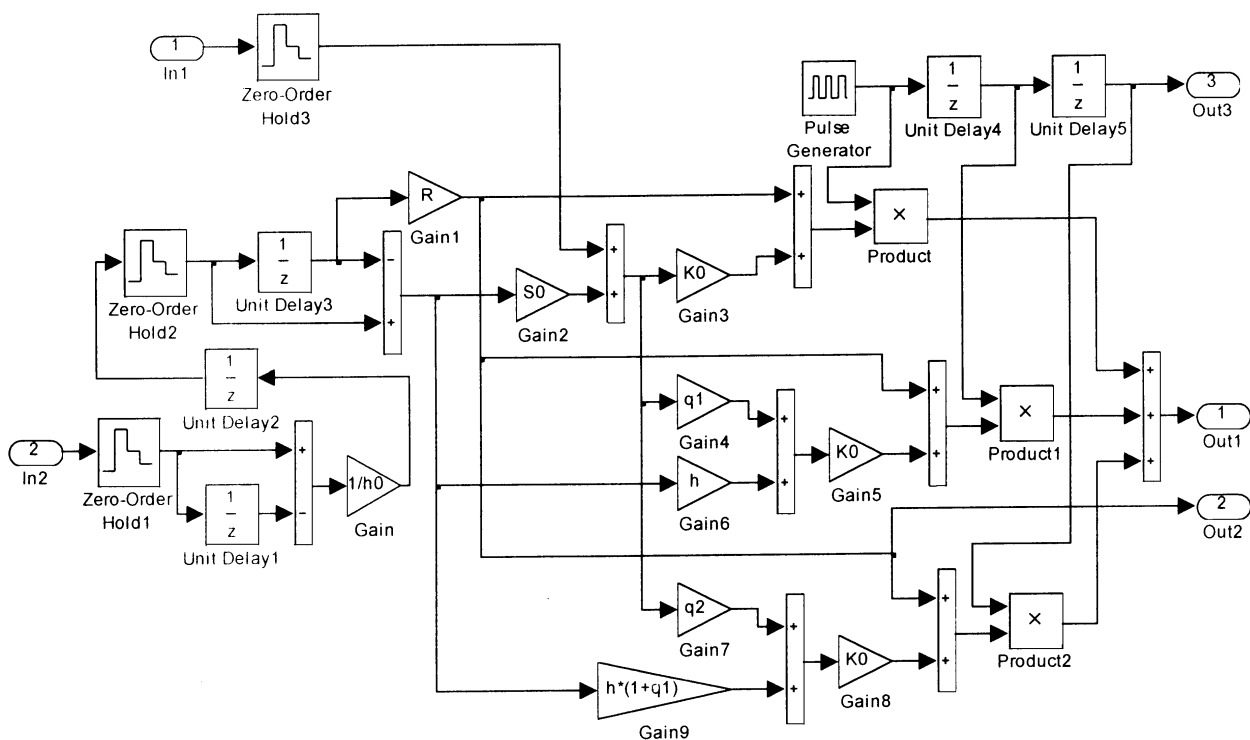


Рисунок 2

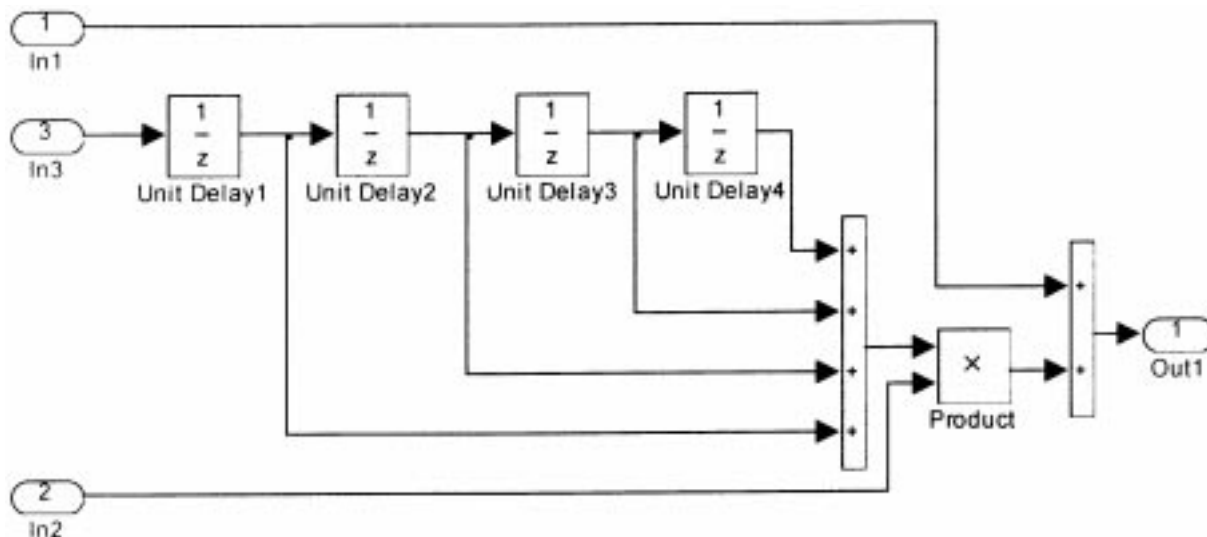


Рисунок 3

Субблок SubSystem2, структурная схема которого приведена на рис.3, представляет дополнительную схему, которая проектируется специально для учета запаздывания в объекте управления. Для оптимального по быстродействию цифрового регулятора типа А, у которого $h = \tau = 4c$, в этом субблоке используется только одно звено запаздывания Unit Delay1 (остальные звенья запаздывания отключены). Для оптимального по быстродействию цифрового регулятора типа Б, у которого $h = \tau/2 = 2c$, в этом субблоке используются два звена запаздывания Unit Delay1 и Unit Delay2 (остальные звенья запаздывания отключены). Для оптимального по быстродействию цифрового регулятора типа В, у которого $h = \tau/4 = 1c$, в этом субблоке используются все четыре звена запаздывания Unit Delay1 - Unit Delay4.

4 РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

На рис.4 показаны переходные процессы (реакции системы на единичное ступенчатое воздействие) при наличии в системе оптимальных по быстродействию цифровых регуляторов типа А, для которого $\tau = h = 4c$ (кривая 1), типа Б, для которого $\tau = 2h = 4c$ (кривая 2) и типа В, для которого $\tau = 4h = 4c$ (кривая 3).

Для системы с оптимальным по быстродействию цифровым регулятором типа А время регулирования $T_p = 16c$; для системы с оптимальным по быстродействию цифровым регулятором типа Б время регулирования $T_p = 10c$; для системы с оптимальным по быстродействию цифровым регулятором типа В время регулирования $T_p = 7c$. Таким образом, быстродействие системы с регуляторами типа Б и В превышает быстродействие системы с регулятором типа А соответственно 1,6 раза и примерно в 2,3 раза. Оптимальные по быстродействию цифровые регуляторы обеспечивают апе-

риодические переходные процессы (без перерегулирования).

Можно еще больше уменьшить время регулирования T_p для системы с оптимальными по быстродействию цифровыми регуляторами, уменьшая шаг квантования (например, приняв $\tau = 8h = 4c$, $h = 0,5c$), но при этом несколько усложняется схема субблока SubSystem2 и, главное, значительно возрастают управляющие импульсы на выходе регулятора, что требует применения дополнительного усиления сигналов.

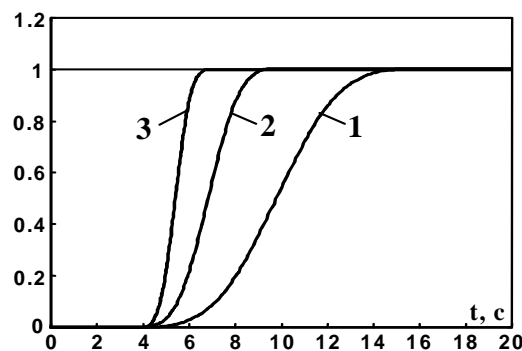


Рисунок 4

При произвольном входном воздействии $u(t)$, которое изменяется с максимальной скоростью ω_{\max} и максимальным ускорением ϵ_{\max} , удобно рассматривать эквивалентное гармоническое воздействие

$$u_{\text{э}}(t) = U_{\text{э}} \sin \omega_{\text{э}} t, \quad (12)$$

параметры которого определяются из соотношений [4]

$$U_{\text{э}} = \frac{\omega_{\max}^2}{\epsilon_{\max}}, \quad \omega_{\text{э}} = \frac{\epsilon_{\max}}{\omega_{\max}}. \quad (13)$$

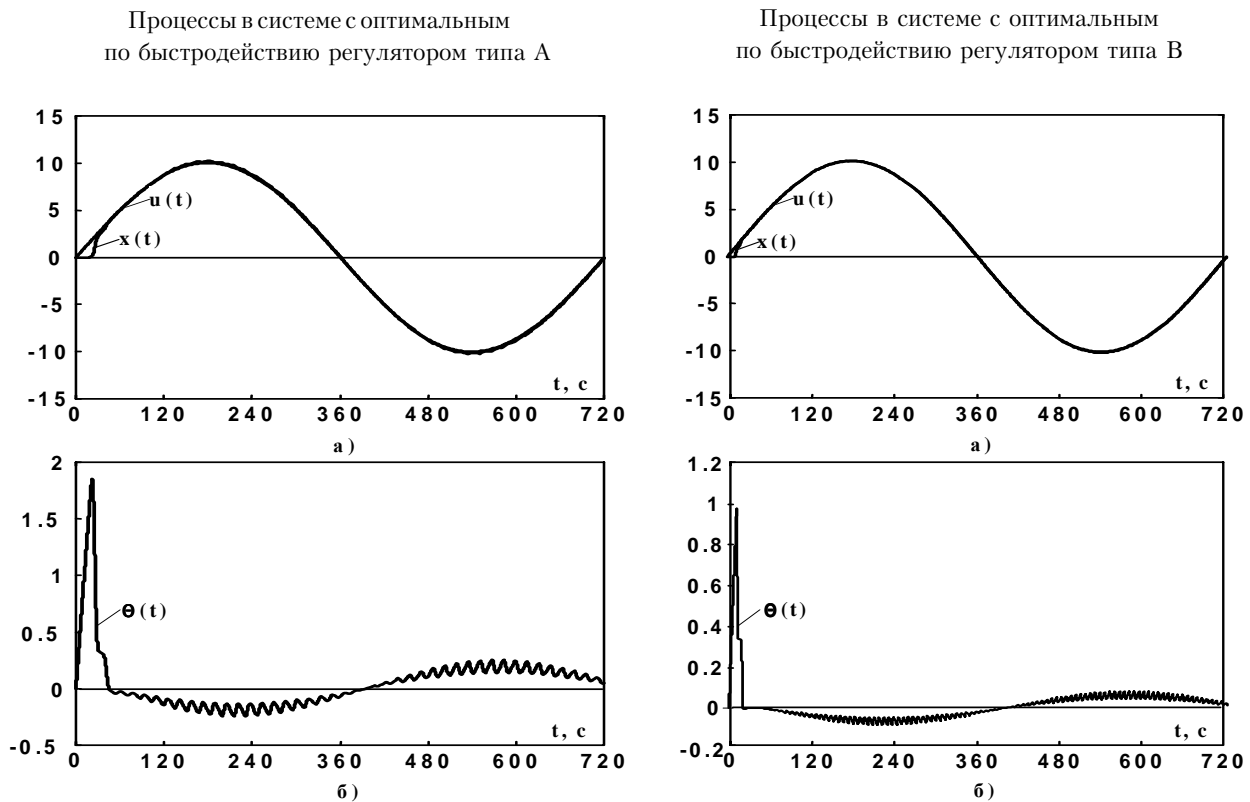


Рисунок 5

В свою очередь, если заданы параметры эквивалентного гармонического воздействия, то максимальная скорость ω_{\max} и максимальное ускорение ε_{\max} произвольного входного воздействия находятся из соотношений

$$\omega_{\max} = U_3 \omega_3; \quad \varepsilon_{\max} = U_3 \omega_3^2. \quad (14)$$

Процессы в системе (см. рис.1) с оптимальными по быстродействию цифровыми регуляторами типа А и В при поступлении на вход системы эквивалентного гармонического воздействия

$$u_3(t) = 10 \sin(\pi t / 360), \quad (15)$$

где $U_3 = 10$; $\omega_3 = \pi / 360 \approx 8,72 \times 10^{-3} \text{ рад} / \text{с}$, приведены на рис.5, где $u(t)$ - вход системы, $x(t)$ - выход системы, $\theta(t)$ - ошибка рассогласования.

Отслеживание входного эквивалентного гармонического воздействия в системах с указанными регуляторами (без учета большого выброса при захвате входного сигнала) достаточно точное: максимальная текущая ошибка в системе с оптимальным по быстродействию регулятором типа А достигает примерно 7% от амплитуды входного воздействия, а максимальная текущая ошибка в системе с оптимальным по быстродействию регулятором типа В составляет примерно 3%.

ВЫВОДЫ

Проведенные исследования процессов в системе автоматического управления температурой электропечи показывают, что оптимальные по быстродействию цифровые регуляторы можно эффективно использовать не только при ступенчатых входных сигналах, но и при произвольных воздействиях с ограниченной скоростью изменения и ускорением. Следует считать, что применение оптимальных по быстродействию цифровых регуляторов для стационарных объектов управления целесообразно и дальнейшая разработка таких регуляторов для различных объектов с чистым запаздыванием перспективна.

ПЕРЕЧЕНЬ ССЫЛОК

1. Ротач В. Я. Теория автоматического управления теплоэнергетическими процессами. М.: Энергоатомиздат, 1985. - 296 с.
2. Ротач В.Я. Расчёт робастной настройки автоматических регуляторов // Теплоэнергетика. - 1994. - №10. - С.7-12.
3. Гостев В.И., Стеклов В.К. Системы автоматического управления с цифровыми регуляторами: Справочник. - К.: "Радиоаматор", 1998. - 704 с.
4. Гостев В.И., Лесовой И.П., Чуприн А.Е. Применение оптимальных по быстродействию цифровых регуляторов для объектов управления с чистым запаздыванием // Радиоэлектроника. Информатика. Управление. - 2000. - N2. - С.6-11.
5. Бесекерский В.А. Динамический синтез систем автоматического регулирования. - М.: Наука, 1970. - 576с.

МИКРОПРОЦЕССОРНАЯ СИСТЕМА УПРАВЛЕНИЯ КОМПЛЕКСОМ ПО ПЕРЕРАБОТКЕ ОТХОДОВ ПОЛИМЕРНЫХ МАТЕРИАЛОВ

В.В.Дядичев

Запропоновано автоматизовану систему керування процесом готування багатокomпонентної соекструзійної суміші з відходів полімерних матеріалів. Розроблено алгоритм функціонування системи, засобів керування комплексом, апаратна і програмна частини мікропроцесорної системи оперативного керування якістю готування соекструзійної суміші.

Предложена автоматизированная система управления процессом приготовления многокомпонентной соекструзионной смеси из отходов полимерных материалов. Разработан алгоритм функционирования системы, средств управления комплексом, аппаратная и программная часть микропроцессорной системы оперативного управления качеством приготовления соекструзионной смеси.

The automated management system of process of cooking multicomponent coextrusion of a mixture from waste of polymer stuffs is offered. The operation algorithm of a system both controls a complex and skeleton diagram of a system of operating control by quality cooking coextrusion of a mixture is designed.

Одним из направлений современного развития экономики Украины является создание новых и интенсификация действующих ресурсосберегающих производств на основе широкого использования передовых технологий, средств и систем автоматизированного управления технологическими процессами.

Проблема. В последнее время все большее значение в мире занимают проблемы загрязнения окружающей среды. И в первую очередь это касается утилизации отходов. Существующие средства и уровень автоматизации комплексов для утилизации отходов полимерных материалов не могут обеспечить необходимую точность подачи материалов смеси, а используемая жесткая установка заданий расхода компонентов смеси без учета их реальных расходов не позволяет достичь требуемого качества смеси из вторичных полимеров [1,2,3]. Общий уровень автоматизированного управления оборудованием на комплексах СНГ уступает уровню аналогичных производств в США, Канаде, Австралии, Швеции, Франции, Финляндии [4,5,6,7].

Такое состояние в уровне автоматизации процесса создания многослойных изделий из использованных полимерных материалов методом соекструзии определило необходимость в разработке автоматизированной системы управления технологическим процессом приготовления соекструзионной смеси, базирующейся на новых методах и принципах управления процессом, современных компьютерных средствах управления, учета и дозирования материалов.

Проведенный анализ существующих технологических схем рециклинга пластмасс и соответственно систем управления этими технологическими процессами позво-

ляет сделать вывод о том, что существующие системы управления процессом приготовления смеси не могут обеспечить необходимое качество (однородность) полимерной смеси в силу следующих причин:

- при приготовлении соекструзионной смеси используются наполненные отходы и вторичный полимер, а они являются трудно дозируемыми материалами. Поэтому использование дозатора вторичных полимеров, наряду с другими дозаторами, в режиме управляемого не обеспечит требуемую эффективность связанного дозирования;

- в существующих схемах не принимается во внимание пространственное размещение средств дозирования компонентов, т.е. при формировании управляющих воздействий не учитываются транспортные запаздывания поступления компонентов в экструдера;

Целью работы является создание современной системы автоматизированного управления процессом приготовления соекструзионной смеси из вторичных полимеров на основе моделирования процесса и с учетом транспортного запаздывания поступления компонентов смеси в экструдера с обеспечением изготовления нового изделия необходимого качества.

Анализ технологической схемы комплекса по переработке отходов полимерных материалов позволяет определить следующие функции системы контроля и управления процессом приготовления смеси: контроль расхода компонентов полимерной смеси (наполненные отходы, вторичный полимер, первичный полимер, наполнитель); контроль плотности выходного материала дробилки; контроль давления воздуха в магистрали и учет расхода компонентов полимерной смеси; расчет объема приготовленного соекструзионного материала за смену, сутки, месяц; стабилизация подачи компонентов; корректировка заданий на расходы компонентов с учетом их наполненности; корректировка расхода первичного полимера по реальному значению подачи компонентов смеси; подача компонентов в заданном соотношении к вяжущему.

Система автоматизированного управления процессом приготовления многокомпонентной соекструзионной смеси аппаратно может быть реализована либо на базе универсальной ПЭВМ и специального набора устройств связи с объектом (УСО), либо на основе специализированной микропроцессорной системы управления с использованием однокристалльного микроконтроллера. Нами был выбран последний вариант, поскольку система управления со встроенным современным микроконтроллером, как по сложности технической реализации, так и с экономической точки зрения является гораздо более эффективной для решения подобных задач [8,9]. Интегрированные в микроконтроллер периферийные устройства позволяют

реализовать все необходимые функции управления и связи с объектами с использованием минимального количества дополнительных аппаратных модулей, получить относительно простую, технологичную в производстве конструкцию с низким энергопотреблением и всеми необходимыми возможностями.

Разработанный алгоритм функционирования системы и средств управления позволяют определить архитектуру системы управления (рис.1), которая коммутирует 9 входных сигналов (0-5 мА), 8 входных и 12 выходных позиционных сигналов, 4 импульсных входных сигналов (0-8 кГц) и 14 аналоговых выходных сигналов с гальванической развязкой (0-5 мА). В состав микропроцессорной системы управления входят: однокристалльный микроконтроллер, постоянная электрически перепрограммируемая память (EEPROM) нормирующие преобразователи аналогового ввода, модули аналогового вывода, ввода/вывода позиционных сигналов, ввода импульсных сигналов, модули связи с принтером и ЭВМ верхнего уровня, дисплей, клавиатура, стабилизатор напряжения питания. Все основные модули системы, кроме клавиатуры и дисплея, выполнены в виде одноплатной конструкции. Расположенные на плате разъемы позволяют при необходимости легко расширить возможности системы путем добавления дополнительных модулей ввода/вывода. Применение комплектующих элементов ведущих мировых фирм-производителей (Microchip, Analog Devices, Power-

tip, Vorpla) обеспечило высокую надежность, компактность и низкое энергопотребление разработанного устройства [8, 9].

Примененный микроконтроллер PIC18C858 производства фирмы Microchip имеет следующие периферийные устройства, используемые системой управления:

- 16-ти канальный 10-ти битный аналого-цифровой преобразователь;
- 4 таймера-счетчика;
- широтно-импульсный модулятор с разрешением 10 бит;
- Синхронный последовательный порт SPI;
- Интерфейс I2C;
- Асинхронный порт USART;
- Сторожевой таймер.

Нормирующие преобразователи обеспечивают подключение к интегрированному в микроконтроллер аналого-цифровому преобразователю одного из 9-ти входных сигналов стандартного уровня (0-5 мА). Аналоговые входы системы подключаются к датчикам: расхода наполненных отходов (весы ВЕ4050), вторичного полимера (весы ВЕ4050), первичного полимера (дозаторы ВД4011), наполнителя (ДМ5085), датчику плотности смеси (ДПИ-103). Возможности расширения предусматривают подключение еще до 7-ми дополнительных входов (датчики заполнения бункеров 1-5, расход воздуха при транспортировке соэкструзионной смеси 6, 7).

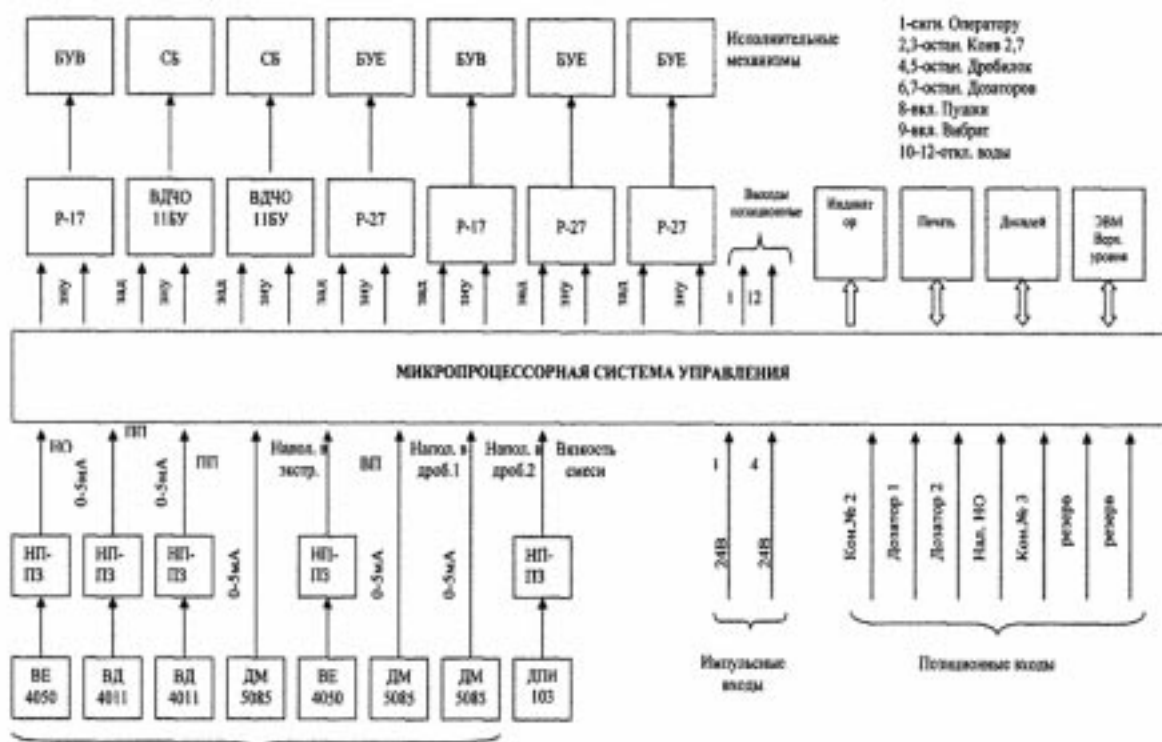


Рисунок 1 - Структурная схема системы оперативного управления качеством приготовления соэкструзионной смеси

Модули аналогового выхода предназначены для преобразования восьмиразрядного, десятиразрядного или двенадцатиразрядного двоичного кода в аналоговый сигнал (0-5 мА) по четырнадцати каналам. Функционально каждый канал выполнен на специализированном интегральном цифро-аналоговом преобразователе TC1322 производства фирмы Microchip и операционном усилителе, обеспечивающем преобразование выходного напряжения ЦАП в токовый сигнал. Управление преобразователями осуществляется по стандартному 2-х проводному интерфейсу I2C, что позволило максимально минимизировать требуемые линии связи для большого количества модулей. В качестве общего для всех аналого-цифровых преобразователей используется прецизионный высокостабильный источник опорного напряжения AD780 фирмы Analog Devices, имеющий очень малый температурный дрейф. Алгоритм управления модулями ЦАП и схемотехническое решение позволяют легко настроить каждый канал на преобразование с необходимой разрешающей способностью (8, 10 или 12 бит).

Модуль ввода/вывода позиционных сигналов предназначен для ввода/вывода восьмиразрядной дискретной информации. В схеме ввода позиционных сигналов для обеспечения гальванической развязки между схемой и датчиками используются оптроны АОТ110. Для гальванической развязки схемы вывода позиционных сигналов и исполнительных механизмов применены твердотельные электронные реле D2425 производства фирмы Crydom. Эти реле имеют очень малый ток управления, позволяющий подключать их непосредственно к выходам микроконтроллера без дополнительных согласующих выходных ключей. Отсутствие дребезга контактов, присущего обычным механическим реле, обеспечивает высокую надежность и помехоустойчивость системы, а большой максимальный коммутируемый ток (до 25 А) позволяет управлять мощной нагрузкой.

Микропроцессорная система может работать автономно без подключения к ней ЭВМ верхнего уровня. Для этого в ней имеется встроенный ЖКИ дисплей для отображения информации и клавиатура, содержащая 22 клавиши. В качестве ЖКИ дисплея применен функционально законченный модуль PC2004В производства фирмы Powertip. Это знаковосинтезирующий ЖКИ модуль, позволяющий отображать цифро-буквенную информацию и специальные графические символы в 4-х строках по 20 символов в каждой. Встроенный контроллер HD44780 обеспечивает все функции, необходимые для работы модуля и управляется от микроконтроллера с помощью стандартного набора команд. Интегрированная в модуль светодиодная подсветка гарантирует высокую контрастность информации на дисплее даже в условиях плохой освещенности цеха. На дисплей выводится информационная и служебная информация, размерность индицируемых величин (кг, кг/час, сек, м³, кг/м куб.), а также результаты самодиагностики системы. Клавиатура устройства позволяет, как непосредственно управлять работой системы, так и вводить необходимые константы и параметры в диалоговом режиме с помощью разработанного интерфейса оператора. Операционная система допускает вывод на дисплей смысловой информации на

трех языках - украинском, русском и английском. Выбор нужного языка, а также настройка диалогового меню осуществляется оператором с клавиатуры.

Постоянная электрически перепрограммируемая память (EEPROM) предназначена для долговременного хранения системных констант, индивидуальных коэффициентов преобразования аналоговых модулей ввода/вывода, параметров настройки и конфигурации, а также пользовательских констант и рецептов смесей. Часть параметров вводятся в EEPROM при общей настройке комплекса, в составе которого используется система управления, и защищена от случайного изменения паролем. Пользовательские же параметры (форма отображения данных на дисплее, язык вывода смысловой информации, размерности физических величин и т.д.) могут быть изменены оператором для обеспечения максимального удобства работы с устройством. Аппаратно EEPROM реализована на микросхеме 24LC16В производства фирмы Microchip, имеющую объем 2к x 8 бит. Операции чтения/записи из/в EEPROM осуществляются микроконтроллером по стандартному 2-х проводному интерфейсу I2C.

Необходимые для работы системы временные интервалы с высокой точностью формируются интегрированными в микроконтроллер таймерами. Входная частота стабилизирована кварцевым резонатором, необходимые индивидуальные корректирующие коэффициенты записываются в EEPROM при тестировании и настройке.

Модуль связи с ЭВМ верхнего уровня и принтером представляет собой два идентичных конвертора логических уровней в стандарт интерфейса RS-232C. Аппаратно он выполнен на двух микросхемах ADM202 фирмы Analog Devices. Внутренняя структура этой микросхемы позволяет получить напряжения ± 12 В, необходимые для интерфейса RS-232C, из напряжения питания +5 В. Протокол приема/передачи данных по интерфейсу реализуется модулем асинхронного порта ввода/вывода USART, который интегрирован в микроконтроллер.

Аппаратная часть микропроцессорной системы управления собрана в универсальном 19-ти дюймовом конструктиве модели "Internorm 19", производимого фирмой BOPLA. Данный конструктив полностью соответствует международным стандартам, габаритный типовой размер и универсальная конструкция легко позволяют встраивать сконструированную в нем систему управления в стандартные промышленные стойки и шкафы. Кроме того, имеющиеся откидывающиеся ножки дают возможность использовать устройство и в автономном настольном варианте. На передней панели системы расположены 22 мембранные клавиши и ЖКИ дисплей. Клавиатура передней панели содержит специальные обозначения и обеспечивает 2-х уровневое взаимодействие пользователя с системой. Лицевая панель конструктива закрыта откидывающейся крышкой из прозрачного материала, обеспечивающей защиту устройства от воздействий пыли и имеющей замок для санкционированного доступа к клавиатуре.

Внутренняя часть конструктива "Internorm 19" разделена перегородкой на два отсека. В первом отсеке располагается сетевой фильтр, сетевой трансформатор и блок стабилизатора напряжения питания. Отсек блока питания

экранирован от остальной части прибора металлическим кожухом, соединенным с клеммой заземления. Во втором отсеке в направляющих размещена основная плата микропроцессорной системы. ЖКИ дисплей оснащен плоским шлейфом RC-16 с разъемом IDC-16 и подключен к разъему IDCC-16, расположенному на плате. С помощью такого же разъема подключена и клавиатура. На плате также расположен разъем, к которому может быть подключен дополнительный модуль расширения каналов аналогового ввода. Плата установлена таким образом, что расположенные в ее торце разъемы доступны для подключения к ним ответных частей с внешней части устройства. К разъемам DB-9M могут быть подключены принтер и ЭВМ верхнего уровня, к разъемам IDCC-30 подключаются дискретные входные сигналы. Позиционные выходные сигналы и сигналы аналогового ввода/вывода подключены к промышленным разъемам V.35.

ВЫВОДЫ

Разработана система автоматизированного управления процессом приготовления соэкструзионной смеси, обеспечивающая реализацию предложенной математической модели процесса и метода управления экструзионным оборудованием, созданы аппаратные и программные средства микропроцессорной системы интегрированного производства рециклинга полимеров. Система обеспечивает: управление технологическим процессом в реальном масштабе времени, возможность восстановления после сбоя, возможность автономной отладки и модификации

математического обеспечения, диагностику работоспособности комплекса. Испытания и опыт эксплуатации системы подтвердили правильность теоретических выводов и эффективность предложенного метода управления, возможность повышения качества соэкструзии с помощью АСУ ТП приготовления смеси, построенной на основе теоретических и практических исследований. Полностью разработана аппаратная часть системы и полная конструкторская документация на опытные образцы, а сама система рекомендована к внедрению на комплексах по переработке вторичных полимерных материалов.

ПЕРЕЧЕНЬ ССЫЛОК

1. Асеев А.А., Баранов В.П., Левенталь В.Ф. Гибридные микросборки для аналого-цифрового преобразователя сигналов тензорезисторных датчиков // Приборы и системы управления, - 1985. - №5. - С.24-25.
2. Видинеев И. Д. Автоматическое непрерывное дозирование сыпучих материалов. - М.: Энергия, 1974. - 118с.
3. Завгородний С.М. Оборудование для переработки пластмасс.- М.: Машиностроение, 1976.- 315 с.
4. Вильгельм Клеер, Система АРА (опыт Австрии)// Упаковка, №6. 2001. - С. 32-34.
5. Смиронов М.Г., Осетинский Л.Г., Многослойные пленки (возможности и преимущества) // Упаковка, №4,2002 - С. 14-15.
6. Франс Люкс, Упаковка и экология, Перев. с англ., М.: изд-во МГУП), 1999.- 220 с.
7. Переработка отходов пластмасс, Е.А. Малишкова, И.И. Потапов.-М.: Авис Оригинал, 1997.-159с.
8. Дядичев В.В. Переработка отходов полимерных материалов методом соэкструзии. - Луганск: изд-во ВНУ им. В.Даля, 2003.-220с.
9. Кагур И.И., Струтинский В.М., Новые возможности ЭВФ.// Упаковка, №3. - 2001 - С. 24-26.

УДК 678.74.002.5

РАЗРАБОТКА ЦИФРОВЫХ АЛГОРИТМОВ ДЛЯ СИСТЕМЫ С ЗАПАЗДЫВАНИЕМ ПО УПРАВЛЕНИЮ

Е.А.Еремин, Л.В.Ильина

В классе самонастраивающихся систем с явной эталонной моделью рассматривается решение задачи синтеза цифровых алгоритмов настройки для объекта с запаздыванием по управлению. Использование критерия гиперустойчивости и метода непрерывных моделей позволяет обеспечить желаемое качество управления в условиях априорной неопределенности, обеспечивая достижение поставленных целей адаптации.

In the class of self-adapting systems with obvious reference model the solution of the task of synthesis of digital algorithms of set-up for object with delay on control is considered. Usage of a hyperstability criterion and method of continuous models allows to supply desirable quality of control in conditions of prior uncertainty, providing reaching of object in views of adapting.

ВВЕДЕНИЕ

В современном мире в силу широкого применения микропроцессорной техники важную роль играет задача управления непрерывными объектами с помощью дискрет-

ных управляющих устройств. Кроме того, актуальной является задача регулирования в условиях априорной неопределенности объектами, содержащими запаздывание в канале управления. Наличие запаздывания отрицательно сказывается на работоспособности системы управления. Существуют различные подходы к решению данной проблемы [1, 2], одним из методов является компенсация запаздывания с помощью различных дополнительных устройств. В данной работе исследуется проблема компенсации в системе запаздывания по управлению с использованием адаптивного прогнозирующего устройства. Помимо основного регулятора в систему вводится вспомогательный контур, выполняющий функцию прогноза. Для построения контура дискретной настройки коэффициентов регулятора используется метод непрерывных моделей [3, 4]. Решение задачи синтеза в настоящей работе достигается в рамках критерия гиперустойчивости [5, 6].

1 ПОСТАНОВКА ЗАДАЧИ

Пусть объект управления описывается уравнением

$$\frac{dx(t)}{dt} = Ax(t) + bu(t-h) + f(t), \quad (1)$$

$$x(0) = x_0, u(s) = 0, s \in [-h, 0],$$

где $x(t) \in R^n$ - вектор состояния, $u(t) \in R$ - управляющее воздействие, h - известное постоянное запаздывание, $f(t)$ - возмущающее воздействие, удовлетворяющее условию

$$\int_0^{\infty} \|f(t)\|^2 dt < \infty. \quad (2)$$

Объект (1) функционирует в условиях априорной неопределенности

$$A = A(\zeta), B = B(\zeta), f(t) = f_{\zeta}(t), \zeta \in \Xi,$$

где ζ - набор неизвестных параметров, принадлежащих заданному множеству Ξ .

Желаемое поведение объекта управления (1) задается с помощью явной эталонной модели

$$\begin{aligned} \frac{dx_m(t)}{dt} &= A_m x_m(t) + b_m r(t), \xi(t) = x_m(t-h), \\ x_m(0) &= x_{m0}, \end{aligned} \quad (3)$$

где $x_m(t) \in R^n$, $r(t) \in R$ - скалярное задающее воздействие, которое является кусочно-непрерывной ограниченной функцией.

Для объекта управления (1) и эталонной модели (3) выполнены условия структурного согласования

$$A = A_m + b_m C_0^T, b = (1 + \mu_0) b_m, \mu_0 > 0, \quad (4)$$

причем матрицы A и A_m - гурвицевы.

Требуется синтезировать систему управления таким образом, чтобы при любых начальных условиях $x(0)$ обеспечивалось выполнение цели управления вида

$$\lim_{t \rightarrow \infty} [x(t) - \xi(t)] = 0. \quad (5)$$

2 МЕТОД РЕШЕНИЯ

Для компенсации запаздывания по управлению в систему вводится дополнительный контур

$$\frac{dy(t)}{dt} = A_m y(t) + b_m (u(t) - u(t-h)) + b_m v(t+h), \quad (6)$$

где $v(t)$ - дополнительное управляющее воздействие.

При решении задачи построения системы прямого

адаптивного управления будем следовать методике, предложенной в работе [6].

2.1 Первый этап синтеза

В рамках исследования системы на гиперустойчивость необходимо получить эквивалентное математическое описание. Для этого вводятся в рассмотрение два вектора рассогласования $e(t) = z(t) - x_m(t)$, $\varepsilon(t) = z(t-h) - x(t)$, где $z(t) = x(t) + y(t)$. Кроме того, для удобства вводятся следующие обозначения:

$$\alpha_0^T = \left[-\frac{C_0^T}{1 + \lambda \mu_0}, \frac{(1-\lambda)C_0^T}{1 + \lambda \mu_0}, -\frac{\mu_0(1-\lambda)}{1 + \lambda \mu_0}, \frac{1}{1 + \lambda \mu_0} \right], \lambda = \text{const} > 0,$$

$$\beta_0^T = [\mu_0, C_0^T], w_1^T = [z^T, y^T, u(t-h), r(t)],$$

$$w_2^T = [u(t) - u(t-h), y^T].$$

Основной и вспомогательный законы управления задаются в виде

$$u(t) = \alpha^T(t) w_1(t), v(t) = \beta^T(t) w_2(t-h),$$

$$\beta(t+h) = \beta(t) - b_m^T H e(t) P_1 w_2(t), E(t) = H e(t), \quad (7)$$

где матрица H выбирается из уравнения $HA_m + A_m^T H = -Q$;

Q - произвольная положительно определенная симметрическая матрица, $P_1 = \text{diag}\{p_{1i}\}$, $i = 1, \dots, n+1$ - произвольная положительно определенная диагональная матрица. Тогда эквивалентное математическое описание для системы (1), (3), (6) можно записать в виде

$$\frac{de(t)}{dt} = A_m e(t) + q_1 + f(t), q_1 = -q_2,$$

$$\begin{aligned} q_2 &= -b_m [(1 + \lambda \mu_0)(\alpha^T(t) - \alpha_0^T) w_1(t) + \\ &+ (\beta^T(t) - b_m^T E(t) P_1 w_2(t) - \lambda \beta_0^T) w_2(t)]; \end{aligned} \quad (8)$$

$$\frac{d\varepsilon(t)}{d(t)} = A \varepsilon(t) + b_m q_3 + f(t-h) - f(t), q_3 = -q_4, \quad (9)$$

$$q_4 = -(\beta^T(t) - \beta_0^T) w_2(t-h).$$

2.2 Второй этап синтеза

На данном этапе основной задачей синтеза является обеспечение условий вещественности и строгой положительности линейной стационарной части системы (1), (3), (6).

Частотные неравенства

$$\text{Re}[H(j\omega I_m - A_m)^{-1}] > 0, \quad (10)$$

$$\text{Re}[g^T (j\omega I_n - A)^{-1} b_m] > 0, \quad \forall \omega \geq 0$$

всегда будут иметь место, т.к. матрицы A и A_m гурвицевы, а матрица H и вектор g выбираются по определенным правилам.

2.3 Третий этап синтеза

Следуя критерию гиперустойчивости, необходимо проверить выполнение интегрального неравенства В.М.Попова (ИНП)

$$\eta(0, t) = \int_0^t E^T(s) q_2 ds + \int_0^t \varepsilon^T(s) g q_4 ds > -\gamma^2, \quad \forall t \geq 0. \quad (11)$$

Представим ИНП в виде суммы трех интегралов

$$\eta(0, t) = \sum_{i=1}^3 \eta_i(0, t)$$

и рассмотрим следующие интегральные слагаемые

$$\eta_1(0, t) = - \int_0^t E^T(s) b_m (1 + \lambda \mu_0) \{ \alpha(t) - \alpha_0 \}^T w_1(s) ds,$$

$$\eta_2(0, t) = - \int_0^t E^T(s) b_m (1 + \lambda \mu_0) \times \\ \times \{ \beta(t) - \lambda \beta_0 - b_m^T E(s) P_1 w_2(s) \}^T w_2(s) ds,$$

$$\eta_3(0, t) = - \int_0^t \varepsilon^T(s) g \{ \beta(t) - \beta_0 \}^T w_2(s-h) ds,$$

которые можно переписать в виде

$$\eta_1(0, t) = 0,5 \sum_{i=1}^{2n+2} \int_0^t (E^T(s) b_m + \delta) (1 + \lambda \mu_0) \{ \alpha_i(t) - \alpha_{0i} \} w_{1i}(s) ds +$$

$$+ 0,5 \sum_{i=1}^{2n+2} \int_0^t (E^T(s) b_m - \delta) (1 + \lambda \mu_0) \{ \alpha_i(t) - \alpha_{0i} \} w_{1i}(s) ds,$$

$$\eta_2(0, t) = 0,5 \sum_{j=1}^{n+1} \int_0^t (E^T(s) b_m + \delta) (1 + \lambda \mu_0) \{ \beta_j(t) - \lambda \beta_{0j} -$$

$$- b_m^T E(s) P_{1j} w_{2j}(s) \} w_{2j}(s) ds +$$

$$+ 0,5 \sum_{j=1}^{n+1} \int_0^t (E^T(s) b_m - \delta) (1 + \lambda \mu_0) \{ \beta_j(t) - \lambda \beta_{0j} -$$

$$- b_m^T E(s) P_{1j} w_{2j}(s) \} w_{2j}(s) ds,$$

$$\eta_3(0, t) = 0,5 \sum_{j=1}^{n+1} \int_0^t (\varepsilon^T(s) g + \Delta) \{ \beta_j(t) - \beta_{0j} \} w_{2j}(s-h) ds +$$

$$+ 0,5 \sum_{j=1}^{n+1} \int_0^t (\varepsilon^T(s) g - \Delta) \{ \beta_j(t) - \beta_{0j} \} w_{2j}(s-h) ds.$$

Если коэффициенты самонастройки основного и вспомогательного законов управления синтезировать в виде

$$\frac{d\alpha(t)}{dt} = -b_m^T E(t) P_2 w_1(t),$$

$$\frac{d\beta(t)}{dt} = -P_3 [b_m^T E(t) w_2(t) + g^T \varepsilon(t) w_2(t-h)], \quad (12)$$

где $P_2 = \text{diag}\{p_{2i}\}$, $P_3 = \text{diag}\{p_{3j}\}$, $i = 1, \dots, 2n+2$, $j = 1, \dots, n+1$ - положительно определенные диагональные матрицы, то для $\eta(0, t)$ будет справедлива искомая оценка

$$\eta(0, t) > -\gamma^2 = \text{const}, \quad \forall t > 0.$$

Дискретизация алгоритмов настройки. Решение задачи синтеза цифровых алгоритмов настройки параметров регуляторов (6) получим с использованием метода непрерывных моделей.

Требуется определить дискретные алгоритмы адаптации $\alpha(t)$, $\beta(t)$ так, чтобы система (1), (3), (7) была адаптивной в заданном классе Ξ . Осуществив переход от непрерывных алгоритмов к цифровым, получим следующее математическое описание:

$$\alpha_i(t_{k+1}) = \alpha_i(t_k) + \gamma_t (b_m^T E(t_k) p_{2i} w_{1i}(t_k)), \quad (13)$$

$$i = 1, \dots, 2n+2,$$

$$\beta_j(t_{k+1}) = \beta_j(t_k) + \gamma_t (p_{3j} b_m^T E(t_k) w_{2j}(t_k) + g^T \varepsilon(t_k) w_{2j}(t_k - h)), \quad j = 1, \dots, n+1, \quad (14)$$

где γ_t - шаг алгоритма настройки, k - номер шага, t_k - дискретный аналог времени.

2.4 Четвертый этап синтеза

Справедливость неравенств (10), (11), рассматриваемых относительно эквивалентной системы управления (8), (9), означает гиперустойчивость системы (1)-(4), (6), (7) и выполнение цели управления (5) и целей адаптации

$$\lim_{t \rightarrow \infty} \alpha(t) = \alpha_0, \quad \lim_{t \rightarrow \infty} \beta(t) = \beta_0.$$

При затухающих возмущениях система сохраняет свою работоспособность. Из выполнения соотношения (2) следует выполнение предельного соотношения

$$\lim_{t \rightarrow \infty} f(t) = 0,$$

поэтому выполнение целевого условия (5) не будет нарушено, так как возмущение является затухающим во времени.

ИЛЛЮСТРАТИВНЫЙ ПРИМЕР

Пусть объект управления (1) имеет следующую структуру матриц

$$A = \begin{bmatrix} 0 & 1 \\ -a_1 & -a_2 \end{bmatrix}, b = \begin{bmatrix} 0 \\ b_1 \end{bmatrix}, x(0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, f(t) = \begin{bmatrix} 0 \\ 0,21e^{-t} \end{bmatrix},$$

$$A_m = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}, b_m = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Параметры матрицы A и вектора b - неизвестные числа, находящиеся в следующих диапазонах $0,1 \leq a_i \leq 4$, $i=1, 2$, $0,1 \leq b_i \leq 2,5$. Постоянное запаздывание $h = 0,3$

известно. Имитационное моделирование осуществлялось при следующем задающем воздействии:

$$r(t) = 2 \sin(0,63t).$$

Выбор матрицы H и вектора g был осуществлен согласно неравенствам (10)

$$H = \begin{bmatrix} 1,25 & 0,25 \\ 0,25 & 0,25 \end{bmatrix}, g = \begin{bmatrix} 3 \\ 2 \end{bmatrix}.$$

Поведение выхода объекта управления (1) - $x(t)$ и выхода эталонной модели (3) - $\xi(t)$ продемонстрировано на рис. 1. Характер изменения сигнала рассогласования $x(t) - \xi(t)$ изображен на рис. 2.

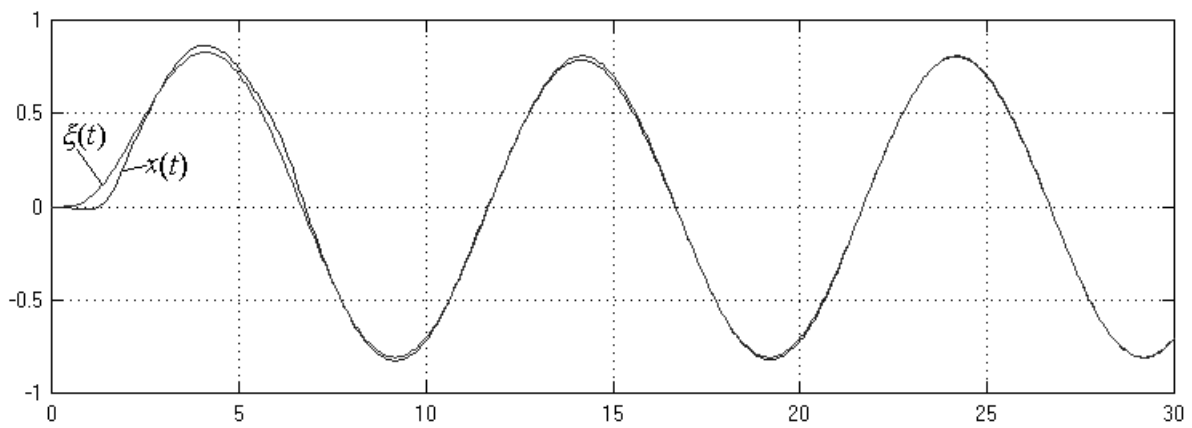


Рисунок 1 - Динамика выходов объекта управления и эталонной модели

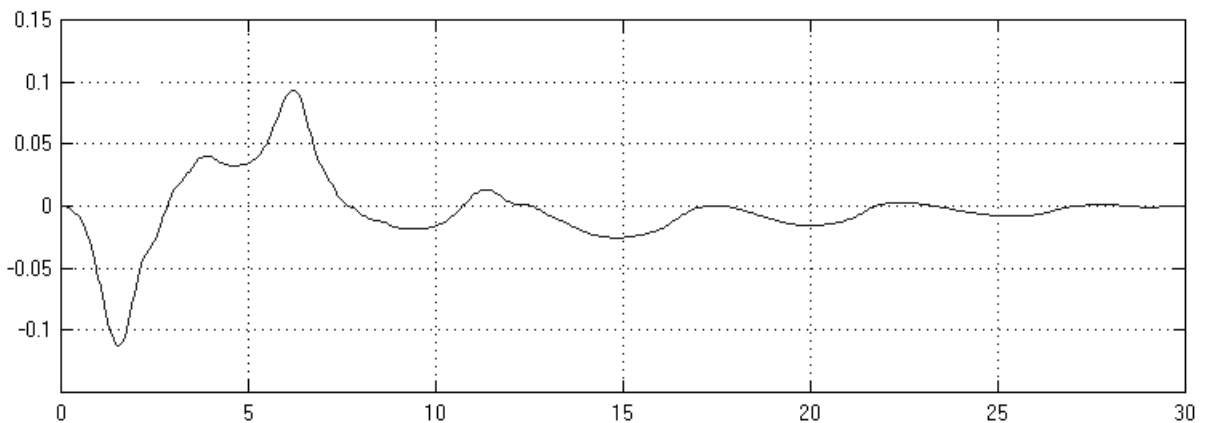
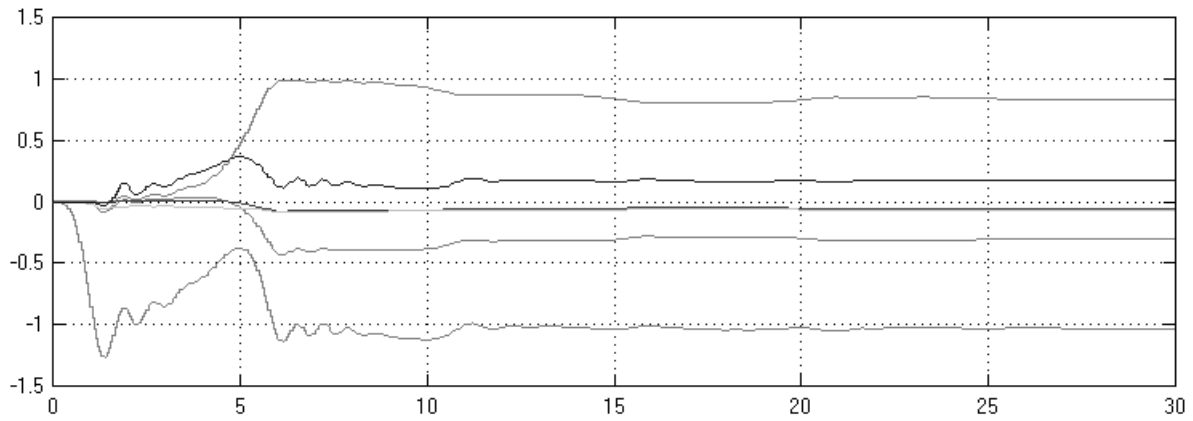
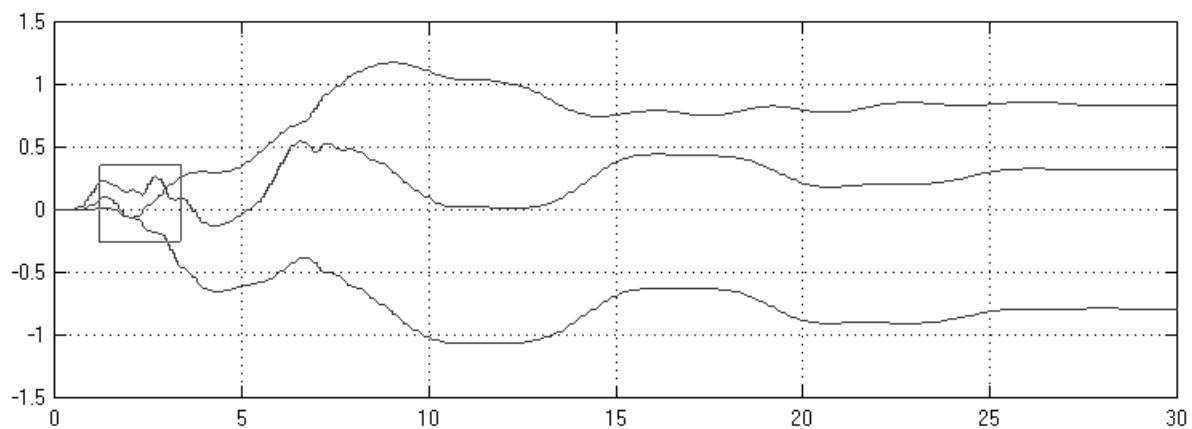


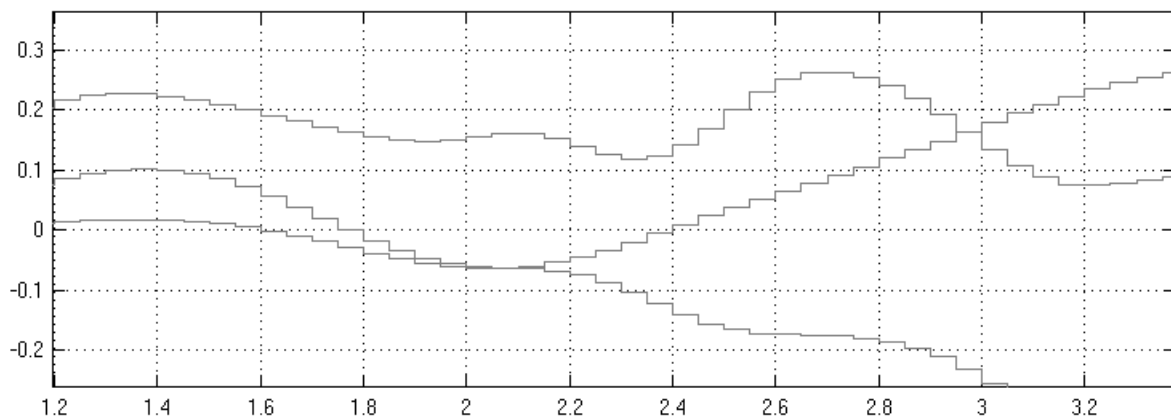
Рисунок 2 - Сигнал рассогласования $x(t) - \xi(t)$



a)



b)



c)

Рисунок 3 - Цифровые процессы настройки параметров регуляторов

Коэффициенты настройки $\alpha(t_k)$, $\beta(t_k)$ законов управления (7), описываемые уравнением (13) при $\gamma_t = 0,05$, представлены на рис. 3а и 3б, для наглядности цифрового процесса настройки коэффициентов фрагмент одного из них приведен на рис.3с.

ЗАКЛЮЧЕНИЕ

Полученные результаты показывают, что использование прогнозирующего устройства и цифровой настройки коэффициентов регуляторов позволяет обеспечить достаточно хорошее качество управления априорно неопределенным объектом, который содержит запаздывание в канале управления.

ПЕРЕЧЕНЬ ССЫЛОК

1. Цыкунов А.М. Адаптивное управление с компенсацией влияния запаздывания в управляющем воздействии// Известия Российской академии наук. Теория и системы управления. -№4. -2000. -С.78-81.
2. Мееров М.В. Адаптивные компенсирующие регуляторы с предиктором Смита//Автоматика и телемеханика. -2000 №10. С.125-135.
3. Деревницкий Д.П., Фрадков А.Л. Прикладная теория дискретных адаптивных систем. М.: Наука, 1981. 216с.
4. Цыкунов А.М. Адаптивное управление объектами с последствием. М.: Наука, 1984. 241с.
5. Еремин Е.Л. Гиперустойчивость систем управления нелинейным объектом с запаздыванием//Автоматизация технологических процессов. -Фрунзе: Фрунз. политехн. ин-т, 1987.
6. Еремин Е.Л., Цыкунов А.М. Синтез адаптивных систем управления на основе критерия гиперустойчивости. - Бишкек: Илим, 1992. 315с.

УДК 629.7:62.50

СИНТЕЗ И АНАЛИЗ АНАЛОГО-ЦИФРОВЫХ АДАПТИВНЫХ НАБЛЮДАТЕЛЕЙ МНОГОФАЗНЫХ НЕОПРЕДЕЛЕННЫХ СИГНАЛОВ

Е.Е.Потапенко

В статті синтезовано аналого-дискретні адаптивні однокаскадні й багатокаскадні спостерігачі, які одночасно оцінюють проєкції на нерухомий базис вектора (струму або напруги) та швидкість плоского обертання цього вектора на фоні вад, які обумовлено широтно-імпульсною модуляцією. Спостерігачі взагалі не викривляють фази корисних складових вихідних сигналів перетворювача частоти, що є найбільш важливим, і ледь викривляють їх амплітуди тільки у швидкопливаючих процесах.

В статье синтезированы аналого-дискретные адаптивные однокаскадный и многокаскадный наблюдатели, одновременно оценивающие проекции на неподвижный базис вектора (тока или напряжения) и скорость плоского вращения этого вектора на фоне помех, обусловленных широтно-импульсной модуляцией. Наблюдатели совершенно не искажают фазы полезных составляющих выходных сигналов преобразователя частоты, что наиболее важно, и слегка искажают их амплитуды только в быстротечных переходных процессах.

Analogue-discrete adaptive single and many cascade observers were synthesized in the article which simultaneously estimate of vector (current or voltage) projections on moveless basis and flat rotation speed of this vector on noise background which are caused by pulse width modulation. Observers don't deform absolutely the useful frequency converter output signal phases that is more important and a little bit deform their amplitudes just in the fast transient processes.

ВВЕДЕНИЕ

Современные электроприводы переменного тока в большинстве случаев содержат преобразователи частоты (ПЧ) с широтно-импульсной модуляцией (ШИМ). Выходные сигналы ПЧ с ШИМ можно представить как суммы полезных сигналов и высокочастотных помех, причем, амплитуды помех могут в десятки раз превышать амплитуды полезных сигналов. Кроме того, помехи представляют собой прямоугольные импульсы малой длительности. Такая форма сигналов не позволяет путем числовой обработки достаточно точно выделить полезные сигналы. Во многих случаях частоты полезных сигналов неизвестны. Постановка обычных аналоговых фильтров приводит к недопустимому искажению полезных сигналов. Насколько известно автору, в настоящее время

отсутствуют фильтры, выделяющие полезные сигналы без искажений в описанных условиях. Это не позволяет создать обратные связи, охватывающие непосредственно ПЧ для компенсации его погрешностей.

Для оценки полезных (главных) гармоник можно было бы попытаться использовать (см. работу [1] и библиографию к ней) ряд методов, таких как рекуррентный метод наименьших квадратов, расширенный фильтр Калмана, дискретное преобразование Фурье, метод Проуни и др. Однако,

1) эти методы связаны с большим объемом вычислений, что требует применения цифровой обработки сигналов,

2) для дискретной фильтрации выходных сигналов ШИМ требуется очень маленький такт счета, что затрудняет применение только цифровой обработки сигналов,

3) для этих методов в случае с ШИМ слишком велико отношение "шум/полезный сигнал".

В работах [2, 3] с участием автора предложены простые адаптивные наблюдатели, содержащие непрерывные инерционные звенья первого порядка (ИЗ1) и цифровой компенсатор искажений полезных сигналов. Тот факт, что в цифровую часть сигналы поступают не непосредственно с выходов ПЧ с ШИМ, а с выходов ИЗ1, позволяет существенно увеличить такт счета и, тем самым, устранить указанные выше недостатки 1) и 2). Помимо недостатка 3) наблюдатели работ [2,3] требуют по 2 преобразователя "код-напряжение", что усложняет аппаратную часть фильтров.

Целью данной статьи является полное устранение перечисленных недостатков.

1 СИНТЕЗ АНАЛОГО-ДИСКРЕТНЫХ АДАПТИВНЫХ НАБЛЮДАТЕЛЕЙ

Для системы

$$\dot{x}_1 = -\omega x_2, \quad \dot{x}_2 = \omega x_1, \quad \omega = (x_1 \dot{x}_2 - x_2 \dot{x}_1) |x|^{-2}, \quad (1)$$

описывающей поведение проекций x_1, x_2 вектора x ,

совершающего плоское вращение со скоростью ω , и измерений

$$x_{1m} = x_1 + \vartheta_1, \quad x_{2m} = x_2 + \vartheta_2, \quad (2)$$

где ϑ_1 и ϑ_2 представляют собой высокочастотные по сравнению с ω гармонические помехи, в работах [2, 3] синтезирован наблюдатель

$$\dot{\hat{x}}_1 = -\hat{\omega}\hat{x}_2 + l_{11}(\hat{x}_1 - x_{1m}) + l_{12}(\hat{x}_2 - x_{2m}), \quad (3)$$

$$\dot{\hat{x}}_2 = \hat{\omega}\hat{x}_1 + l_{21}(\hat{x}_1 - x_{1m}) + l_{22}(\hat{x}_2 - x_{2m}), \quad (4)$$

где $l_{ij}(i, j = 1, 2)$ - коэффициенты усиления наблюдателя, а индекс “^” означает оценку соответствующей переменной. Для определения $\hat{\omega}$ умножим (3) на \hat{x}_1 и вычтем из полученного уравнение (4), умноженное на \hat{x}_2 . В результате получим

$$\hat{\omega} = \frac{\hat{x}_1 \dot{\hat{x}}_2 - \hat{x}_2 \dot{\hat{x}}_1 + (l_{21}\hat{x}_1 - l_{11}\hat{x}_2)(\hat{x}_1 - x_{1m}) + (l_{22}\hat{x}_1 - l_{12}\hat{x}_2)(\hat{x}_2 - x_{2m})}{\hat{x}_1^2 + \hat{x}_2^2}. \quad (5)$$

Так как наблюдатель (3), (4) дает соотношения $\hat{x}_1 \approx x_1$, $\hat{x}_2 \approx x_2$, то на основании (2) выражение (5) можно переписать в виде

$$\hat{\omega} = \frac{\hat{x}_1 \dot{\hat{x}}_2 - \hat{x}_2 \dot{\hat{x}}_1 - (l_{21}\hat{x}_1 - l_{11}\hat{x}_2)\vartheta_1 - (l_{22}\hat{x}_1 - l_{12}\hat{x}_2)\vartheta_2}{\hat{x}_1^2 + \hat{x}_2^2}. \quad (6)$$

Для повышения точности оценки ω выход выражения (6) надо пропустить через низкочастотный фильтр, уменьшающий влияние высокочастотных помех ϑ_1 и ϑ_2 . В результате при наличии указанного фильтра вместо (5), (6) можно полагать

$$\hat{\omega} = \frac{\hat{x}_1 \dot{\hat{x}}_2 - \hat{x}_2 \dot{\hat{x}}_1}{\hat{x}_1^2 + \hat{x}_2^2}. \quad (7)$$

При $\hat{x}_1 \approx x_1$, $\hat{x}_2 \approx x_2$ будем иметь $\hat{\omega} \approx \omega$.

Пусть $l_{12} = \hat{\omega} = -l_{21}$, $l_{11} = l_{22} = l_1$. Тогда наблюдатель примет вид

$$\dot{\hat{x}}_1 = l_1(\hat{x}_1 - x_{1m}) - \hat{\omega}x_{2m},$$

$$\dot{\hat{x}}_2 = l_1(\hat{x}_2 - x_{2m}) + (\hat{\omega}x_{1m}). \quad (8)$$

Структурную схему системы (7), (8) можно представить в виде рис. 1.

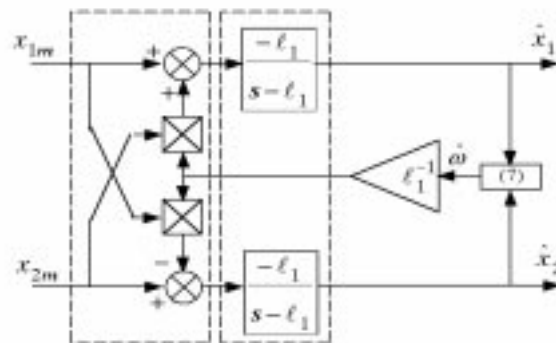


Рисунок 1

Поменяв местами в линейной схеме на рис. 1 выделенные пунктиром звенья, придем к схеме на рис. 2.

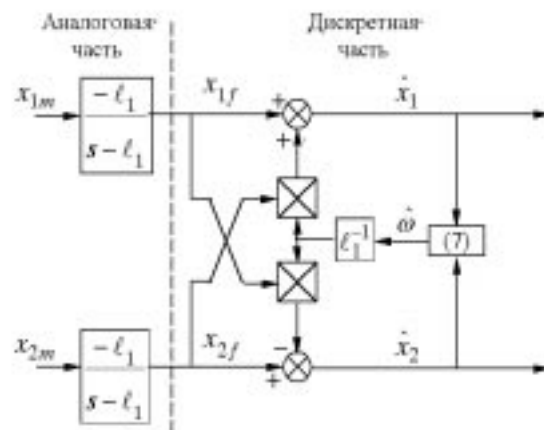


Рисунок 2

В схеме на рисунке 2 аналоговые инерционные звенья устраняют высокочастотные помехи, а дискретная часть устраняет сдвиги по фазе и искажения амплитуды полезных составляющих сигналов, возникающие при прохождении сигналов через инерционные звенья. Устранение высокочастотных помех позволяет существенно увеличить такт счета микропроцессора и, тем самым, удешевить его. Следует отметить, что фильтр на рис. 1 более помехоустойчив, чем фильтр на рис.2, так как на рис.1 сигнал $\hat{\omega}$, получаемый путем дифференцирования сигналов \hat{x}_1, \hat{x}_2 , пропускается через инерционные звенья, а на рис. 2 - не пропускается. Поэтому при дискретном исполнении фильтра лучше использовать схему рис.1.

При необходимости улучшения фильтрующих свойств фильтр можно сделать состоящим из двух и более каскадов. Двухкаскадный фильтр представлен на рис.3.

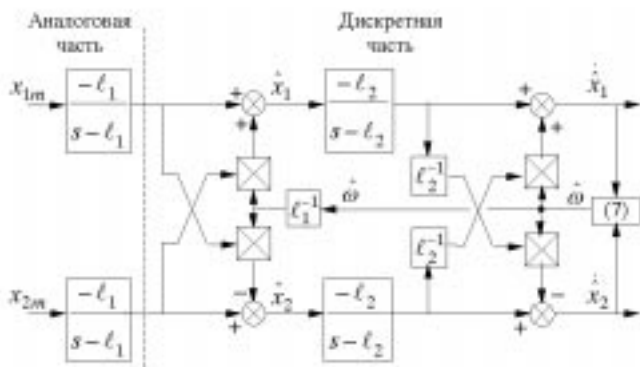


Рисунок 3

Для улучшения фильтрующих свойств и понижения такта счета структурную схему можно преобразовать к виду рис. 4. На основании рис. 2 и 4 можно построить многокаскадный фильтр.

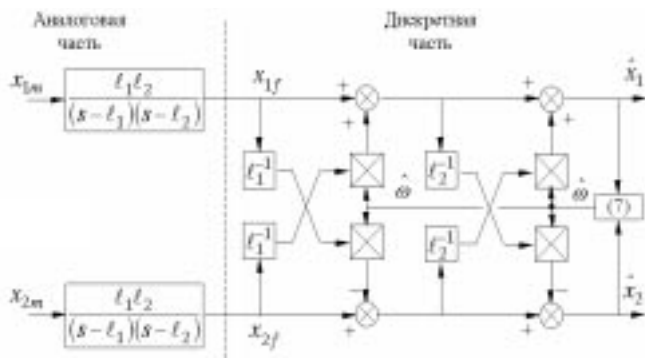


Рисунок 4

2 РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

На рис. 5 представлены графики x_{2m} , x_2 , \hat{x}_2 в установившемся режиме. Как видно из рис. 5, имеет место точная оценка основной гармоники сигнала. Во всех последующих результатах моделирования помеха полагалась с теми же параметрами (амплитуда 100, частота 10^4 c^{-1}) и зашумленные сигналы не показываются. На рис. 6 продемонстрирован переходный процесс оценки неизвестной частоты $\omega = 300 \text{ c}^{-1}$ с помощью адаптивного наблюдателя, демонстрирующий его высокое быстродействие и точность.

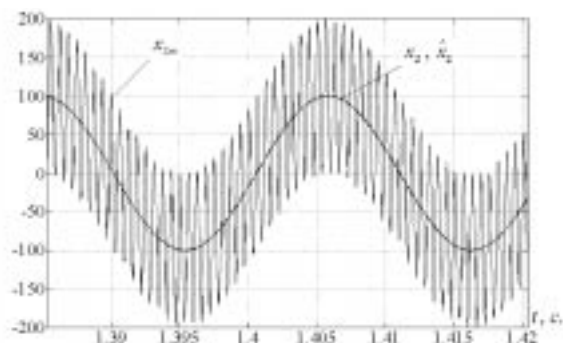


Рисунок 5

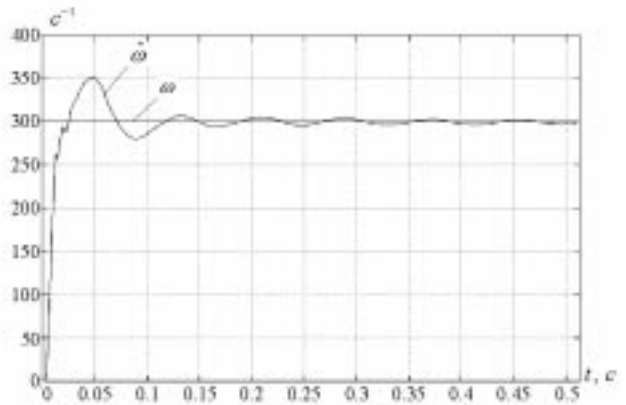


Рисунок 6

На рис. 7, 8 представлены результаты моделирования работы фильтра при изменении частоты от 0 c^{-1} до 300 c^{-1} и увеличении модуля x от 0 до 100 за $0,5 \text{ c}$.

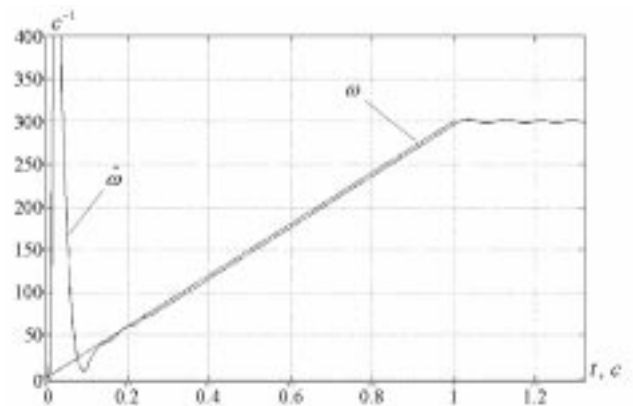


Рисунок 7

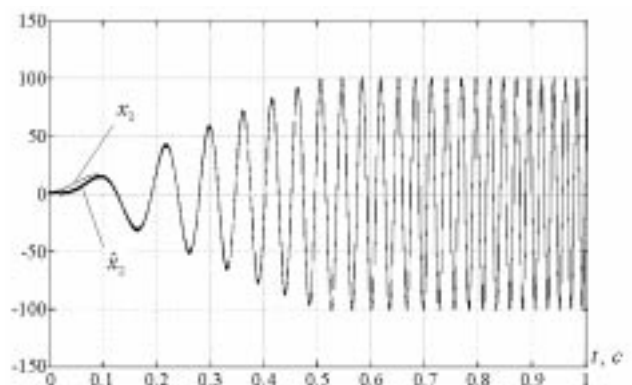


Рисунок 8

Как следует из рис. 7, 8, имеет место точная оценка неизвестных искоемых переменных. На рис. 9, 10 изображены те же искоемые переменные, но при изменении частоты от 0 c^{-1} до 1 c^{-1} (1 оборот за 6 c). Рис. 9 свидетельствует о точной оценке проекций вектора.

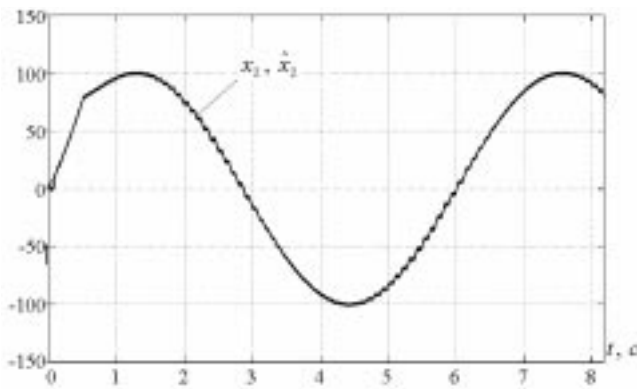


Рисунок 9

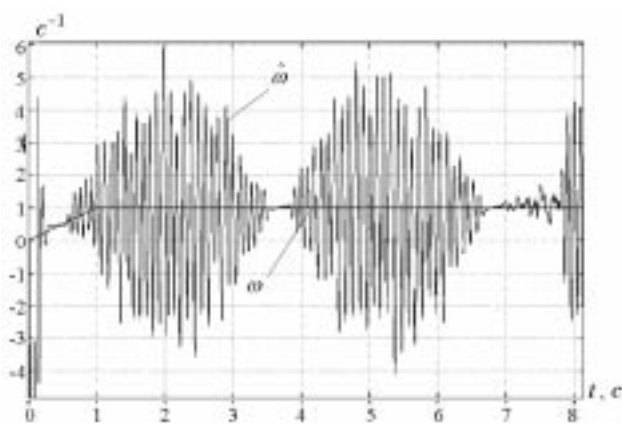


Рисунок 10

Оценка же частоты неудовлетворительна. Это свидетельствует о необходимости включения в фильтр второго каскада, как это сделано на рис. 4. Как показывают результаты моделирования, двухкаскадный фильтр полностью устраняет пульсации в оценке скорости.

ЗАКЛЮЧЕНИЕ

В соответствии с целью статьи синтезированы аналого-дискретные адаптивные однокаскадный (рис. 2) и многокаскадный (рис. 4) наблюдатели, одновременно оценивающие проекции на неподвижный базис вектора (тока или напряжения) и скорость плоского вращения этого вектора на фоне больших и "неудобных" помех, обусловленных ШИМ. Наблюдатели совершенно не искажают фазы полезных составляющих выходных сигналов ПЧ, что наиболее важно, и слегка искажают их амплитуды только в быстротечных переходных процессах. Один каскад фильтра требует лишь 6 операций сложения/вычитания и 7 операций умножения/деления. Каждый последующий каскад нуждается лишь в двух операциях сложения/вычитания и четырех операциях умножения/деления. Таким образом, синтезированные фильтры устраняют все три проблемы потенциально возможных аналогов, приведенные во введении. В дальнейшем предполагается исследовать контуры регулирования тока и напряжения с синтезированными фильтрами при векторном управлении двигателями переменного тока.

ПЕРЕЧЕНЬ ССЫЛОК

1. Rosolovski E., Staszewski J. Recursive Method for Power System Frequency Measurement // Технічна електродинаміка. Тематичний випуск. - 2002. - Часть. 6. - С. 114-117.
2. Потапенко Е.Е. Определение скорости плоского вращения вектора, заданного его проекциями // Радиозлектроника. Информатика. Управление. - 2002. - №1. - С. 143-146.
3. Потапенко Е.М., Потапенко Е.Е. Фильтрация многофазных нестационарных неопределенных гармонических сигналов // Электротехника // 2003. - №3. - С.55-57.

УДК 681. 514. 6

СУММАРНЫЕ ФУНКЦИОНАЛЫ КАЧЕСТВА ИТЕРАЦИОННЫХ ДВУХКАНАЛЬНЫХ СИСТЕМ УПРАВЛЕНИЯ С ЭТАЛОННОЙ НАСТРОЙКОЙ И ПРИ СЛУЧАЙНЫХ ВОЗДЕЙСТВИЯХ

А.А.Худяев

Розглядається побудова сумарних функціоналів якості для оцінки ефективності та аналізу процесів самонастроювання ітераційних двоканальних відтворюючих систем із заданими еталонними операторами каналів керування і при випадкових входніх впливах.

Рассматривается получение суммарных функционалов качества для оценки эффективности и анализа процессов самонастройки итерационных двухканальных воспроизводя-

щих систем с заданными эталонными операторами каналов управления и при случайных входных воздействиях.

The determination of the quality functional equations for efficiency estimation and analysis of self-adjusting processes of the iterative type two-channel reproduction control systems with basic standard operators of channels and by random input actions is considered.

ВВЕДЕНИЕ

В настоящее время ко многим системам автоматического управления, в том числе к воспроизводящим системам, предъявляются все более высокие требования в отношении полосы пропускаемых частот, порядка астатизма, установившейся динамической точности, максимального ускорения и других динамических показателей.

Воспроизводящие системы, обладающие высокой динамической точностью, необходимы для управления современными быстропротекающими производственными процессами, а также для создания разного рода информационно-измерительных и контролирующих устройств. Известно, что возможности системы в отношении динамической точности, в конечном счете, определяются предельными значениями координат и производных координат, характеризующих поведение входящих в систему элементов. В первую очередь, важны ограничения, наложенные на скорости и ускорения, а иногда, и на рывки привода системы. В одноканальных САУ эти ограничения часто не позволяют увеличивать динамическую точность системы и полосу пропускаемых ею частот до требуемых значений.

Для повышения точности воспроизведения полезного сигнала могут быть использованы многоканальные системы, работающие по замкнутому и разомкнутому циклам [1]. Последующее развитие таких систем привело к созданию многоканальных систем, построенных по принципу "грубого" и "точного" управления и получивших название итерационных многоканальных систем [2]. В классе итерационных многоканальных систем воспроизведение задающих воздействий осуществляется последовательными приближениями (итерациями), реализуемыми соответствующими каналами управления [2,3]. Это позволяет при стационарных воздействиях потенциально обеспечить максимально высокую заданную точность работы всей многоканальной системы со стандартной (эталонной) настройкой отдельных каналов и в большинстве случаев получать качество воспроизведения, недостижимое в одноканальных системах управления [4-6].

Преимущественное применение в промышленных автоматизированных электромеханических системах [7] получили системы управления унифицированными электроприводами с эталонной (или типовой) настройкой контуров управления и с возможностью перенастройки основных параметров регуляторов. Для синтеза таких систем используют метод эталонных операторов, развитый применительно к итерационным многоканальным САУ в работах [8, 10].

Особые требования предъявляются к качеству высокочастотных итерационных воспроизводящих систем, работающих при случайных внешних воздействиях. При этом с точки зрения технической реализации наиболее просты двухканальные структуры итерационных САУ [5-7]. Однако, с изменением характеристик внешних воздействий (например, с изменением соотношения интенсивностей помех: $\varphi(t)$ - в точном и $f(t)$ - в грубом каналах) с точки зрения достижения наиболее высокой точности воспроизведения системы желательно изменять соответствующим образом динамические свойства второго,

точного (или *компенсирующего*) канала W_2^* , для чего нужно изменять (перенастраивать) его параметры.

Для формализации процессов самонастройки в работе [9] введен показатель качества итерационной двухканальной системы, характеризующий эффективность использования второго, точного, канала в виде *выигрыша в точности G*. В работе [10] в рамках задачи синтеза двухканальной системы управления с самонастройкой точного канала определены необходимые функционалы качества и проанализировано влияние полосы пропускания второго (компенсирующего) канала и характеристик отдельных случайных внешних воздействий на точность итерационной двухканальной воспроизводящей системы с эталонной настройкой.

Используя результаты, полученные в [9,10], определим суммарные функционалы качества, позволяющие анализировать влияние *формирующего параметра* r_2 точного канала и характеристик случайных входных сигналов $x(t)$, $f(t)$, $\varphi(t)$ на эффективность и процессы самонастройки итерационных двухканальных систем с заданными эталонными операторами каналов.

1 ПОКАЗАТЕЛИ ЭФФЕКТИВНОСТИ ТОЧНОГО КАНАЛА

Пусть внешние воздействия $x(t)$ и $f(t)$, $\varphi(t)$ - полезный сигнал и две помехи, итерационной двухканальной воспроизводящей системы управления представляют некоррелированные между собой центрированные стационарные случайные процессы, заданные типовыми спектральными плотностями $S_x(\omega)$ и $S_f(\omega)$, $S_\varphi(\omega)$ вида [2,8]

$$S_x(\omega) = \frac{S_x(0)}{\omega^2(T_x^2\omega^2 + 1)}; \quad (1)$$

$$S_f(\omega) = \frac{S_f(0)}{T_f^2\omega^2 + 1}, \quad S_\varphi(\omega) = \frac{S_\varphi(0)}{T_\varphi^2\omega^2 + 1}, \quad (2)$$

где $S_x(0)$, $S_f(0)$, $S_\varphi(0) = \rho^2 S_f(0)$ и T_x , T_f , T_φ - начальные (при $\omega = 0$) значения спектральных плотностей и постоянные времени корреляции случайных процессов $\dot{x}(t)$, $f(t)$, $\varphi(t)$; $\dot{x}(t)$ - стационарный случайный процесс, соответствующий первой производной изменения полезного сигнала; ρ^2 - заданный коэффициент соотношения интенсивностей помех в точном $S_\varphi(0)$ и в грубом $S_f(0)$ каналах; $\omega = 2\pi\Omega$ - угловая частота (скорость), рад/с.

Для оценки эффективности второго, точного, канала W_2^* в работе [9] введен выигрыш в точности G , равный отношению установившихся значений дисперсий ошибок

первого, грубого (или *основного*) канала $\bar{\delta}^2$ и двух-канальной системы в целом $\bar{\varepsilon}^2$

$$G = \frac{\bar{\delta}^2}{\varepsilon^2} = (1 + \mu) \frac{1}{\mu g_x^{-1} + g_f^{-1} + \lambda g_\phi^{-1}} = G_0 \left(\frac{1}{1 - \Delta G} \right), \quad (3)$$

где

$$G_0 = (1 + \mu_0) \frac{1}{\mu_0 g_{0x}^{-1} + g_{0f}^{-1} + \lambda_0 g_{0\phi}^{-1}}; \quad (4)$$

$$\mu = \frac{\bar{\delta}_x^2}{\bar{\delta}_f^2} = \mu_0 (1 - \Delta \mu), \quad \lambda = \frac{\bar{\delta}_\phi^2}{\bar{\delta}_f^2} = \lambda_0 (1 - \Delta \lambda); \quad (5)$$

$$g_x = \frac{\bar{\delta}_x^2}{\varepsilon_x^2} = g_{0x} \left(\frac{1}{1 - \Delta g_x} \right), \quad (6)$$

$$g_f = \frac{\bar{\delta}_f^2}{\varepsilon_f^2} = g_{0f} \left(\frac{1}{1 - \Delta g_f} \right), \quad g_\phi = \frac{\bar{\delta}_\phi^2}{\varepsilon_\phi^2} = g_{0\phi} \left(\frac{1}{1 - \Delta g_\phi} \right). \quad (7)$$

Здесь обозначены: $\bar{\delta}_x^2$, $\bar{\delta}_f^2$, $\bar{\delta}_\phi^2$ - дисперсии ошибок основного, грубого, канала W_1^* при воздействии на него соответственно полезного сигнала $x(t)$ или помехи $f(t)$, приведенной ко входу этого канала, или помехи $\phi(t)$, приведенной ко входу дополнительного, точного, канала W_2^* ; ε_x^2 , ε_f^2 и ε_ϕ^2 - составляющие дисперсии ошибки итерационной системы, обусловленные сигналами $x(t)$, $f(t)$ и $\phi(t)$ соответственно.

Показатели g_x и g_f характеризуют эффективность использования точного канала для компенсации ошибок по полезному сигналу $x(t)$ и помехе $f(t)$; соответственно g_ϕ определяет нормированную дополнительную ошибку, вносимую помехой $\phi(t)$. Величины μ , λ характеризуют относительную точность первого, грубого, канала. В (3)-(7) входят G_0 , μ_0 , λ_0 , g_{0x} , g_{0f} , $g_{0\phi}$ - первые приближения к величинам G , μ , λ , g_x , g_f , g_ϕ соответственно; ΔG , $\Delta \mu$, $\Delta \lambda$, Δg_x , Δg_f , Δg_ϕ - поправки, учет которых позволяет при необходимости определять точные значения. Из (3), (5)-(7) видно, что значениями поправок можно пренебречь, если их модули достаточно малы по сравнению с единицей.

Нормированное уравнение (3), выраженное в безразмерных величинах, удобно для анализа влияния характеристик полезного сигнала и помех на показатель эффективности G и для оптимизации параметров второго, точного, канала при стационарном первом.

2 ПЕРВЫЙ (ИЗМЕРИТЕЛЬНЫЙ) ВАРИАНТ ДВУХКАНАЛЬНОЙ СИСТЕМЫ

Рассмотрим случай, когда эталонные операторы первого и второго каналов различны и соответствуют колеба-

тельному и апериодическому звеньям. Тогда реальные операторы каналов $W_1^*(p)$, $W_2^*(p)$, где $p \triangleq \frac{d}{dt}$ - оператор дифференцирования по времени, и спектральные плотности внешних воздействий (1), (2) могут быть представлены в виде [10]

$$W_1^*(p) = \frac{1}{r_1 p^2 + 2\xi_1 r_1 p + 1}, \quad W_2^*(p) = \frac{z}{r_1 p + z}; \quad (8)$$

$$S_x(\omega) = \frac{S_x(0)}{\omega^2 (\gamma_x^2 r_1^2 \omega^2 + 1)};$$

$$S_f(\omega) = \frac{S_f(0)}{\gamma_f^2 r_1^2 \omega^2 + 1}, \quad S_\phi(\omega) = \frac{S_\phi(0)}{\gamma_\phi^2 r_1^2 \omega^2 + 1}, \quad (9)$$

где

$$z = \frac{r_1}{r_2} = \frac{\Omega_2}{\Omega_1} = \frac{\omega_2}{\omega_1}, \quad \xi_1 = \xi_\omega; \quad (10)$$

$$\gamma_x = \frac{T_x}{r_1} = \frac{\Omega_1}{\Omega_x} = \frac{\omega_1}{\omega_x}, \quad \Omega_x = T_x^{-1}; \quad (11)$$

$$\gamma_f = \frac{T_f}{r_1} = \frac{\Omega_1}{\Omega_f} = \frac{\omega_1}{\omega_f}, \quad \Omega_f = T_f^{-1}; \quad (12)$$

$$\gamma_\phi = \frac{T_\phi}{r_1} = \frac{\Omega_1}{\Omega_\phi} = \frac{\omega_1}{\omega_\phi}, \quad \Omega_\phi = T_\phi^{-1}. \quad (13)$$

Здесь: r_1 , r_2 - *формирующие параметры* (масштабные множители) соответственно первого и второго каналов, имеющие размерность времени; в этом случае собственная частота $\Omega_i = r_i^{-1}$, ($i = 1, 2$) определяет полосу пропускания соответствующего i -го канала с эталонной настройкой, c^{-1} ; ξ_ω - коэффициент демпфирования первого канала с эталонной настройкой. Из (10)-(13) следует, что безразмерный параметр z характеризует отношение полос пропускания точного и грубого каналов и определяется отношением их собственных частот; параметры γ_x и γ_f , γ_ϕ - отношение полосы пропускания основного (грубого) канала к ширине спектра производной полезного сигнала Ω_x и спектров помех Ω_f , Ω_ϕ соответственно. Обычно для воспроизводящих систем выполняются следующие условия: полезный сигнал узкополосен, т.е. $T_x \gg r_1$ и $\gamma_x^{-1} \ll 1$, а помехи $f(x)$, $\phi(t)$ близки к процессам типа белого шума, т.е. $T_f \ll r_1$, $T_\phi \ll r_1$ и $\gamma_f \ll 1$, $\gamma_\phi \ll 1$.

Расчет и анализ показателей эффективности (составляющих выигрыша в точности G) g_x (6) и g_f , g_ϕ (7), выполненный в [10] при $0 < \xi_1 < 1$, показал, что компенсирующее действие точного канала на ошибки по

полезному сигналу $x(t)$ и помехе $f(t)$ тем выше, чем больше z (меньше значение параметра r_2), и при $z > 1$ величины g_x, g_f пропорциональны z^2 . Вместе с тем, показатель g_φ с ростом z ($z > 1$) уменьшается, поскольку величина g_φ пропорциональна z^{-1} .

Компенсация помехи $f(t)$ эффективна при $z > 1$, когда $g_f > 3$. Но для значений $z > 1$ показатель $g_x > \gamma_x$ и при $\gamma_x^{-1} = 0,01$ (что имеет место в случае узкополосных полезных сигналов, характерных для большинства промышленных воспроизводящих САУ) $g_f > 100$, т.е. дисперсия ошибки по задающему воздействию благодаря наличию дополнительного компенсирующего канала уменьшается на два порядка и более, и может быть пренебрежимо малой. Кроме того, динамические характеристики основной (грубой) воспроизводящей системы обычно таковы, что дисперсии ошибок по полезному сигналу $\overline{\delta_x^2}$ и помехе $\overline{\delta_f^2}$, имеют один порядок малости (при сохранении неравенства $\overline{\delta_f^2} < \overline{\delta_x^2}$), тогда согласно (5) $1 < \mu < 10$. Так, например, при $\gamma_x = 100$ $\mu = 5$. Тогда составляющей μg_x^{-1} в (3) можно пренебречь ($\mu g_x^{-1} \ll g_f^{-1} + \lambda g_\varphi^{-1}$); при этом

$$G \approx G_1 = (1 + \mu) \frac{1}{g_f^{-1} + \lambda g_\varphi^{-1}}. \quad (14)$$

Для оценки погрешности такого представления выигрыша в точности G воспользуемся расчетными данными, полученными в [10]. Для $\xi_1 = 0,5$ и $r_1 = r_1^* = 0,1c$ при $z \geq 1$, $\gamma_x^{-1} \ll 1$, $0 < \gamma_f < 1$ имеем

$$g_x \gamma_x^{-1} > 1, \quad g_f \approx 2, \quad \mu \approx \mu_0 = \frac{8\xi_1^3 r_1^2}{\gamma_x} \cdot \frac{S_x(0)}{S_f(0)}.$$

Так, при $z = 1$, $\gamma_x^{-1} = 0,01$ ($T_x = 10c$)

$$g_x > \gamma_x = 100, \quad g_f \approx 3, \quad \mu = 5. \quad (15)$$

Для показателя g_φ при $z \geq 1$, $0 < \gamma_\varphi < 1$,

$$0 < g_\varphi < 1,33. \quad (16)$$

Кроме того, при $\xi_1 \leq 1$ и ширине спектров помех, большей полосы пропускания основного канала ($\gamma_f, \gamma_\varphi < 1$) выполняется: $|\Delta\lambda| < 1$, и в первом приближении

$$\lambda \approx \lambda_0 = \frac{S_\varphi(0)}{S_f(0)} = \rho^2.$$

Погрешность $\Delta G = (G_1 - G)/G$ вычисления G по формуле (14) обратно пропорциональна интенсивности помехи $\varphi(t)$ в точном канале. Влияние $\varphi(t)$ можно оценить по величине λ_0 . Как правило, $\lambda_0 \ll 1$, однако для полноты оценки ΔG рассмотрим 4 случая: $\lambda_0 = 0; 0,01; 0,1; 1$. Учитывая (15) и выбирая в (16) самый неблагоприятный случай $g_\varphi = 1,33$, для погрешности ΔG получим

$$\Delta G = \frac{\mu g_x^{-1}}{g_f^{-1} + \lambda g_\varphi^{-1}} <$$

$$\begin{aligned} & \left. \begin{aligned} & 0,47 \text{ (или } \Delta G < 4,7\% \text{) при } \lambda \approx \lambda_0 = 1 \\ & \text{(т. е. интенсивности помех} \\ & \text{точном и грубом каналах одинаковы),} \\ & 0,124 \text{ (или } \Delta G < 12,4\% \text{) при } \lambda \approx \lambda_0 = 0,1 \\ & \text{(т.е. } S_\varphi(0) = 0,1 S_f(0), \rho \approx 0,32 \text{),} \\ & 0,149 \text{ (или } \Delta G < 14,9\% \text{) при } \lambda \approx \lambda_0 = 0,01 \\ & \text{(т.е. } S_\varphi(0) = 0,01 S_f(0), \rho \approx 0,1 \text{),} \\ & 0,152 \text{ (или } \Delta G < 15,2\% \text{) при } \lambda \approx \lambda_0 = 0 \\ & \text{(т.е. помеха } \varphi(t) \text{ в точном канале} \\ & \text{практически отсутствует).} \end{aligned} \right\} \quad (17) \end{aligned}$$

Соотношения (17) дают оценку погрешности сверху вычисления значений *выигрыша в точности* G по приближенной формуле (14) для $z \geq 1$ и при различных λ . С увеличением γ_x^{-1} ($\gamma_x^{-1} < 1$) погрешность ΔG возрастает и при $\gamma_x^{-1} = 0,1$ ($T_x = 1c$) для $z \geq 1$ достигает $\Delta G \leq 0,395$ (или $\Delta G \leq 39,5\%$) уже при $\lambda = 1$.

Для того чтобы при $\gamma_x^{-1} = 0,1$ ($T_x = 1c$) и $0 \leq \lambda \leq 1$ погрешность ΔG не превышала значений, определенных формулой (17), безразмерный параметр z должен удовлетворять неравенству $z \geq 5,5$, а при $\gamma_x^{-1} = 0,5$ ($T_x = 0,2c$) и $0 \leq \lambda \leq 1$ для z должно уже выполняться условие $z \geq 18$.

Рассмотрим случай (при достаточно малых значениях γ_x^{-1} и $z \geq 1$), когда составляющую ошибки итерационной системы от полезного сигнала можно не учитывать. Тогда выигрыш в точности G (3) с точностью до составляющей μg_x^{-1} будет определяться выражением

$$G_1 = G_{01} \left(\frac{1}{1 - \Delta G_1} \right), \quad (18)$$

где с учетом (4), (5), (7)

$$G_{01} = (1 + \mu_0) \frac{1}{g_{0f}^{-1} + \lambda_0 g_{0\varphi}^{-1}}, \quad (19)$$

$$G_1 = \frac{1}{1 + \mu_0(1 - \Delta\mu)} \times \left[\frac{G_0}{g_{0f}} \Delta g_f + \frac{G_0}{g_{0f}} \lambda_0 (\Delta\lambda + \Delta g_\varphi - \Delta\lambda \Delta g_\varphi) - \mu_0 \Delta\mu \right]. \quad (20)$$

Как отмечалось, для итерационных двухканальных воспроизводящих систем характерно $\gamma_x^{-1} \ll 1$, $\gamma_f \ll 1$, $\gamma_\varphi \ll 1$ и, как правило, $z \geq 1$ (что обеспечивается законом-мерным расширением полосы пропускания компенсирующего канала по сравнению с основным). Тогда для модулей поправок $\Delta\mu$, $\Delta\lambda$, Δg_f , Δg_φ , как показано в [10], выполняется $|\Delta\mu| \ll 1$, $|\Delta\lambda| \ll 1$, $|\Delta g_f| \ll 1$ и $|\Delta g_\varphi| \ll 1$ (при $z < \gamma_\varphi^{-1}$), и в качестве первых приближений можно принять

$$\mu \approx \mu_0, \quad \lambda \approx \lambda_0, \quad g_f \approx g_{0f} = z^2 + 2\xi_1 z + 1, \quad g_\varphi \approx g_{0\varphi} = \frac{1}{2\xi_1 z}. \quad (21)$$

Пусть в наиболее неблагоприятном случае помехи $f(t)$ и $\varphi(t)$ представляют собой белый шум, при этом [10]

$$\gamma_f = \gamma_\varphi = 0, \quad \Delta\lambda = \Delta g_f = \Delta g_\varphi = 0 \quad (22)$$

и

$$\lambda = \lambda_0, \quad g_f = g_{0f}, \quad g_\varphi = g_{0\varphi}, \quad \mu = \mu_0 \frac{1 + a_{\xi_1} \gamma_x^{-1}}{1 + 2\xi_1 \gamma_x^{-1} + \gamma_x^{-2}}, \quad (23)$$

где

$$a_{\xi_1} = \frac{1 + 4\xi_1^2}{8\xi_1^3}. \quad (24)$$

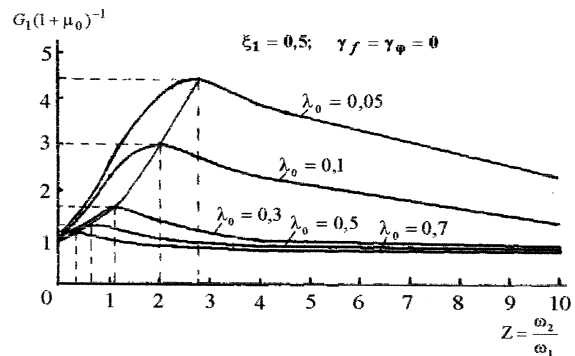
При $\gamma_x^{-1} \ll 1$ и $2\gamma_x^{-1/2} \leq \xi_1 \leq 1$, как видно из (5), (23), (24), имеют место соотношения $|\Delta\mu| \ll 1$ и $\mu \approx \mu_0$, откуда для поправки ΔG_1 согласно (20), (22) выполняется неравенство $|\Delta G_1| \ll 1$. Пренебрегая значениями поправки $\Delta\mu$ в рассматриваемом случае и полагая $\mu = \mu_0$, получим

$$\Delta G_1 = 0. \quad (25)$$

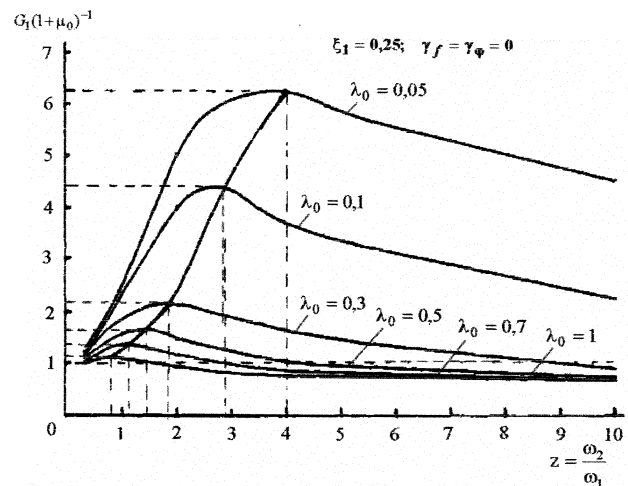
Тогда для оценки выигрыша в точности G_1 из (18), (19), (21), (25) окончательно найдем

$$G_1 = (1 + \mu_0) \frac{1}{g_{0f}^{-1} + \lambda_0 g_{0\varphi}^{-1}} = \frac{1 + \mu_0}{2\xi_1 \lambda_0} \times \frac{z^2 + 2\xi_1 z + 1}{z^3 + 2\xi_1 z^2 + z + \frac{1}{2\xi_1 \lambda_0}}. \quad (26)$$

На рис. 1, а, б приведены графики зависимости величины $G_1(1 + \mu_0)^{-1}$ от параметров z и λ_0 при заданных значениях $\xi_1 = 0,5$; $\xi_1 = 0,25$ соответственно. Из графиков видно, что более глубокий экстремум выигрыша в точности G_1 имеет место при достаточно малых λ_0 ($\lambda_0 \leq 0,3$), при которых оптимальное значение параметра z_{opt} > 1 . Линия, соединяющая точки экстремума, определяет оптимальные значения z_{opt} и $G_{1opt}(1 + \mu_0)^{-1}$ для различных значений λ_0 .



а)



б)

Рисунок 1 - Графики зависимости величины $G_1(1 + \mu_0)^{-1}$ от параметров z и λ_0 при заданных значениях $\xi_1 = 0,5$ (а) и $\xi_1 = 0,25$ (б), и при $\gamma_f = \gamma_\varphi = 0$

В ряде практических случаев случайная помеха $f(t)$, воздействующая на основной канал, имеет относительно узкополосный спектр $S_f(\omega)$, ширина которого ($\omega_f = 2\pi T_f^{-1}$) сопоставима с полосой пропускания основного канала ($\omega_1 = 2\pi r_1^{-1}$) [2, 8]. Полагая $\gamma_f \neq 0$ и $\gamma_\varphi = 0$, и используя результаты, полученные в [10], имеем

$$\mu = \mu_0 \frac{(2\xi_1 + c_{\gamma_f})(1 + a_{\xi_1} \gamma_x^{-1})}{2\xi_1(1 + 2\xi_1 \gamma_x^{-1} + \gamma_x^{-2})}, \quad \lambda = \lambda_0 \left(1 + \frac{c_{\gamma_\lambda}}{2\xi_1} \right),$$

$$g_f = g_{0f} \left(1 - \frac{c_{\gamma_f z}}{\gamma_f z + 1} \right),$$

$$g_{\varphi} = g_{0\varphi}, \quad (27)$$

$$\text{где } c_{\gamma_f} = \frac{a_{\xi_1} \gamma_f^2}{1 + 2\xi_1 \gamma_f}.$$

Тогда для выигрыша в точности G_1 согласно (14), (21), (27) получим

$$G_1 = \frac{1 + \mu}{2\xi_1 \lambda_0 \left(1 + \frac{c_{\gamma_f}}{2\xi_1} \right)} \times \frac{(z^2 + 2\xi_1 z + 1)(\gamma_f z + 1)}{z(z^2 + 2\xi_1 z + 1)(\gamma_f z + 1) + \left[2\xi_1 \lambda_0 \left(1 + \frac{c_{\gamma_f}}{2\xi_1} \right) \right]^{-1} \left[\gamma_f - c_{\gamma_f} \right] z + 1}. \quad (28)$$

Заметим, что при $\xi_1 = \text{const}$ и $\gamma_f = \text{const}$ характер изменения кривых зависимости $G_1(1 + \mu)^{-1} = \Psi(z, \lambda_0)$ аналогичен характеру кривых на рис.1; а при $\gamma_f = \text{var}$ ($0 < \gamma_f < 1$) значения характеристик отличаются друг от друга незначительно.

3 ВТОРОЙ (УПРАВЛЯЮЩИЙ) ВАРИАНТ ДВУХКАНАЛЬНОЙ СИСТЕМЫ

Предположим теперь, что структуры эталонных операторов первого и второго каналов одинаковы и соответствуют колебательным звеньям с одинаковыми частотами собственных колебаний $\Omega_{ki} = \Omega_{\vartheta} = T_{\vartheta}^{-1} = 1$, ($i = 1, 2$). Такое сочетание динамики каналов более характерно для итерационных двухканальных управляющих систем, выход которых связан с силовой нагрузкой (см., например, рис. 1.10, 1.11 в [7]). Тогда операторы каналов (8) с учетом (10) примут вид

$$W_1^*(p) = \frac{1}{r_1 p^2 + 2\xi_1 r_1 p + 1}, W_2^*(p) = \frac{z^2}{r_1^2 p^2 + 2\xi_2 z r_1 p + z^2}, \quad (29)$$

где $\xi_1 = \xi_{\vartheta 1}$, $\xi_2 = \xi_{\vartheta 2}$ - коэффициенты демпфирования соответственно первого и второго каналов с эталонной настройкой. При $\xi_1 = \xi_2 = \sqrt{2}/2 \approx 0,707$, $r_1 = T_{\vartheta}$ и $z = 1$ операторы $W_1^*(p)$ и $W_2^*(p)$ (29) совпадают с оператором фильтра Баттерворта 2-го порядка.

Расчет и анализ показателей эффективности g'_x , g'_f , g'_{φ} , выполненный в [10] при $0 \leq \xi_1 \leq 1$ и $0 \leq \xi_2 \leq 1$, показал, что в случае задания операторов

автономных каналов в виде (29) для выигрыша в точности G при $z \geq 1$ и $\gamma_x^{-1} \ll 1$ также выполняется допущение, аналогичное (14).

Параметры μ и λ характеризуют относительную точность первого (основного) канала, оператор которого $W_1^*(p)$ (29) во втором варианте не изменился. Поэтому результаты определения μ и λ , полученные для первого варианта, полностью справедливы и для второго варианта задания операторов автономных каналов. Кроме того, при $\gamma_f \ll 1$, $\gamma_{\varphi} \ll 1$ и $z \geq 1$ для модулей поправок $\Delta g'_f$ и $\Delta g'_{\varphi}$, как показано в [10], также выполняются неравенства: $|\Delta g'_f| \ll 1$ и $|\Delta g'_{\varphi}| \ll 1$ (при $z < \gamma_{\varphi}^{-1}$). Тогда в качестве первых приближений можно принять

$$\mu \approx \mu_0, \quad \lambda \approx \lambda_0, \quad g'_{\varphi} \approx g'_{0\varphi} = \frac{\xi_2}{\xi_1 z},$$

$$g'_f \approx g'_{0f} = \xi_2 z^2 \times \frac{b_{f_0} z^6 + b_{f_1} z^5 + b_{f_2} z^4 + b_{f_3} z^3 + b_{f_4} z^2 + b_{f_5} z + b_{f_6}}{a_{f_0} z^6 + a_{f_1} z^5 + a_{f_2} z^4 + a_{f_3} z^3 + a_{f_4} z^2}, \quad (30)$$

где

$$\begin{aligned} a_{f_0} &= 4\xi_2^3 \gamma_f^2, \quad a_{f_1} = 8\xi_2^4 (1 + 2\xi_1 \gamma_f) \gamma_f, \\ a_{f_2} &= \xi_2 [4\xi_2^2 + 2\xi_1 (1 + 8\xi_2^2) \gamma_f + (1 + 16\xi_1^2 \xi_2^2) \gamma_f^2], \\ a_{f_3} &= \xi_1 (1 + 4\xi_2^2) + a_{\xi_2} \gamma_f + \xi_1 (1 + 4\xi_2^2) \gamma_f^2, \\ a_{f_4} &= \xi_2 (1 + 2\xi_1 \gamma_f); \end{aligned} \quad (31)$$

$$\begin{aligned} b_{f_0} &= b_{\gamma_f} \gamma_f^2, \quad b_{f_1} = 2\xi_2 b_{\gamma_f} (1 + 2\xi_1 \gamma_f) + \gamma, \\ b_{f_2} &= 1 + 2\xi_1 (1 + 4\xi_2^2) \gamma_f + (2a_{\xi_2} - 1) \gamma_f^2 + \\ &+ 4\xi_1 (b_{\xi} - 2\xi_1^2 + 1) \gamma_f^3 + b_{\xi} \gamma_f^4, \end{aligned}$$

$$\begin{aligned} b_{f_3} &= 4\xi_2 [\xi_1 + (b_{\xi} - 2\xi_2^2 + 1) \gamma_f + 4\xi_1 (\xi_1^2 + \xi_2^2) \gamma_f^2 + \\ &+ (b_{\xi} - 2\xi_2^2 + 1) \gamma_f^3 + \xi_1 \gamma_f^4], \end{aligned}$$

$$\begin{aligned} b_{f_4} &= b_{\xi} + 4\xi_1 (b_{\xi} - 2\xi_1^2 + 1) \gamma_f + (2a_{\xi_2} - 1) \gamma_f^2 + \\ &+ 2\xi_1 (1 + 4\xi_2^2) \gamma_f^3 + \gamma_f^4, \end{aligned}$$

$$b_{f_5} = 2\xi_2 b_{\gamma_f} (2\xi_1 + \gamma_f), \quad b_{f_6} = b_{\gamma_f}; \quad (32)$$

$$a_{\xi_2} = 2(\xi_1^2 + \xi_2^2 + \xi_1^2 \xi_2^2), \quad (33)$$

$$b_{\xi} = 2(2\xi_1^2 + \xi_2^2 - 1), \quad b_{\gamma_f} = 1 + 2\xi_1 \gamma_f + \gamma_f^2. \quad (34)$$

Полагая далее $\mu = \mu_0$, $\lambda = \lambda_0$, $g'_f = g'_{0f}$, $g'_{\varphi} = g'_{0\varphi}$

из (14), (30) для оценки выигрыша в точности G получим

$$G \approx G_2 = (1 + \mu_0) \frac{1}{(g'_0 f)^{-1} + \lambda_0 (g'_0 \phi)^{-1}} = \frac{\xi_2}{\xi_1} \cdot \frac{1 + \mu_0}{\lambda_0} \times$$

$$\times \left\{ b_{f_0} z^6 + b_{f_1} z^5 + b_{f_2} z^4 + b_{f_3} z^3 + b_{f_4} z^2 + b_{f_5} z + b_{f_6} \right\} /$$

$$\left\{ b_{f_0} z^7 + b_{f_1} z^6 + b_{f_2} z^5 + \left\{ \frac{a_{f_0}}{\xi_1 \lambda_0} + b_{f_3} \right\} z^4 + \left\{ \frac{a_{f_1}}{\xi_1 \lambda_0} + b_{f_4} \right\} z^3 + \right.$$

$$\left. + \left\{ \frac{a_{f_1}}{\xi_1 \lambda_0} + b_{f_3} \right\} z^2 + \left\{ \frac{a_{f_3}}{\xi_1 \lambda_0} + b_{f_6} \right\} z + \frac{a_{f_4}}{\xi_1 \lambda_0} \right\}. \quad (35)$$

В качестве примера, упрощающего функционал (35), рассмотрим достаточно типичный частный случай, когда эталонные операторы каналов соответствуют оператору фильтра Баттерворта 2-го порядка, а помехи $f(t)$ и $\phi(t)$ представляют собой белый шум, при этом

$$\gamma_f = \gamma_\phi = 0, \quad \xi_1 = \xi_2 = \frac{\sqrt{2}}{2}, \quad r_1 = T_\varepsilon. \quad (36)$$

Тогда согласно (31)-(34), (36) выигрыш в точности G_2 (35) примет вид

$$G_2 = \frac{1 + \mu_0}{\lambda_0} \cdot \frac{z^3 + z^2 + z + 1}{z^4 + z^3 + z^2 + \left(1 + \frac{2}{\lambda_0}\right) z + \frac{1}{\lambda_0}}. \quad (37)$$

График зависимости величины $G_2(1 + \mu_0)^{-1}$ (37) от параметров z и λ_0 при заданных $\xi_1 = \xi_2 = \sqrt{2}/2$ и $\gamma_f = \gamma_\phi = 0$ приведен на рис.2. Из рисунка видно, что, как и в первом варианте задания эталонных операторов каналов, более глубокий экстремум выигрыша в точности G_2 имеет место при достаточно малых значениях λ_0 ($\lambda_0 \leq 0,1$), при которых оптимальные значения параметра $z'_{opt} > 3$

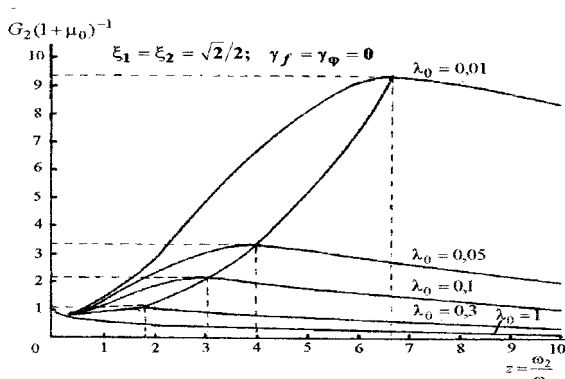


Рисунок 2 - График зависимости величины $G_2(1 + \mu_0)^{-1}$ от параметров z и λ_0 при заданных значениях $\gamma_f = \gamma_\phi = 0$

Отметим, что во втором варианте эталонных операторов каналов при таком же задании $\xi_1 = \xi_2 = \text{const}$, что и в первом варианте, для достижения экстремума выигрыша в точности G_2 требуется уже большее расширение полосы пропускания второго, точного канала по сравнению с первым каналом. Так, например, в первом варианте при $\xi_1 = \sqrt{2}/2$, $\lambda_0 = 0,05$ имеем $z_{opt} = 2,5$; $G_{1max}(1 + \mu_0)^{-1} = 4$; а во втором варианте при $\xi_1 = \xi_2 = \frac{\sqrt{2}}{2}$, $\lambda_0 = 0,05$ получаем: $z'_{opt} = 4$; $G_{2max}(1 + \mu_0)^{-1} = 3,3$.

ВЫВОДЫ

1. Для различных вариантов задания эталонных операторов автономных каналов управления итерационных двухканальных воспроизводящих систем при случайных входных воздействиях получены суммарные функционалы качества для оценки выигрыша в точности G таких систем по сравнению с одноканальными системами.

2. Показано, что при достаточно узкополосных полезных сигналах и широкополосных помехах в каналах составляющей ошибки итерационных систем от полезного сигнала можно пренебречь. Это позволяет значительно упростить функционалы качества без существенных погрешностей в оценке выигрыша в точности G .

3. Полученные суммарные функционалы (26), (28), (35) и (37) позволяют при заданном первом, грубом канале анализировать влияние полосы пропускания второго, точного, канала и характеристик входных сигналов на эффективность и процессы экстремальной самонастройки итерационных двухканальных систем при случайных воздействиях.

ПЕРЕЧЕНЬ ССЫЛОК

1. Клубникин П.Ф. Объединенные следящие системы с двумя приводами // Автоматика и телемеханика. - 1959. - Т. XX. - № 2. - С. 161-175.
2. Осмоловский П.Ф. Итерационные многоканальные системы автоматического управления. - М.: Сов. радио, 1969. - 256с.
3. Следящие приводы: В 2-х кн./ Под ред. Б.К.Чемоданова. , - М.: Энергия, 1976. - Кн. Первая: - 480с.
4. Никольский А.А. Точные двухканальные следящие электроприводы с пьезокомпенсаторами. - М.: Энергоатомиздат, 1988. - 160 с.
5. Осмоловский П.Ф., Худяев А.А. Влияние запаздывания входных координат на динамическую точность двухканальной итерационной измерительной системы // Автоматика, - 1990. - Ж2. - С.35-42.
6. Двухякорный линейный синхронный привод обрабатывающего центра / Б.И. Кузнецов, А.А. Худяев, И.М. Некрасов, В.И. Русаев // Электротехника. - 1993. - №4. - С. 11-18.
7. Многоканальные итерационные системы управления: Учебное пособие / Б.И. Кузнецов, А.А. Худяев, И.Н. Богаенко и др. - К.: НПК "КИА", 1998. - 224 с.
8. Худяев А.А., Гвоздева Е.В. Автоматизированное проектирование итерационных многоканальных систем управления с эталонной настройкой каналов // Вестник ХГПУ. Сборник научных трудов. Тематический выпуск 113. - Харьков: ХГПУ, 2000. - С. 49-56.
9. Худяев А.А. Критерии эффективности итерационных двухканальных воспроизводящих систем при случайных

УДК 681.511.46

МОДИФИКАЦИЯ МЕТОДА КОНЕЧНОГО СОСТОЯНИЯ ДЛЯ СИНТЕЗА УПРАВЛЕНИЯ РЕАЛЬНОГО ВРЕМЕНИ В МНОГОМЕРНЫХ НЕЛИНЕЙНЫХ ТЕРМИНАЛЬНЫХ СИСТЕМАХ С ДИФФЕРЕНЦИРУЕМЫМИ НЕЛИНЕЙНОСТЯМИ

Е.А.Шушляпин, А.Е.Шушляпина

Запропоновано модифікацію методу кінцевих станів для синтезу керувань реального часу стосовно до багатомірних нелінійних термінальних систем з диференційованими нелінійностями. Модифікований метод застосовано для моделювання керування перехідним процесом при включенні електродвигуна постійного струму послідовного збудження.

Предложена модификация метода конечных состояний для синтеза управлений реального времени применительно к многомерным нелинейным терминальным системам с дифференцируемыми нелинейностями. Модифицированный метод применен для моделирования управления переходным процессом при включении электродвигателя постоянного тока последовательного возбуждения.

There was offered a terminal state method modification for synthesis real time controls in multidimensional nonlinear terminal systems with differentiable nonlinearities. The modified method was applied for modeling of control a transient process on switching on a direct current electric motor of consequent drive.

ВВЕДЕНИЕ И ПОСТАНОВКА ЗАДАЧИ

Управление нелинейными системами является одной из актуальных задач управления, поскольку иногда объекты управления функционируют в таких режимах работы, когда соответствующие линейные описания становятся неадекватными. Такие режимы имеют место при разгоне и торможении электрических машин, в химической кинетике, при управлении движущимися объектами, управлении потоками жидкостей и газов и в других отраслях. В ряде работ [1-6] развивается подход к синтезу управляющих воздействий для многомерных терминальных нелинейных систем с дифференцируемыми нелинейностями и аддитивными управлениями вида

$$\frac{dx}{dt} = \Phi(t, x) + B(t)u(t), \quad (t \in [t_0, t_f], x(t_0)) = x^0, \quad (1)$$

$$x \in R^n \times 1, u \in R^r \times 1,$$

Упомянутый подход, названный методом конечных состояний, применяется для задач терминального управления с моделью (1) и критерием

$$J = J(x(t_f, t)) \rightarrow J^*, \quad (2)$$

где J^* - заданное значение, а целевая функция $J(x)$ предполагается дифференцируемой.

Алгоритм расчета управления методом конечных состояний имеет вид

$$u(t) = G^+(t_f, t) \cdot f_j(\bar{J}(t_f, t), J^*),$$

$$G^+(t_f, t) = \frac{G^T}{GG^T} \in R^{r \times 1},$$

$$G = \left. \frac{\partial J(x)}{\partial x} \right|_{x = \bar{x}(t_f, t)} = W(t_f, t, x(t))B(t) \in R^{1 \times r},$$

$$\frac{dW(\vartheta, t, x(t))}{d\vartheta} = \left[\left. \frac{\partial \Phi(\vartheta, x)}{\partial x} \right|_{x = \bar{x}(\vartheta, t)} \right] \cdot W(\vartheta, t, x(t)), \quad (3)$$

$$\frac{d\bar{x}(\vartheta, t)}{d\vartheta} = \Phi(\vartheta, t),$$

$$\vartheta \in [t, t_f], W(t, t, x(t)) = I, \bar{x}(t, t) = x(t).$$

В алгоритме (3) I - единичная матрица, W - $(n \times n)$ -нелинейная весовая матрица, $f_j(\bar{J}(t_f, t), J^*)$ - правая часть скалярного дифференциального уравнения относительно так называемой критериальной функции $\bar{J}(t_f, t)$, обеспечивающего ее приближение при $t \rightarrow t_f$ к J^* . Критериальная же функция - это целевая функция критерия, где вектор состояния $x(t)$ заменен вектором переменных конечного состояния $\bar{x}(t_f, t)$. Переменные конечного состояния, как видно из (3), представляют собой прогнозные значения конечного состояния системы (1), находящейся в момент t в состоянии $x(t)$, когда на интервале $\vartheta \in [t, t_f]$

управление обнулено. Одним из возможных видов дифференциального уравнения для желаемой критериальной функции может быть

$$\frac{d\bar{J}(t_f t)}{dt} = \frac{J^* - \bar{J}(t_f t)}{T}. \quad (4)$$

При этом $\bar{J}(t_f t)$ по экспоненте с постоянной времени T приближается к заданному значению критерия J^* , а в силу равенства переменных состояния и переменных конечного состояния при одинаковых значениях первого и второго аргументов целевая функция критерия также будет близка к J^* . Параметр T влияет как на точность приближения к J^* , так и на энергию управления (чем меньше T , тем выше точность, но больше энергия и наоборот). Как и в других методах терминального управления, заданное значение J^* может оказаться недостижимым, что проявляется в отклонении фактического поведения критериальной функции от экспоненциального (4). При этом можно идти двумя путями: подбирать достижимое значение J^* , удаляя его от желаемого значения, либо оставить желаемое значение, согласившись с отклонением фактического поведения критериальной функции от желаемой траектории. Опыт применения метода конечного состояния показывает, что чаще более предпочтителен второй путь, особенно при наличии амплитудных ограничений на компоненты вектора управлений.

Из (3) видно, что для расчета управления в каждый момент времени t требуется интегрировать систему дифференциальных уравнений порядка $n(1+n)$ на интервале $[t, t_f]$. В распространенном на практике частном случае неотрицательно определенной квадратической целевой функции $J = x^T(t_f) F x(t_f)$ с диагональной матрицей F , а также раздельного вхождения компонент вектора управляющих воздействий в уравнения (1), количество требуемых столбцов весовой матрицы равно размерности r вектора управлений. В этом случае порядок системы (3) может быть снижен до $n(1+r)$. Поскольку моментов t , в которые рассчитывается управление, бесчисленное множество (практически - десятки и сотни), то затрачиваемое на расчет управлений время относительно велико. Поэтому часто данный подход неприменим для синтеза управлений, т.е. для расчета сигналов обратных связей, формируемых управляющим компьютером в реальном времени. В настоящей статье предлагается приближенная процедура расчета управлений, пригодная для их синтеза в реальном времени. При этом предполагается, что весь вектор состояния известен (т.е. измеряется) в каждый момент времени.

МОДИФИКАЦИЯ МЕТОДА КОНЕЧНЫХ СОСТОЯНИЙ

Совместная система уравнений для определения нелинейной весовой матрицы W и переменных конечного состояния \bar{x} в (3) представляет собой систему нелинейных дифференциальных уравнений с дифференцируемыми функциями правых частей без внешних воздействий. Это соответствует системе вида

$$\frac{dy_i(\vartheta)}{d\vartheta} = R_{i0}(\vartheta, y(\vartheta)), \vartheta \in [t, t_f], y(t) = y^t, y = \{y_i\}, \\ i = 1, 2, \dots, n, \quad (5)$$

где компоненты вектора y - требуемые для расчета управления столбцы весовой матрицы, а также переменные конечного состояния, R_{i0} - непрерывные и дифференцируемые функции своих аргументов. Начальные условия y^t - это в соответствии с (3) единицы либо нули для элементов весовой матрицы и $x(t)$ для переменных конечного состояния.

Предположим, что непрерывные и дифференцируемые k_i раз по всем своим переменным правые части системы (5) таковы, что существуют сходящиеся тейлоровские разложения

$$R_{i0}(\vartheta, y(\vartheta)) = R_{i0}(t, y(t)) + R_{i1}(t, y(t))(\vartheta - t) + \dots \\ R_{ik_i}(t, y(t))(\vartheta - t)^{k_i} + \dots, \quad (6)$$

$$R_{ik}(t, y(t)) = \frac{1}{k!} \frac{d^k R_{i0}}{dt^k}, k = 1, 2, \dots, k_i.$$

Составим дифференциальные уравнения для $R_{ik}(\vartheta, y(\vartheta))$, получив предварительно из (6) выражения для коэффициентов. Для сокращения записей аргументы покажем только в первом соотношении.

$$R_{i1}(\vartheta, y(\vartheta)) = \frac{dR_{i0}(\vartheta, y(\vartheta))}{d\vartheta} = \frac{\partial R_{i0}(\vartheta, y(\vartheta))}{\partial \vartheta} + \\ + \frac{\partial R_{i0}(\vartheta, y(\vartheta))}{\partial y} \frac{dy(\vartheta, y(\vartheta))}{d\vartheta},$$

$$R_{i2} = \frac{1}{2} \frac{d^2 R_{i0}}{d\vartheta^2} = \frac{1}{2} \frac{dR_{i1}}{d\vartheta} = \frac{1}{2} \left(\frac{\partial R_{i1}}{\partial \vartheta} + \frac{\partial R_{i1}}{\partial y} \frac{dy}{d\vartheta} \right),$$

$$R_{i3} = \frac{1}{3!} \frac{d^3 R_{i0}}{d\vartheta^3} = \frac{1}{3!} \frac{d}{d\vartheta} \left(\frac{\partial^2 R_{i0}}{\partial \vartheta^2} \right) = \frac{1}{3!} \frac{d}{d\vartheta} (2R_{i2}) = \\ = \frac{1}{3} \left(\frac{\partial R_{i2}}{\partial \vartheta} + \frac{\partial R_{i2}}{\partial y} \frac{dy}{d\vartheta} \right),$$

$$R_{i4} = \frac{1}{4!} \frac{d^4 R_{i0}}{d\vartheta^4} = \frac{1}{4!} \frac{d}{d\vartheta} \left(\frac{d^3 R_{i0}}{d\vartheta^3} \right) = \frac{1}{4!} \frac{d}{d\vartheta} (3! R_{i3}) =$$

$$= \frac{1}{4} \left(\frac{\partial R_{i3}}{\partial \vartheta} + \frac{\partial R_{i3}}{\partial y} \frac{dy}{d\vartheta} \right), \dots,$$

$$R_{ik} = \frac{1}{k} \left(\frac{\partial R_{i(k-1)}}{\partial \vartheta} + \frac{\partial R_{i(k-1)}}{\partial y} \frac{dy}{d\vartheta} \right), \dots \quad (7)$$

Из (7) следует система дифференциальных уравнений для тейлоровских коэффициентов (вновь аргументы запишем только в первом из них)

$$\frac{\partial R_{i0}}{\partial \vartheta}(\vartheta, y(\vartheta)) = R_{i1}(\vartheta, y(\vartheta)) \quad (8)$$

$$\frac{dR_{i1}}{d\vartheta} = 2R_{i2}, \quad \frac{dR_{i2}}{d\vartheta} = 3R_{i3}, \dots, \quad \frac{dR_{ik_i}}{d\vartheta} = (k_i + 1)R_{i(k_i+1)}, \dots$$

Система (8) - незамкнута, поэтому для ее практического использования обнулим правые части k_i -х уравнений для каждого i . Основанием для этого является то, что для сходящихся тейлоровских разложений при $k > k_i$ коэффициенты $R_{ik} \approx 0$. В итоге получится система

$$\frac{dy_i(\vartheta, t)}{d\vartheta} = R_{i0}(\vartheta, y(\vartheta)),$$

$$\frac{dR_{i0}}{d\vartheta} = R_{i1}, \quad \frac{dR_{i1}}{d\vartheta} = 2R_{i2}, \quad \frac{dR_{i2}}{d\vartheta} = 3R_{i3}, \dots, \quad (9)$$

$$\frac{dR_{ik_i}}{d\vartheta} = 0, \quad \vartheta \in [t, t_f],$$

$$i = 1, 2, \dots, n.$$

Номер k_i зависит от i , так как для каждой из нелинейностей количество учитываемых членов тейлоровских разложений может быть различно. Таким образом, общий порядок системы (9) - $n + \sum_{i=1}^n k_i$. Систему (9) назовем расширенной, а ее вектор состояния обозначим через

$$\bar{y} = \left(y_1, y_2, \dots, y_n, R_{10}, \dots, R_{1k_1}, \dots, R_{n0}, \dots, R_{nk_n} \right)^T.$$

Отметим также то, что разложение (6) здесь используется лишь для приведенного выше обоснования способа перехода от незамкнутой системы (8) к замкнутой системе (9).

Для расчета начальных условий при $\vartheta = t$ расширенной системы воспользуемся (7).

$$y_i(t) = y_i^t,$$

$$R_{i0}(t) - \text{известно,}$$

$$R_{ik}(t) = \frac{1}{k} \left(\frac{\partial^k R_{i(k-1)}}{\partial \vartheta^k} + \sum_{j=1}^n \left(\frac{\partial^k R_{i(k-1)}}{(\partial y_j)^k} \right) \frac{dy_j}{d\vartheta} \right) \Bigg|_{\vartheta=t}, \quad (10)$$

$$k = 1, 2, \dots, k_i, \quad i = 1, 2, \dots, n.$$

Простые с виду формулы (10) на самом деле составляют главную трудность применения данного подхода, поскольку трудоемкость безошибочного определения аналитических выражений для производных высокого порядка в силу уравнений (5) лавинообразно возрастает с ростом k . Заметим, что однажды полученные аналитические выражения для начальных условий системы (9) в дальнейшем без всяких изменений используются для всех моментов t . Указанные выражения в конечном счете зависят от вектора состояния $x(t)$ управляемой системы (1). Следует отметить, что данный способ преобразования нелинейной системы в линейную систему увеличенного порядка в принципе можно использовать и непосредственно для моделей вида (1), представляя тейлоровскими разложениями элементы вектор функции Φ . Отметим также, что система вида (9) не является приближенным эквивалентом исходной нелинейной системы вида (5). Эти системы эквивалентны только при определенных, рассчитываемых по (10), начальных условиях.

Для расширенной система (9) нетрудно получить аналитическое решение, которое для первых его компонент имеет вид

$$y_i(t_f) = y_i(t) + \sum_{k=0}^{k_i} \frac{1}{(k+1)!} R_{ik}(t) (t_f - t)^{k+1}, \quad i = 1, 2, \dots, n. \quad (11)$$

Остальные компоненты могут быть также получены, однако для расчета управления они не нужны и выражения для них мы не приводим.

В сравнении с исходным алгоритмом (3), где при расчете каждого текущего значения управляющего вектора требуется интегрировать систему дифференциальных уравнений порядка $n(1+n)$ на интервале $[t, t_f]$, в модифицированном методе в каждый момент времени требуется:

- 1) рассчитать $\bar{y}(t)$ по аналитическим выражениям (10);
- 2) определить по выражению (11) конечные (при $\vartheta = t_f$) значения переменных конечного состояния и столбцов весовой матрицы;
- 3) определить вектор управления по первым трем соотношениям из (3).

Неизвестные заранее количества k_i с теоретической точки зрения должны задаваться достаточно большими, однако практически это лимитировано возможностью определения аналитических выражений для начальных условий расширенной системы (9). Кроме того, при этом увеличивается порядок расширенной системы, однако этот фактор менее важен.

ПРИМЕР

Для демонстрации возможностей модифицированного метода конечных состояний рассмотрим пример из [2], где метод конечного состояния в форме (3) применялся для управления разгоном электродвигателя постоянного тока последовательного возбуждения с приведенной ниже математической моделью

$$L_A \frac{di_A}{dt} + r_A i_A + k_i i_A \omega = u, J_M \frac{d\omega}{dt} + k_M \omega = k_i i_A^2, \quad (12)$$

где i_A - ток якоря, ω - скорость вращения, u - управляющее напряжение на выводах двигателя, r_A, L_A - сопротивление и индуктивность якоря, k_1, k_2 - электрическая и механическая постоянные, k_M - коэффициент пропорциональности момента нагрузки и скорости, J_M - момент инерции, приведенный к валу двигателя. В процессе разгона начальные значения переменных нулевые, а конечные значения определяются из установившегося стационарного режима и удовлетворяют уравнениям

$$(i_A^*)^3 + i_A^* \frac{r_A k_M}{k_1 k_2} - U_{ном} \frac{k_M}{k_1 k_2} = 0, \omega^* = (i_A^*)^2 \frac{k_2}{k_M}, \quad (13)$$

где $U_{ном}$ - номинальное значение напряжения двигателя.

Элементы представления (1) для модели (12) следующие:

$$\Phi = \begin{pmatrix} -\frac{r_A}{L_A} i_A - \frac{k_1 i_A \omega}{L_A} \\ -\frac{k_M}{J_M} \omega + \frac{k_2 i_A^2}{J_M} \end{pmatrix}, B = \begin{pmatrix} 1 \\ L_A \\ 0 \end{pmatrix}, x = \begin{bmatrix} i_A \\ \omega \end{bmatrix}. \quad (14)$$

Как и в [2], цель управления отобразим квадратическим терминальным критерием с неотрицательными коэффициентами F_A, F_ω :

$$J = F_A (i_A(t_f) - i_A^*)^2 + F_\omega (\omega(t_f) - \omega^*)^2 \rightarrow J^*, \quad (15)$$

$$\Delta i_A = i_A(t_f) - i_A^*, \Delta \omega = \omega(t_f) - \omega^*,$$

где i_A, ω^* определяются из нелинейных уравнений (13).

Управление по методу конечных состояний (частный случай (3)) определяется выражением [2]

$$u(t) = \frac{L_A}{2T} \frac{J^* - F_A \Delta i_A^2 - F_\omega \Delta \omega^2}{F_A \Delta i_A^2 W_{11} + F_\omega \Delta \omega W_{21}}, \quad (16)$$

$$\Delta \bar{i}_A \equiv \bar{i}_A(t_f, t) - i_A^*, \Delta \bar{\omega} \equiv \bar{\omega}(t_f, t) - \omega^*,$$

$$W_{i1} \equiv W_{i1}(t_f, t, x(t)), i = 1, 2, x = (i_A, \omega)^T,$$

где желаемое $J^* = 0$, что соответствует точному приведению фазовых координат к заданным значениям согласно критерию (15), T - постоянная времени желаемого экспоненциального поведения критериальной функции (см. пояснения к алгоритму (3)). Как следует из (16), в данном случае нужны переменные конечного состояния $\bar{i}_A, \bar{\omega}$ и первый столбец нелинейной весовой матрицы $W_{i1}, i = 1, 2$. Отсюда следуют представления векторов y и R_0 модели (5):

$$y = (\bar{i}_A, \bar{\omega}, W_{11}, W_{21})^T,$$

$$R_{10} = a_{11} \bar{i}_A + a_{12} \bar{i}_A \bar{\omega}, R_{20} = a_{22} \bar{\omega} + a_{21} \bar{i}_A^2,$$

$$R_{30} = a_{11} W_{11} + a_{12} (\bar{\omega} W_{11} + \bar{i}_A W_{21}),$$

$$R_{40} = 2a_{21} \bar{i}_A W_{11} + a_{22} W_{21},$$

$$a_{11} = -\frac{r_A}{L_A}, a_{12} = -\frac{k_1}{L_A}, a_{21} = \frac{k_2}{J_M}, a_{22} = -\frac{k_M}{J_M}.$$

Элементы расширенной системы (9) при $k = 1, i = 1..2$ следующие:

$$y = (\bar{i}_A, \omega, W_{11}, W_{21}, R_{10}, R_{20}, R_{30}, R_{40}, R_{11}, R_{21}, R_{31}, R_{41})^T,$$

$n = 4$.

Найдем по выражениям (10) начальные условия.

$$\bar{y}_1(t) \equiv \bar{i}_A(t, t) = i_A(t), \bar{y}_2(t) \equiv \bar{\omega}(t, t) = \omega(t),$$

$$\bar{y}_3(t) \equiv W_{11}(t, t, x(t)) = 1, \bar{y}_4(t) \equiv W_{12}(t, t, x(t)) = 0,$$

$$\bar{y}_5(t) \equiv R_{10}(t) = a_{11} i_A + a_{12} i_A(t) \omega(t),$$

$$\bar{y}_6(t) \equiv R_{20}(t) = a_{21} i_A^2(t) + a_{22},$$

$$\bar{y}_7(t) \equiv R_{30}(t) = a_{11} W_{11}(t, t, x(t)) + a_{12} (\omega(t) W_{11}(t, t, x(t)) + i_A(t) W_{21}(t, t, x(t))) = a_{11} + a_{12} \omega(t),$$

$$\bar{y}_8(t) \equiv R_{40}(t) = 2a_{21} i_A(t) W_{11}(t, t, x(t)) +$$

$$+ a_{22} W_{21}(t, t, x(t)) = 2a_{21} i_A(t), \quad (17)$$

$$\bar{y}_9(t) \equiv R_{11}(t) = a_{11} R_{10}(t) + a_{12} R_{10}(t) \omega(t) + i_A(t) (R_{20}(t)),$$

$$\bar{y}_{10}(t) \equiv R_{21}(t) = 2a_{21} i_A(t) R_{10}(t) + a_{22} R_{20}(t),$$

$$\bar{y}_{11}(t) \equiv R_{31}(t) = a_{11} R_{30}(t) + a_{12} (R_{20}(t) W_{11}(t, t, x(t)) +$$

$$+ \omega(t) R_{30}(t) + a_{12} (R_{10}(t) W_{21}(t, t, x(t)) + i_A(t) (R_{40}(t)) =$$

$$= a_{11} R_{30}(t) + a_{12} (R_{20}(t) + \omega(t) R_{30}(t) + i_A(t) (R_{40}(t)),$$

$$\bar{y}_{12}(t) \equiv R_{41}(t) = 2a_{21} [R_{10}(t) W_{11}(t, t, x(t)) + i_A(t) R_{30}(t)] +$$

$$+ a_{22} R_{40}(t) = 2a_{21} [R_{10}(t) + i_A(t) R_{30}(t)] + a_{22} R_{40}(t).$$

Рассчитанные по этим формулам значения подставляются в выражение (11), из которого определяются все требуемые для расчета управления (16) данные, а именно:

$$\bar{i}_A(t_f, t) \equiv y_1(t_f), \bar{\omega}(t_f, t) \equiv y_2(t_f),$$

$$W_{11}(t_f, t, x(t)) \equiv y_3(t_f), W_{21}(t_f, t, x(t)) \equiv y_4(t_f).$$

Графики переходных процессов системы (12) для различных режимов коммутации приведены на рисунке 1.

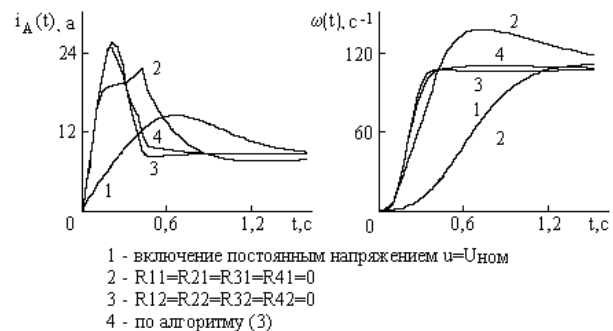


Рисунок 1 - Переходные процессы по току и скорости при разгоне двигателя

Зависимости получены для следующих параметров двигателя мощностью 5 кВт и алгоритмов управления: $r_A = 2 \text{ Ом}$, $L_A = 3 \text{ Гн}$, $k_1 = 0,1 \text{ Ом} \cdot \text{с}$, $k_2 = 1 \text{ н} \cdot \text{м} \cdot \text{А}^{-2}$, $J_M = 1,05 \text{ кг} \cdot \text{м}^{-2}$, $k_M = 0,7 \text{ н} \cdot \text{м} \cdot \text{с}$, $T = 0,01 \text{ с}$, $t_f = 0,4 \text{ с}$, $J^* = 0$. Интегрирование дифференциальных уравнений выполнялось методом Рунге-Кутты 4 порядка на интервале $[0, t_f]$ с управлением (16) и далее продолжалось до $1,5 \text{ с}$ при номинальном входном напряжении $u = U_{ном} = 110 \text{ В}$. Весовые коэффициенты $F_A = F_B = 1$, $F_\omega = 25$, что объясняется желанием в большей мере управлять скоростью двигателя. Управляющие напряжения ограничивались значением $U_{max} = 500 \text{ В}$. Номинальные значения тока и скорости, определяемые уравнениями (13), для перечисленных значений параметров равны: $i_A^* = 8,66 \text{ А}$, $\omega^* = 106,58 \text{ с}^{-1}$.

Кривые 1 соответствуют неуправляемому режиму коммутации, когда на двигатель подается постоянное питающее напряжение, равное номинальному. Кривые 2, 3 получены для управляемой коммутации по алгоритму (16), (11), (17) с различным количеством учитываемых членов тейлоровских разложений (6), а кривые 4 - управляемой коммутации по алгоритму (3). В последнем случае управление также рассчитывалось по выражению (16), однако переменные конечного состояния и весовые функции определялись многократным интегрированием системы уравнений (см. комментарии к алгоритму (3)). По кривым 2 и 3 хорошо видно, как использование большего количества членов тейлоровских разложений приближает процессы к точной реализации (3) метода конечных состояний, представленной кривыми 4. Так, кривые 3, соответствующие учету членов R_{i0} и R_{i1} , практически совпадают с кривыми 4, в то время как при учете только одного члена R_{i0} (кривые 2) наблюдается заметное различие. Конечные значения тока и скорости для режима коммутации, соответствующего кривым 4, равны: $i_A(t_f) = 8,7 \text{ А}$, $\omega(t_f) = 106,7 \text{ с}^{-1}$, что весьма близко к значениям для номинального режима. Время переходных процессов по току и скорости равно $t_f = 0,4 \text{ с}$, что более чем в три раза меньше времени переходного процесса при неуправляемой коммутации. Недостатком алгоритмов управляемой коммутации является увеличенное до 27 А пиковое значение тока, что больше соответствующего значения 14 А при включении постоянным напряжением.

И наконец, сравнивая базовый (3) и модифицированный (16), (11), (17) алгоритмы по времени, приведем следующие данные: моделирование $1,5 \text{ с}$ работы системы (12) с базовым и модифицированным алгоритмами управления на компьютере класса 486DX66 заняло

соответственно $2,5 \text{ с}$ и $0,6 \text{ с}$ машинного времени. Отметим также, что модифицированный алгоритм содержит 46 операций умножения и деления, 26 операций сложения и вычитания и несколько логических операций. Приведенные данные свидетельствуют, во-первых, о существенном уменьшении времени расчета модифицированного управления в сравнении с временем расчета базового управления, и, во-вторых, о возможности реализации модифицированного управления в реальном времени.

Следует отметить также, что достаточное количество учитываемых членов тейлоровского разложения (6) в данном случае оказалось весьма малым.

ВЫВОД

Предложенная в настоящей статье модификация метода конечного состояния намного проще и требует меньших вычислительных ресурсов при реализации, чем базовый алгоритм (3). Из рассмотренного примера видно, что данная модификация может использоваться для управления достаточно быстрыми процессами в реальном времени при использовании в качестве управляющих персональных компьютеров малой производительности.

ПЕРСПЕКТИВЫ

Алгоритмы управления на основе модифицированного метода конечного состояния имеют большие перспективы для практического применения ввиду их достаточной универсальности, несложной подготовительной работы и возможностей использования в реальном времени. Самостоятельную ценность представляет способ линеаризации, на основе которого получен модифицированный метод. Возможность приближенной замены нелинейной модели линейной с рассчитываемыми по исходной модели начальными условиями для дополнительных переменных открывает большие перспективы для использования достижений теории линейных систем в теории и практике управления нелинейными системами вида (1).

ПЕРЕЧЕНЬ ССЫЛОК

1. Шушляпин Е.А. Синтез линейных и нелинейных систем управления конечным положением на основе моделей конечного состояния // Проблемы управления и информатики. - 1997. - №3. С.10-16.
2. Шушляпин Е.А., Канов Л.Н. Синтез управления переходными процессами в нелинейных системах электропривода // Радиоэлектроника. Информатика. Управление. - 1999. - №1. - С.136-140.
3. Шушляпин Е.А., Канов Л.Н. Модели конечного состояния для непрерывно-дискретных систем // Радиоэлектроника. Информатика. Управление. - 1999. - №2. - С.129-132.
4. Шушляпин Е.А., Канов Л.Н. Синтез терминального управления методом конечного состояния // Известия вузов "Электромеханика". - 2000. - №1. - С.72-75.
5. Шушляпин Е.А. Терминальное управление системами с дифференциально-алгебраическими ограничениями методом конечного состояния // Радиоэлектроника. Информатика. Управление. - 2000. - №1. - С.166-172.
6. Шушляпин Е.А., Канов Л.Н. Терминальное управление возбуждением синхронного генератора методом конечных состояний // Известия вузов "Электромеханика". - 2002. - №2. - С.69-72.

АВТОРЫ НОМЕРА

Jamil Ahmad

Dr., Department of Computer Science College of Information Technology, University of Bahrain Kingdom of Bahrain

Вершина Александр Иванович

кандидат технических наук, доцент Запорожского национального технического университета

Chao-Hsing Hsu

Ph.D., Professor at Department of Electronic Engineering, Chienkuo Institute of Technology, Taiwan

Власов Виктор Алексеевич

доктор физико-математических наук, проректор Томского политехнического университета

Ho Dac Loc

The Technical University, Vietnam

Гавриляко Владимир Михайлович

кандидат физико-математических наук, доцент Харьковского национального университета

Kang K. Yen

Ph.D., Professor at Department of Electrical & Computer Engineering, Florida International University, U.S.A.

Горавский Сергей Павлович

аспирант Херсонского государственного технического университет

Karasawa Shinji

Ph.D., Professor at Department of electrical engineering Miyagi National College of technology, Japan

Горбачев Модест Николаевич

кандидат технических наук, старший научный сотрудник Института общей энергетики НАН Украины, г.Киев

Tadeusz Babi

Ph.D., Professor at Department of Electrical & Computer Engineering, Florida International University, U.S.A.

Гостев Владимир Иванович

доктор технических наук, заведующий кафедрой Государственного университета информационно-коммуникационных технологий, г.Киев

Балыкова Александра Юрьевна

аспирантка Пензенского государственного университета

Дзензерский Виктор Александрович

доктор технических наук, директор Института транспортных систем и технологий НАН Украины "Трансмаг", г.Днепропетровск

Баркалов Александр Александрович

доктор технических наук, профессор Донецкого национального технического университета

Дядичев Валерий Владиславович

кандидат технических наук, доцент Восточноукраинского национального университета им. В. Даля, г. Луганск

Берестнева Ольга Григорьевна

кандидат технических наук, заведующая лабораторией Томского политехнического университета

Еремин Евгений Леонидович

доктор технических наук, профессор Амурского государственного университета, г.Благовещенск

Васечкина Елена Федоровна

старший научный сотрудник Морского гидрофизического института НАН Украины

Жигачев Василий Маратович

аспирант Пензенского государственного университета

Веревкин Леонид Леонидович

старший преподаватель Запорожской государственной инженерной академии

-
- Ильина Лариса Владимировна*
аспиранта Амурского государственного университета, г. Благовещенск
- Заварзин Андрей Владимирович*
сотрудник ГИС-Ассоциации, г. Москва
- Карпуков Леонид Матвеевич*
кандидат технических наук, доцент Запорожского национального технического университета
- Крайнев Виктор Владимирович*
кандидат технических наук, докторант Киевского военного института телекоммуникации и информатизации НТУУ “КПИ”
- Красичков Алексей Александрович*
аспирант, ассистент Донецкого национального технического университета
- Кузовенкова Наталья Александровна*
студентка Запорожского национального технического университета
- Куликов Александр Федорович*
кандидат физико-математических наук, доцент Запорожского национального технического университета
- Литвиненко Владимир Иванович*
кандидат технических наук, доцент Херсонского государственного технического университета
- Меньшиков Григорий Григорьевич*
доктор технических наук, заведующий кафедрой Санкт-Петербургского государственного университета
- Месюра Владимир Иванович,*
кандидат технических наук, декан Винницкого государственного технического университета
- Мисюра Андрей Алексеевич*
аспирант Запорожского национального технического университета
- Митрохин Сергей Александрович*
ассистент Восточноукраинского национального университета им. В. Даля, г. Луганск
- Новотарский Михаил Анатольевич*
кандидат технических наук, старший научный сотрудник Института математики НАН Украины
- Онуфриенко Владимир Михайлович*
кандидат физико-математических наук, декан Запорожского национального технического университета
- Орлов Алексей Алексеевич*
доктор технических наук, профессор Томского политехнического университета
- Пархоменко Анжелика Владимировна*
кандидат технических наук, доцент Запорожского национального технического университета
- Пархоменко Лариса Александровна*
кандидат физико-математических наук, доцент Запорожского национального технического университета
- Пица Наталья Дмитриевна*
аспирантка Запорожского национального технического университета
- Плаксин Сергей Викторович*
кандидат физико-математических наук, старший научный сотрудник Института транспортных систем и технологий НАНУ “Трансмаг”, г. Днепропетровск
- Поправка Олег Николаевич*
врач-офтальмолог Городской детской больницы №1, г. Запорожье
- Потапенко Евгений Евгеньевич*
аспирант Запорожского национального технического университета
- Посунько Александр Павлович*
ассистент Запорожской государственной инженерной академии
- Псарев Валентин Иванович*
доктор физико-математических наук, профессор, пенсионер
- Пулов Роман Дмитриевич*
главный специалист отдела банка “БИГ Энергия”, г. Запорожье
- Романенко Сергей Николаевич*
кандидат физико-математических наук, доцент Запорожского национального технического университета

Самойлин Евгений Александрович
адъюнкт Ростовского военного института
ракетных войск

Серпенинов Олег Витальевич
кандидат технических наук, начальник
кафедры Ростовского военного института
ракетных войск

Сизиков Виктор Петрович
кандидат технических наук, старший науч-
ный сотрудник Омского филиала Институ-
та математики им. С. Л. Соболева СО РАН

Скуртов Сергей Николаевич
инженер ОАО “Информационно-компью-
терные системы”, г. Киев

Соколовский Иван Иванович
кандидат физико-математических наук, ве-
дущий научный сотрудник Института тран-
спортных систем и технологий НАНУ
“Трансмаг”, г. Днепропетровск

Солдатов Борис Трофимович
начальник центра информационных техно-
логий Запорожского национального техни-
ческого университета

Субботин Сергей Александрович
преподаватель Запорожского национально-
го технического университета

Фефелов Андрей Александрович
аспирант Херсонского государственного
технического университета

Худяев Александр Андреевич
кандидат технических наук, доцент Укра-
инской инженерно-педагогической акаде-
мии, г. Харьков

Чувькин Борис Викторович
доктор технических наук, профессор
Пензенского государственного университета

Швец Евгений Яковлевич
кандидат технических наук, проректор За-
порожской государственной инженерной
академии

Шинкарук Оксана Васильевна
студентка Запорожского национального
технического университета

Шушляпина Анна Евгеньевна
студентка Севастопольского национального
технического университета

Шушляпин Евгений Андреевич
доктор технических наук, доцент Севасто-
польского национального технического
университета

Эль-Джейби Ахмад Камаль
аспирант Донецкого национального техни-
ческого университета

Ярин Владимир Дмитриевич
старший научный сотрудник Морского
гидрофизического института НАН Украины

