УДК 004.93

Субботин С. А.

Д-р техн. наук, профессор, профессор кафедры программных средств Запорожского национального технического университета, Запорожье, Украина

МЕТОД СИНТЕЗА ДИАГНОСТИЧЕСКИХ МОДЕЛЕЙ НА ОСНОВЕ РАДИАЛЬНО-БАЗИСНЫХ НЕЙРОННЫХ СЕТЕЙ С ПОДДЕРЖКОЙ ОБОБЩАЮЩИХ СВОЙСТВ

В работе решена актуальная проблема автоматизации синтеза радиально-базисных нейронных сетей на основе набора прецедентов для принятия решений в диагностировании. Предложен метод синтеза радиально-базисных нейронных сетей, который формирует в начале по одному эталону класса, которые при необходимости дополняет новыми эталонами, формируемыми на основе ошибочно распознанных экземпляров, а далее оперирует расстояниями от экземпляров до эталонов кластеров. На основе полученных координат эталонов далее в автоматическом режиме синтезируется структура и настраиваются параметры сети, которые дополнительно для повышения обобщающих свойств и интерпретабельности подвергается контрастированию весов. Предложенный метод не требует задания пользователем числа кластеров, не имеет неопределенности выбора числа нейронов в первом слое и выбора начальных значений весов сети, стремится минимизировать размер сети, характеризуется приемлемым временем обучения, благодаря использованию процедуры оптимизации сети позволяет получать безызбыточные контрастные, интерпретабельные нейромодели. Разработано программное обеспечение, реализующее предложенный метод, а также проведены эксперименты, подтвердившие работоспособность разработанного математического обеспечения и позволяющие рекомендовать его для использования на практике при решении задач построения диагностических моделей по прецедентам для автоматизации принятия решений в технической и биомедицинской диагностики.

Ключевые слова: нейронная сеть, радиально-базисная сеть, обучение, синтез, диагностика.

НОМЕНКЛАТУРА

РБНС – радиально-базисная нейронная сеть;

C – множество эталонов;

C^q – q-й эталон;

C_j^q – j-я координата эталона q-го кластера;

E – среднеквадратическая ошибка;

$f()$ – пользовательский критерий, характеризующий качество аргумента относительно решаемой задачи;

$F()$ – структура РБНС;

F_g – показатель обобщения нейромодели;

K – число классов;

M – число слоев РБНС (M=2);

N – число диагностических признаков;

N_M – число выходных признаков;

N_w – число параметров (весов) РБНС;

$N_{w=0}$ – число весов РБНС, равных нулю;

$N_з$ – число узлов в з-м слое;

opt – оптимальное (желаемое или приемлемое) значение функционала $f()$ для решаемой задачи;

Q – число кластеров;

$R(x^s, C^q)$ – расстояние от экземпляра x^s до эталона C^q;

S' – число экземпляров в конфликтном наборе;

S – число прецедентов в выборке;

S^q – число экземпляров, попавших в q-й кластер;

S'^q– число экземпляров q-го класса в конфликтном наборе;

$t_{об.}$ – время обучения модели;

w – параметры РБНС;

w – множество управляемых (настраиваемых) параметров (весов) РБНС;

$w_j^{(з,i)}$ – значение j-го настраиваемого параметра или веса j-го входа i-го узла з-го слоя;

$x_j^{\min}$, $x_j^{\max}$ – минимальное и максимальное значения j-го признака, соответственно;

x_j^s – значение j-го диагностического признака x_j, характеризующее прецедент (экземпляр) x^s;

x'^s – s-й экземпляр конфликтного набора;

x'^s_j – значение j-го диагностического признака x_j, характеризующее прецедент (экземпляр) x'^s;

X' – конфликтный набор;

y^s – значение выходного признака, сопоставленное прецеденту x^s;

y^{s*} – расчетное значение на i-м выходе РБНС для экземпляра x^s, поданного на ее входы;

y^s_i – значение i-го выходного признака для экземпляра x^s;

$y^s_{расч.}$ – расчетный номер класса для s-го экземпляра относительно сформированного множества эталонов;

β_q – коэффициент, регулирующий величину q-го кластера;

$\varphi^{(\eta,i)}$, $\psi^{(\eta,i)}$ – соответственно, дискриминантная (весовая) и активационная функции i-го узла η-го слоя.

ВВЕДЕНИЕ

Искусственные нейронные сети [3–11] благодаря своим способностям способностями обучаться по примерам, посредством обобщения и неявного извлечения знаний из данных, отображая их в структуру и параметры (веса) сети, получили широкое распространение на практике при решении задач диагностирования и распознавания образов.

Объектом исследования являлся процесс синтеза искусственных нейронных сетей.

Среди известных моделей нейронных сетей и методов их обучения [3–11] для задач автоматизации приня-

DOI 10.15588/1607-3274-2016-2-8

тия решений целесообразно использовать преимущественно сети прямого распространения сигнала, которые не содержат боковых и обратных связей и легче поддаются последующему анализу [6, 9, 10]. В свою очередь, среди сетей данного класса особо можно выделить радиально-базисные нейронные сети (РБНС), которые легко интерпретируются в терминологии кластерного анализа, что упрощает последующий анализ полученной на их основе модели или ее решений [3, 6, 11, 12].

Предметом исследования являлись методы построения РБНС.

Известные методы синтеза РБНС обладают такими недостатками, как зависимость качества получаемой модели от заданных человеком параметром, а также низкий уровень обобщения получаемых моделей и большие затраты времени на процесс построения моделей.

Целью работы являлось создание метода структурно-параметрического синтеза РБНС, обеспечивающего сокращение вовлечения человека в процесс построения нейромодели, а также гарантирующего поддержку обобщающих свойств модели.

1 ПОСТАНОВКА ЗАДАЧИ

Пусть задана обучающая выборка прецедентов $X = \langle x, y\rangle$, где $x=\{x^s\}$, $y=\{y^s\}$, $x=\{x_j\}$, $x^s=\{x^s_j\}$, $s = 1, 2, ..., S$, $j = 1, 2, ..., N$.

Тогда задача структурно-параметрического синтеза РБНС заключается в том, чтобы получить $\langle f(), w\rangle$: $y^{s*} = F(w, x^s), f(F(), w, \langle x, y\rangle) \to opt$.

Здесь $\langle F(), w\rangle$ – диагностическая модель на основе РБНС, которая задается кортежем $\langle M, \{N_\eta\}, \{y^{(\eta,i)}(x^{(\eta,i)})\}\rangle$ и описывается функционально формулами:

$$y_i^s = y^{(M,i)}(y^{(M-1,i)}(...y^{(1,i)}(x^s))), \quad y^{(\eta,i)} = \psi^{(\eta,i)}(\varphi^{(\eta,i)}),$$

$$i=1,2,...,N_{\eta-1};\ w^{(\eta,i)} = \{w_j^{(\eta,i)}\}, \quad w = \{w^{(\eta,i)}\} = \{w_j^{(\eta,i)}\},$$

$$y^{(0,j)} = \psi^{(0,j)} = x_j^s, \quad N_0 = N, \quad x_j^{(1,j)} = x_j^s,$$

$$i=1,2,...,N_{\eta-1},\ \eta=1,2,...,M, \quad j=1,2,...,N;$$

$$y^s = \{y^s{}_i\}, y^{s*} = \psi^{(M,i)}(x^s),\ i = 1, 2, ..., N_M.$$

Для задач классификации: $y^s \in \{q\}$, $q = 1, 2, ..., K$, $K>1$.

В простейшем случае f определяют: $f = E$, где

$$E = \frac{1}{2}\sum_{s=1}^{S}(y^s - \psi^{(M,i)}(x^s))^2 \to \min.$$

2 ЛИТЕРАТУРНЫЙ ОБЗОР

Известные методы построения РБНС [8, 10–12] можно разделить на две группы.

Методы первой группы предполагают, что число нейронов РБНС в первом слое задается пользователем, а число нейронов во втором слое определяется размерностью выхода сети. После чего РБНС рассматривается как частный случай многослойной нейронной сети прямого распространения сигнала (многослойного персептрона), которую обучают с помощью градиентных методов многомерной нелинейной безусловной оптимизации [8, 12], а частные производные целевой функции определяют на основе техники обратного распространения ошибки [8, 10, 11]. При этом существует неопределенность выбора числа нейронов в первом слое, что может привести к получению избыточной сети или, наоборот, к невозможности построения модели, обладающей требуемой точностью. Другим недостатком данных методов является неопределенность выбора начальных значений весов сети, что может привести к невозможности решить задачу обучения РБНС за ограниченное время.

Методы второй группы предполагают, в структуру РБНС отображается обучающая выборка. Это, как правило, реализуется простым запоминанием всей выборки, либо на основе кластер-анализа, в результате которого определяется число и координаты центры кластеров, которые заносятся в память РБНС. Число нейронов в первом слое РБНС задают равным числу кластеров, а в веса нейронов первого слоя заносят координаты центров кластеров. Далее РБНС обучают, корректируя веса нейронов второго слоя на основе методов многомерной нелинейной безусловной оптимизации [8, 12]. Если сеть получают путем отображения выборки, то она, как правило, не обеспечивает обобщение. Если же используют кластер-анализ, то качество полученной нейромодели существенно зависит от качества результатов кластер-анализа, сеть может оказаться избыточной и проявлять низкие обобщающие свойства из-за чрезмерной детализации разбиения признакового пространства.

Поэтому необходимо разработать метод структурно-параметрического синтеза РБНС, обеспечивающего сокращение вовлечения человека в процесс построения нейромодели, а также гарантирующего поддержку обобщающих свойств модели.

3 МАТЕРИАЛЫ И МЕТОДЫ

Прямое отображение выборки в структуру РБНС не обеспечивает обобщающих свойств. Для того, чтобы нейросетевая модель обладала обобщающими свойствами необходимо, чтобы число ее параметров (весов) N_w было меньше размерности обучающей выборки NS.

Число весов в радиально-базисной сети определяется числом кластеров Q, на которые разбивается обучающая выборка $\langle x, y\rangle$, числом признаков N и числом классов K и рассчитывается по формуле: $N_w = (N+1)Q + (Q+1)K$.

Здесь первое слагаемое определяет число параметров нейронов первого слоя РБНС, каждый из которых в весах хранит координаты центра соответствующего кластера, а второе слагаемое определяет параметры связей между нейронами первого и второго слоя.

Исходя из этого выражения, определим Q при заданных N, K и S:

$$(N+1)Q + (Q+1)K < NS \Rightarrow Q < \frac{NS-K}{N+K+1}.$$

В то же время, очевидно, что число кластеров Q должно быть не меньше, чем число классов K. Тогда получим диапазон допустимых значений числа кластеров:

$$K \le Q < \frac{NS-K}{N+K+1}.$$

Для того, чтобы обеспечить высокие обобщающие свойства нейромодели, необходимо в процессе ее построения стремиться минимизировать число используемых кластеров и контролировать соблюдение приведенного выше условия при добавлении новых эталонов.

С другой стороны, многие известные методы кластер-анализа являются высоко итеративными и требуют расчета расстояний между всем экземплярами выборки, а затем перебора большого числа различных вариантов разбиений выборки, что приводит к большим затратам времени на обучение, а также может требовать начального задания числа кластеров пользователем. Для устранения данного недостатка предлагается вначале формировать по одному эталону класса (предполагая классы компактными и состоящими из одного кластера), которые при необходимости дополнять новыми эталонами, формируемыми на основе ошибочно распознанных экземпляров. Это позволит сразу найти грубое решение задачи, не выполняя расчета расстояний между всеми экземплярами выборки, а далее оперировать расстояниями от экземпляров до эталонов кластеров, что позволит сократить время расчетов.

На основе полученных координат эталонов предлагается в автоматическом режиме синтезировать РБНС, которую затем для повышения обобщающих свойств и интерпретабельности предлагается отконтрастировать, удалив малозначимые связи между нейронами.

На основе данных идей формально изложим метод синтеза РБНС.

Этап 1. Инициализация. Задать допустимый порог ошибки ε, $0 \le \varepsilon << 1$, а также обучающую выборку прецедентов $<x, y>$. Определить $x_j^{\min}$, $x_j^{\max}$, $j=1, 2,..., N$. Положить $Q=K$. Сформировать Q эталонов классов, координаты центров которых определяются по формуле:

$$C_j^q = \frac{1}{S^q}\sum_{j=1}^{S}\left\{x_j^s \mid y^s = q\right\},\ q=1,2,...,K;\ j=1,2,...,N.$$

Этап 2. Контроль числа кластеров. Если $Q-1 < \left\lfloor \frac{NS-K}{N+K+1} \right\rfloor$, т.е. имеется возможность добавить еще один кластер при сохранении обобщающих свойств, тогда перейти к этапу 3, в противном случае – к этапу 4.

Этап 3. Выделение конфликтного набора экземпляров. Относительно имеющегося множества эталонов $C=\{C^q\}$, $C^q=\{C_j^q\}$, для экземпляров выборки определить расстояния до эталонов:

$$R(x^s, C^q) = R(C^q, x^s) = \sqrt{\sum_{j=1}^{N}(x_j^s - C_j^q)^2},\ s=1,2,...,S;$$

$$q=1,2,...,Q.$$

Определить расчетный номер класса для каждого экземпляра относительно сформированного множества эталонов:

$$y^s_{расч.} = \arg\min_{q=1,2,...,Q}\{R(x^s, C^q)\}.$$

Все экземпляры, для которых $y^s \ne y^s_{расч.}$, выделить в конфликтный набор:

$$X' = \bigcup_{s=1}^{S}\{< x^s, y^s > \mid y^s \ne y^s_{расч.}\}.$$

Для экземпляров конфликтного набора определить общее число экземпляров S', а также число экземпляров в q-м классе S'^q.

Этап 4. Расширение набора эталонов. Если достигнута требуемая точность ($S'/S<\varepsilon$), тогда перейти к этапу 5, в противном случае – положить $Q=Q+1$ и сформировать новый эталон для наиболее частотного класса в конфликтном наборе:

$$C_j^Q = \frac{1}{S'^q}\sum_{j=1}^{S'}\left\{x'^s_j \mid y'^s = q\right\},$$

после чего перейти к этапу 2.

Этап 5. Определить значение коэффициента, регулирующего величину каждого кластера:

$$\beta_q = \frac{S^q}{\max\limits_{\substack{q=1,2,..,Q;\\ p=q+1,...,Q}}\left\{\sqrt{\sum_{j=1}^{N}\left(C_j^q - C_j^p\right)^2}\right\}},\ q=1,2,...,Q.$$

Этап 6. Синтез структуры РБНС. Задать на первом (скрытом) слое РБНС Q нейронов, использующих в качестве весовых (постсинаптических функций) функции расстояния

$$\varphi^{(1,i)}(x^{(1,i)}, w^{(1,i)}) = w_0^{(1,i)}\sqrt{\sum_{j=1}^{N}(x_j^{(1,i)} - w_j^{(1,i)})^2},$$

где $w_0^{(1,i)}$ – параметр, регулирующий ширину радиально-базисной функции i-го нейрона первого слоя, а в качестве активационных функций – функции Гаусса: $\psi^{(1,i)}(\varphi^{(1,i)}) = \exp(-\varphi^{(1,i)})$.

На втором (выходном) слое РБНС задать число нейронов, равное размерности выходного вектора. Определить в качестве весовых функций нейронов второго слоя сети взвешенные суммы

$$\varphi^{(2,i)}(x^{(2,i)}, w^{(2,i)}) = \sum_{j=1}^{N_M} w_j^{(2,i)} x_j^{(2,i)} + w_0^{(2,i)},$$

где $w_0^{(1,i)}$ – порог, а в качестве активационных функций – линейные функции: $\psi^{(2,i)}(\varphi^{(2,i)}) = \varphi^{(2,i)}$.

Этап 6. Настройка параметров РБНС.

Занести в j-й вес q-го нейрона первого слоя РБНС значение j-й координаты центра q-го кластера: $w_j^{(1,q)} = C_j^q, j = 1, 2, ..., N, q=1, 2, ..., Q$. Занести в нулевые веса нейронов первого слоя параметры, регулирующие ширину кластеров: $w_0^{(1,q)} = \beta_q$, $q=1, 2, ..., Q$.

Определить значения весов нейронов второго слоя РБНС путем решения системы линейных алгебраических уравнений вида $y_i = \psi^{(1,i)}\left(\varphi^{(1,i)}(x)\right)w^{(2)}$, где $y_i = (\{y_i^s\})^T$ – вектор выходных значений, $w^{(2,i)}$ – вектор весов *i*-го выходного нейрона [9], либо на основе обратного распространения ошибки [8, 10, 11].

Этап 7. Оптимизация (контрастирование) РБНС. Выполнить распознавание обучающей выборки с помощью построенной РБНС и определить ошибку сети *E*. До тех пор, пока ошибка является приемлемой ($E \leq \varepsilon$) выполнять последовательно в цикле контрастирование весов второго слоя сети: среди весов второго слоя определить вес с наименьшим абсолютным значением, неравным нулю, положить его равным нулю, выполнить распознавание обучающей выборки, оценить ошибку распознавания *E*. В результате выполнения этих действий часть весов будет обнулена, т.е. можно считать, что будут удалены соответствующие связи, что позволит повысить интерпретабельность полученной модели, а также ее уровень обобщения по отношению к исходным данным и дообучение.

Предложенный метод позволяет автоматизировать процесс синтеза РБНС, сокращая зависимость от пользователя, а также обеспечивает поддержание приемлемых обобщающих свойств сети при приемлемой точности.

4 ЭКСПЕРИМЕНТЫ

Для проверки работоспособности и практической применимости предложенного метода он был программно реализован как дополнение к автоматизированной ситеме диагностирования [13].

Таблица 1 – Характеристики решавшихся задач

Задача	Характеристики задачи		
	N	*S*	*K*
Диагностирование лопаток газотурбинных авиадвигателей по спектрам свободных затухающих колебаний после ударного возбуждения	100	32	2
Прогнозирование прочности лопаток газотурбинных авиадвигателей после упрочняющей обработки	16	57	2
Диагностирование хронического обструктивного бронхита	28	205	2

Разработанная программа использовалась для построения диагностических моделей в задачах технического и медицинского диагностирования [13]: диагностирование лопаток газотурбинных авиадвигателей по спектрам свободных затухающих колебаний после ударного возбуждения, прогнозирование прочности лопаток газотурбинных авиадвигателей после упрочняющей обработки, диагностирование хронического обструктивного бронхита. Характеристики решавшихся задач приведены в табл. 1.

5 РЕЗУЛЬТАТЫ

В табл. 2. приведены результаты, полученные в ходе вычислительных экспериментов.

Для оценивания обобщающих свойств полученных моделей в таблице использован показатель:

$$F_g = \frac{NS}{N_w - N_{w=0}}.$$

Чем больше будет значение данного показателя, тем выше уровень обобщения сетью данных обучающей выборки, тем меньше и компактнее модель, тем она более интерпретабельная.

6 ОБСУЖДЕНИЕ

Как видно из табл. 1, предложенный метод характеризуется меньшими затратами времени на построение РБНС по сравнению с использованием четкого переборного кластерного анализа и по сравнению с отображением выборки в кластеры. При этом предложенный метод обеспечивает в среднем более высокую точность моделей и существенно более высокий уровень обобщения по сравнению с известными методами.

Более высокий уровень обобщения моделей, синтезированных на основе предложенного метода, обеспечивается тем, что кроме обобщения данных эталонами в процессе кластер-анализа, после синтеза сети ее веса контрастируются, а структура сети оптимизируется путем удаления малозначимых связей между нейронами. Это позволяет упростить сеть, снизив ее структурную и параметрическую сложность, а также повысить интерпретабельность нейромодели.

ВЫВОДЫ

В работе решена актуальная проблема автоматизации синтеза РБНС на основе набора прецедентов для принятия решений в диагностировании.

Научная новизна работы состоит в том, что впервые предложен метод синтеза РБНС, который, в отличие от

Таблица 2 – Характеристики методов синтеза РБНС

Задача	Методы синтеза РБНС								
	Прямое отображение выборки в кластеры, обратное распространение ошибки			Кластер-анализ, обратное распространение ошибки			Предложенный метод		
	$t_{об.}$, с	*E*	F_g	$t_{об.}$, с	*E*	F_g	$t_{об.}$, с	*E*	F_g
Диагностирование лопаток	1,3	$0{,}98\cdot10^{-6}$	3,48	4,9	$0{,}94\cdot10^{-6}$	4,05	3,73	$0{,}92\cdot10^{-6}$	4,63
Прогнозирование прочности лопаток	0,8	$0.96\cdot10^{-6}$	3,37	3,1	$0{,}89\cdot10^{-6}$	3,62	2,95	$0{,}83\cdot10^{-6}$	4,21
Диагностирование бронхита	1,6	$0{,}91\cdot10^{-6}$	4,07	6,2	$0{,}93\cdot10^{-6}$	5,15	4,96	$0{,}94\cdot10^{-6}$	6,88

известных методов, не требует задания пользователем числа кластеров, не имеет неопределенности выбора числа нейронов в первом слое и выбора начальных значений весов сети, стремится минимизировать размер сети, характеризуется приемлемым временем обучения, благодаря использованию процедуры оптимизации сети позволяет получать безызбыточные контрастные, интерпретабельные нейромодели.

Практическая значимость результатов работы заключается в том, что разработано программное обеспечение, реализующее предложенный метод, а также проведены эксперименты, подтвердившие работоспособность разработанного математического обеспечения и позволяющие рекомендовать его для использования на практике при решении задач построения диагностических моделей по прецедентам для автоматизации принятия решений в технической и биомедицинской диагностики.

Перспективы дальнейших исследований состоят в исследовании предложенного математического обеспечения на более широком наборе практических задач диагностики и распознавания образов.

БЛАГОДАРНОСТИ

Работа выполнена в рамках в рамках госбюджетных научно-исследовательских тем Запорожского национального технического университета «Интеллектуальные информационные технологии диагностирования и автоматической классификации» и «Интеллектуальные методы диагностирования систем управления удаленными техническими объектами» при поддержке международного проекта «Centers of Excellence for young RESearchers» программы «Темпус» Европейской Комиссии (№ 544137-TEMPUS-1-2013-1-SK-TEMPUS-JPHES).

СПИСОК ЛИТЕРАТУРЫ

1. Intelligent fault diagnosis and prognosis for engineering systems / [G. Vachtsevanos, F. Lewis, M. Roemer et al.]. – New Jersey: John Wiley & Sons, 2006. – 434 p.
2. Price C. Computer based diagnostic systems / C. Price. – London: Springer, 1999. – 136 p.
3. Дли М. И. Нечеткая логика и искусственные нейронные сети / М. И. Дли. – М. : Физматлит, 2003. – 225 с.
4. Рутковская Д. Нейронные сети, генетические алгоритмы и нечеткие системы / Д. Рутковская, М. Пилиньский, Л. Рутковский ; пер. с польск. И. Д. Рудинского. – М. : Горячая линия – Телеком, 2004. – 452 с.
5. Rao V. B. C++ neural networks and fuzzy logic / V. B. Rao. – Foster City : M&T Books, 1995. – 549 p.
6. Bishop C. Neural networks for pattern recognition / C. Bishop. – New York : Oxford University Press, 1995. – 482 p.
7. Круглов В. В. Искуственные нейронные сети. Теория и практика / В. В. Круглов, В. В. Борисов. – М. : Горячая линия – Телеком, 2001. – 382 с.
8. Осовский С. Нейронные сети для обработки информации / С. Осовский. – М. : Финансы и статистика, 2004. – 344 с.
9. Руденко О. Г. Штучні нейронні мережі / О. Г. Руденко, Є. В. Бодянський. – Харків : Компанія СМІТ, 2006. – 404 с.
10. Хайкин С. Нейронные сети: полный курс / С. Хайкин. – СПб : Вильямс, 2005. – 1104 с.
11. Субботін С. О. Нейронні мережі : навч. посіб. / С. О. Субботін, А. О. Олійник; за ред. С. О. Субботіна. – Запоріжжя : ЗНТУ, 2014. – 132 с.
12. Олійник А. О. Інтелектуальний аналіз даних : навчальний посібник / А. О. Олійник, С. О. Субботін, О. О. Олійник. – Запоріжжя : ЗНТУ, 2012. – 278 с.
13. Интеллектуальные информационные технологии проектирования автоматизированных систем диагностирования и распознавания образов : монография / [С. А. Субботин, Ан. А. Олейник, Е. А. Гофман и др.] ; под ред. С. А. Субботина. – Харьков : ООО «Компания Смит», 2012. – 317 с.

Статья поступила в редакцию 21.01.2016.

Субботін С. О.
Д-р техн. наук, професор, професор кафедри програмних засобів Запорізького національного технічного університету, Запоріжжя, Україна

МЕТОД СИНТЕЗУ ДІАГНОСТИЧНИХ МОДЕЛЕЙ НА ОСНОВІ РАДІАЛЬНО-БАЗИСНИХ НЕЙРОННИХ МЕРЕЖ З ПІДТРИМКОЮ УЗАГАЛЬНЮВАЛЬНИХ ВЛАСТИВОСТЕЙ

У роботі вирішено актуальну проблему автоматизації синтезу радіально-базисних нейронних мереж на основі набору прецедентів для прийняття рішень у діагностуванні. Запропоновано метод синтезу радіально-базисних нейронних мереж, що формує на початку по одному еталону класу, які за необхідності доповнює новими еталонами, формованими на основі помилково розпізнаних екземплярів, а далі оперує відстанями від екземплярів до еталонів кластерів. На основі отриманих координат еталонів далі в автоматичному режимі синтезується структура і налаштовуються параметри мережі, які додатково для підвищення узагальнювальних властивостей та інтерпретабельності піддається контрастуванню. Запропонований метод не вимагає задавання користувачем числа кластерів, не має невизначеності вибору кількості нейронів у першому шарі та вибору початкових значень ваг мережі, прагне мінімізувати розмір мережі, характеризується прийнятним часом навчання, завдяки використанню процедури оптимізації мережі дозволяє одержувати безнадлишкові контрастні, інтерпретабельні нейромоделі. Розроблено програмне забезпечення, яке реалізує запропонований метод, а також проведені експерименти, що підтвердили працездатність розробленого математичного забезпечення і дозволяють рекомендувати його для використання на практиці при вирішенні задач побудови діагностичних моделей за прецедентами для автоматизації прийняття рішень у технічній і біомедичній діагностиці.

Ключові слова: нейронна мережа, радіально-базисна мережа, навчання, синтез, діагностика.

Subbotin S. A.
Dr.Sc., Professor, Professor of department of software tools, Zaporizhzhya National Technical University, Zaporizhzhya, Ukraine

THE METHOD OF DIAGNOSTIC MODEL SYNTHESIS BASED ON RADIAL BASIS NEURAL NETWORKS WITH THE SUPPORT OF GENERALIZATION PROPERTIES

Urgent problemof automation of radial basis neural network synthesis based on a set of precedents for decision-making in the diagnosis is solved in the paper. The method for the synthesis of radial basis neural network is proposed. It forms at the beginning one class pattern, which, if necessary, supplemented with new patterns formed on the basis of wrongly recognized instances, and then operates with the distance from the instances to the patterns of the clusters. On the basis of the obtained pattern coordinates it further automatically synthesize structure and

adjust the weights of the network, which is further optimized to improve the generalizing and interpretability properties by weights contrasting. The proposed method does not require the user specify the number of clusters, has no uncertainty in selection the number of neurons in the first layer and in the choice of the initial values of the network weights, seeks to minimize the size of the network, and characterized by an acceptable time of learning through the use of network optimization procedure allows to obtain nonredundant, contrast, and interpretable neural models. The software implementing proposed method has been developed. The experiments confirming efficiency of developed software have been conducted. They allow to recommend the proposed method for use in practice in solving the problems of diagnostic model constructing by precedents to automate the decision-making in technical and biomedical diagnostics.

Keywords: neural network, radial base network, training, synthesis, diagnostics.

REFERENCES

1. Vachtsevanos G., Lewis F., Roemer M. et al Intelligent fault diagnosis and prognosis for engineering systems. New Jersey, John Wiley & Sons, 2006, 434 p.
2. Price C. Computer based diagnostic systems. London, Springer, 1999, 136 p.
3. Dli M. I. Nechotkaya logika i iskusstvennyye neyronnyye seti. Moscow, Fizmatlit, 2003, 225 p.
4. Rutkovskaya D., Pilin'skiy M., Rutkovskiy L.; per. s pol'sk. I. D. Rudinskogo Neyronnyye seti, geneticheskiye algoritmy i nechotkiye sistemy. Moscow, Goryachaya liniya, Telekom, 2004, 452 p.
5. Rao V. B. C++ neural networks and fuzzy logic. Foster City: M&T Books, 1995, 549 p.
6. Bishop C. Neural networks for pattern recognition. New York, Oxford University Press, 1995, 482 p.
7. Kruglov V. V., Borisov V. V. Iskusstvennyye neyronnyye seti. Teoriya i praktika. Moscow, Goryachaya liniya, Telekom, 2001, 382 p.
8. Osovskiy S. Neyronnyye seti dlya obrabotki informatsii. Moscow, Finansy i statistika, 2004, 344 p.
9. Rudenko O. H., Bodyanskyy YE. V. Shtuchni neyronni merezhi. Kharkiv, Kompaniya SMIT, 2006, 404 p.
10. Khaykin S. Neyronnyye seti: polnyy kurs. Sankt-Peterburg, Vil'yams, 2005, 1104 p.
11. Subbotin S. O., Oliynyk A. O.; za red. S. O. Subbotina. Neyronni merezhi : navch. posib. Zaporizhzhya, ZNTU, 2014, 132 p.
12. Oliynyk A. O., Subbotin S. O., Oliynyk O. O. Intelektualnyy analiz danykh : navchal'nyy posibnyk. Zaporizhzhya, ZNTU, 2012, 278 p.
13. Subbotin S. A., Oleynik An. A., Gofman Ye. A. i dr. pod red. S. A. Subbotina Intellektual'nyye informatsionnyye tekhnologii proyektirovaniya avtomatizirovannykh sistem diagnostirovaniya i raspoznavaniya obrazov : monografiya. Khar'kov, OOO «Kompaniya Smit», 2012, 317 p.

УДК 621.391

Штовба С. Д.[1], Галущак А. В.[2]

[1]Д-р техн. наук, професор, професор кафедри комп'ютерних систем управління, Вінницький національний технічний університет, Вінниця, Україна
[2]Асистент кафедри комп'ютерних систем управління, Вінницький національний технічний університет, Вінниця, Україна

КРИТЕРІЇ НАВЧАННЯ НЕЧІТКОГО КЛАСИФІКАТОРА НА ОСНОВІ ВІДСТАНІ МІЖ ГОЛОВНИМИ КОНКУРЕНТАМИ

Класифікація це віднесення об'єкта за деякими ознаками до одного з класів. До класифікації зводяться різноманітні задачі прийняття рішень в інженерії, економіці, медицині, соціології та в інших областях. В нечітких класифікаторах залежність «входи – вихід» описуються за допомогою лінгвістичних правил <Якщо – тоді>, антецеденти яких містять нечіткі терми «низький», «середній», «високий» тощо. Для підвищення безпомилковості нечіткий класифікатор навчають за експериментальними даними. В даній роботі запропоновано нові критерії навчання нечіткого класифікатора, які враховують різницю належностей нечіткого висновку лише до головних конкурентів. За правильної класифікації головним конкурентом прийнятого рішення є клас, що має другий за величиною ступінь належності. У випадку неправильної класифікації помилково прийняте рішення є головним конкурентом правильного класу. Проведені комп'ютерні експерименти із навчання нечіткого класифікатора для розпізнавання трьох сортів італійських вин засвідчили суттєву перевагу нових критеріїв. Серед нових критеріїв помірну перевагу має критерій на основі квадратичної відстані між головними конкурентами з штрафом за помилкове рішення. Нові критерії можуть застосовуватися не лише для навчання нечітких класифікаторів, але і для навчання деяких інших моделей, наприклад, нейронних мереж.

Ключові слова: класифікація, нечітка база знань, навчання, критерії навчання, головні конкуренти.

НОМЕНКЛАТУРА

$\widetilde{a}_{ij}$ – нечіткий терм, яким оцінюється ознака x_i в j-му правилі, $i=\overline{1,n}$, $j=\overline{1,k}$;

$\Delta_r(\mathbf{K})$ – мітка коректності класифікації r-го об'єкту за нечіткою моделлю з параметрами $\mathbf{K}$;

$\mu_j(X^*)$ – ступінь виконання j-го правила з бази знань;

$\mu_{l_s}(y^*)$ – ступінь належності вхідного вектору $\mathbf{X}^*$ до класу l_s;

$\mu_j(x_i^*)$ – ступінь належності значення x_i^* нечіткому терму $\widetilde{a}_{ij}$;

$\mu_{l_s}(y_r)$ – ступінь належності r-го об'єкту навчальної вибірки до класу l_s;

$\mu_{l_s}(\mathbf{K},\mathbf{X}_r)$ – ступінь належності висновку до класу l_s, який розраховано за нечіткою моделлю з параметрами $\mathbf{K}$ для вхідного вектора $\mathbf{X}_r$;

$\mu_{win}(\mathbf{X}_r)$ – ступінь належності вхідного вектора $\mathbf{X}_r$ класу переможцю;

$\mu_{vicewin}(\mathbf{X}_r)$ – ступінь належності вхідного вектора $\mathbf{X}_r$ до класу з другим рангом;

$\mu(x)$ – функція належності для змінної x;

$\wedge$ – t-норма;

b – координата максимуму гаусової функції належності;

c – коефіцієнт концентрації гаусової функції належності.

$Crit_1$ – перший критерій навчання;

$Crit_2$ – другий критерій навчання;

$Crit_3$ – третій критерій навчання;

$Crit_4$ – четвертий критерій навчання;

$Crit_5$ – п'ятий критерій навчання;

$D_r(\mathbf{K})$ – відстань між бажаною та дійсною вихідними нечіткими множинами при класифікації r-го об'єкту на основі нечіткої бази знань з параметрами ;

d_j – категоріальне значення консеквента j-го правила;

F – нечітка модель;

k – кількість правил;

$\mathbf{K}$ – вектор параметрів нечіткої бази знань, які налаштовуються;

l_m – мітка класа з номером m;

M – обсяг навчальної вибірки;

p – штрафний коефіцієнт;

$\mathbf{P}$ – вектор параметрів функцій належності термів бази знань;

smax – операція знаходження другого за величиною елемента множини;

T – обсяг тестової вибірки;

Vicewin – клас з другим за величиною ступенем належності;

$\mathbf{W}$ – вектор вагових коефіцієнтів правил нечіткої бази знань;

win – клас-переможець з максимальним ступенем належності;

w_j – ваговий коефіцієнт j-го правила;

$\mathbf{X}$ – вектор вхідних атрибутів;

$\mathbf{X}_r$ – вхідні атрибути r-го об'єкту;

y – результат класифікації;

y_r – клас r-го об'єкту.

ВСТУП

Класифікація це віднесення об'єкта за деякими ознаками до одного з класів. До класифікації зводяться різноманітні задачі прийняття рішень в інженерії, економіці, медицині, соціології, політиці, спорті та в інших областях.

DOI 10.15588/1607-3274-2016-2-9

Останнім часом все більш популярними стають нечіткі класифікатори, тобто класифікатори, в процесі функціонування або навчання яких використовуються нечіткі множини [1]. Сьогодні переважно застосовуються нечіткі класифікатори на основі логічного виведення по базі продукційних правил, антецеденти яких містять нечіткі терми «низький», «середній», «високий» тощо. Кожне правило задає область вхідних атрибутів, в межах якої об'єкти належать одному класу. Границі цих областей нечіткі, тому один і той же об'єкт може одночасно належати декільком класам, але з різним ступенем.

Для підвищення безпомилковості нечіткий класифікатор навчають по експериментальним даним. Для цього змінюють його параметри, щоб мінімізувати відстань між експериментальними даними та результатами нечіткого виведення. Цю відстань, яку назвемо критерієм навчання, можна визначити у різний спосіб [2]. Критерії навчання впливають на зміну параметрів нечіткого класифікатора на кожній ітерації алгоритму налаштування і, відповідно призводять до різних результатів. Тому, метою статі є знаходження такого критерія, використання якого забезпечує найкращу результативність навчання нечіткого класифікатора.

1 ПОСТАНОВКА ЗАДАЧІ

Нечіткий класифікатор являє собою відображення $\mathbf{X} = (x_1, x_2, \ldots, x_n) \xrightarrow{F} y \in \{l_1, l_2, \ldots, l_m\}$, яке реалізується логічним виведенням по базі нечітких правил. Навчання нечіткої бази знань здійснимо за експериментальними даними. Відповідно до принципу зовнішнього доповнення [3] експериментальні дані розіб'ємо на навчальну вибірку з M рядків «входи – вихід» та тестову вибірку з T таких рядків:

$$(\mathbf{X}_r, y_r),\ r = \overline{1, M}, \tag{1}$$

$$(\mathbf{X}_r, y_r),\ r = \overline{1, T}\ \ r = \overline{1, T}, \tag{2}$$

де $\mathbf{X}_r = (x_{r1}, x_{r2}, \ldots, x_{rn});\ y_r \in \{l_1, l_2, \ldots, l_m\}$.

Задача полягає у знаходженні такого критерію навчання *Crit*, застосування якого під час налаштування нечіткої бази знань *F* на вибірці (1) забезпечує мінімальну частоту помилок класифікації на тестовій вибірці (2).

2 ОГЛЯД ЛІТЕРАТУРИ

Класифікація на основі нечітких множин була запропонована 50 років тому в статті [4]. Перші роботи із ідентифікації залежностей за допомогою нечітких класифікаторів з'явилися в середині 90-х років. Вперше задача структурної ідентифікації шляхом відбору правил нечіткого класифікатора за показниками безпомилковості та складності бази знань розглянута в роботі [5]. Пізніше в [6] на основі критерію безпомилковості здійснили і параметричну ідентифікацію шляхом налаштування ваг правил. При зміні лише ваги правил можна отримати швидкі алгоритми навчання [7], але вони не гарантують високої безпомилковості нечіткого класифікатора. Для підвищення безпомилковості нечіткого класифікатора на етапі параметричної ідентифікації налаштовують не лише ваги правил, а й функції належності. Першими роботами з налаштування функцій належності нечіткого класифікатора є статті [8, 9, 10]. В них критерієм навчання виступає квадратична нев'язка між двома нечіткими множинами – бажаними та реальними результатами класифікації. В сучасних роботах навчання нечітких класифікаторів здійснюють саме за цим критерієм (дивись, наприклад, [11]).

В роботі [7] запропоновано новий критерій навчання, який поєднує частоту помилок та квадратичну нев'язку між нечіткими бажаними та дійсними результатами логічного виведення. У випадку помилкової класифікація квадратична нев'язка зважується штрафним коефіцієнтом. Комп'ютерні експерименти в роботах [2, 7] показали перевагу цього критерію навчання над частотою помилок та над квадратичною нев'язкою. Але ця перевага є незначною і навчання нечітких класифікаторів не завжди є результативним. Тому виникає зацікавленість у нових критеріях, навчання за якими забезпечує кращу безпомилковість нечіткого класифікатора.

Зміна параметрів нечіткого класифікатора може відбуватися за логічним висновком для одного об'єкту навчальної вибірки або за результатами логічного виведення за усією навчальною вибіркою [6]. Ми розглядаємо пакетний режим навчання, коли параметри класифікатора модифікуються за логічними висновками за усією вибіркою (1).

Ґрунтуючись на [2] базу правил нечіткого класифікатора запишемо так:

$$\text{Якщо } (x_1 = \tilde{a}_{1j} \text{ та } x_2 = \tilde{a}_{2j} \text{ та } \ldots \text{ та } x_n = \tilde{a}_{nj} \text{ з вагою } w_j),$$

$$\text{тоді } y = d_j,\ \ j = \overline{1, k}, \tag{3}$$

де $w_j \in [0, 1]$, $j = \overline{1, k}$; $d_j \in \{l_1, l_2, \ldots, l_m\}$.

Класифікація об'єкта з атрибутами $\mathbf{X}^* = (x_1^*, x_2^*, \ldots, x_n^*)$ здійснюється таким чином. Спочатку розраховується ступінь виконання *j*-го правила з бази (3):

$$\mu_j(X^*) = w_j \cdot \left(\mu_j(x_1^*) \wedge \mu_j(x_2^*) \wedge \ldots \wedge \mu_j(x_n^*)\right),$$

$$j = \overline{1, k}. \tag{4}$$

Ступінь належності $\mathbf{X}^*$ до класів $l_1, l_2, \ldots, l_m$ розраховується так:

$$\mu_{l_s}(y^*) = \max_{\forall j:\, d_j = l_s} \left(\mu_j(X^*)\right),\ s = \overline{1, m}. \tag{5}$$

Нечітким рішенням задачі класифікації буде нечітка множина

$$\tilde{y}^* = \left(\frac{\mu_{l_1}(y^*)}{l_1}, \frac{\mu_{l_2}(y^*)}{l_2}, \ldots, \frac{\mu_{l_m}(y^*)}{l_m}\right). \tag{6}$$

Кінцевим результатом виведення оберемо ядро нечіткої множини (6), тобто клас з максимальним ступенем належності:

$$y^* = \arg \max_{\{l_1, l_2, \ldots, l_m\}} \max_{s = \overline{1, m}} \left(\mu_{l_s}(y^*)\right).$$

Навчання нечіткого класифікатора полягає в знаходженні такого вектора $\mathbf{K} = (\mathbf{P}, \mathbf{W})$, який мінімізує частоту помилок класифікації на тестовій вибірці. При цьому для зміни координат вектора $\mathbf{K}$ використовується лише навчальна вибірка (1). Координати вектора $\mathbf{K}$ змінюємо на кожній ітерації алгоритму оптимізації за відстанню між результатами логічного виведення та експериментальними даними з вибірки (1). Цю відстань, яку назвемо критерієм навчання, можна визначити у різний спосіб.

Сьогодні найвідомішими є 3 критерія навчання нечіткого класифікатора: на основі частоти помилок [6], на основі відстані між нечіткими результатами класифікації і експериментальними даними [8–11], та їх комбінації [2, 7].

Критерій 1 – частота помилок класифікації:

$$Crit_1 = \frac{1}{M} \sum_{r=\overline{1, M}} \Delta_r(\mathbf{K}), \qquad (7)$$

де $\Delta_r(\mathbf{K}) = \begin{cases} 1, & якщо \;\; y_r \neq F(\mathbf{K}, \mathbf{X}_r) \\ 0, & якщо \;\; y_r = F(\mathbf{K}, \mathbf{X}_r) \end{cases}$.

Переваги критерію полягають в його простоті та ясній змістовній інтерпретації. Але цільова функція в задачі оптимізації за цим критерієм приймає дискретні значення, що ускладнює застосування швидких градієнтних методів оптимізації, особливо за малих вибірок даних.

Критерій 2 – квадратична нев'язка між двома нечіткими множинами – бажаними та реальними результатами класифікації. Для її розрахунку значення вихідної змінної y в навчальній вибірці фаззифікують таким чином:

$$\left.\begin{array}{ll} \widetilde{y} = \left(\frac{1}{l_1}, \frac{0}{l_2}, ..., \frac{0}{l_m}\right), & якщо \;\; y = l_1 \\ \widetilde{y} = \left(\frac{0}{l_1}, \frac{1}{l_2}, ..., \frac{0}{l_m}\right), & якщо \;\; y = l_2 \\ \vdots & \\ \widetilde{y} = \left(\frac{0}{l_1}, \frac{0}{l_2}, ..., \frac{1}{l_m}\right), & якщо \;\; y = l_m \end{array}\right\}. \qquad (8)$$

Критерій навчання враховує відстань між логічним висновком у формі нечіткої множини (6) та бажаним нечітким значенням вихідної змінної (8):

$$Crit_2 = \sum_{r=\overline{1, M}} D_r(\mathbf{K}). \qquad (9)$$

Для розрахунку $D_r(\mathbf{K})$ використовується евклідова метрика:

$$D_r(\mathbf{K}) = \sum_{s=\overline{1, m}} \left(\mu_{l_s}(y_r) - \mu_{l_s}(\mathbf{K}, \mathbf{X}_r)\right)^2, \qquad (10)$$

Перевага критерію $Crit_2$ полягає в урахуванні міри впевненості в прийнятому рішенні на основі ступенів належності об'єкту різним класам. В критерії $Crit_1$ вважається, що результат класифікації об'єкту є абсолютно достовірним, тобто неважливо наскільки ступінь належності у рішення більший, ніж у інших альтернатив – на 0,0001 чи на 1. Крім того, цільова функція в задачі навчання за критерієм (9) не має довгих плато, тому вона придатна до оптимізації градієнтними методами. Але близькі до границь класів об'єкти вносять майже однаковий вклад в критерій навчання (9) як за правильної, так і за помилкової класифікації, тому навчання може бути нерезультативним.

Критерій 3 – квадратична нев'язка між нечіткими бажаними та реальними результатами класифікації з додатковим штрафом за помилкове рішення. Ідея полягає в збільшенні відстані D для помилково класифікованих об'єктів:

$$Crit_3 = \sum_{r=\overline{1, M}} \left(\Delta_r(\mathbf{K}) \cdot p + 1\right) \cdot D_r(\mathbf{K}), \qquad (11)$$

де $p > 0$.

3 МАТЕРІАЛИ І МЕТОДИ

Нижче пропонуються нові критерії навчання нечіткого класифікатора, які враховують різницю належностей нечіткого висновку лише до головних конкурентів.

Критерій 4 – відстань між головними конкурентами з штрафом за помилкове рішення. Ідея цього нового критерію полягає у врахуванні різниці належностей нечіткого висновку лише до головних конкурентів. За алгоритмом виведення рішенням обирається клас з максимальним ступенем належності. Позначимо цей клас-переможець через *win*. У випадку правильної класифікації головним конкурентом прийнятого рішення є *vicewin* – клас з другим за величиною ступенем належності. Чим більша різниця між належністю до класів *win* та *vicewin*, тим більша впевненість у логічному висновку, і тим далі об'єкт знаходиться від границі розділу класів.

Для r-го об'єкту з вибірки (1) $\mu_{win}(\mathbf{X}_r) = \max\limits_{s=\overline{1, m}} \left(\mu_{l_s}(\mathbf{X}_r)\right)$ та $\mu_{vicewin}(\mathbf{X}_r) = \operatorname{smax}\limits_{s=\overline{1, m}} \left(\mu_{l_s}(\mathbf{X}_r)\right)$. Відповідно, різниця між головними конкурентами дорівнює $\mu_{win}(\mathbf{X}_r) - \mu_{vicewin}(\mathbf{X}_r)$.

В критерії навчання враховуватимемо відносні показники, розділивши різницю на ступінь належності класу-переможцю. За правильної класифікації відносна різниця дорівнює $D_r^1 = \dfrac{\mu_{win}(\mathbf{X}_r) - \mu_{vicewin}(\mathbf{X}_r)}{\mu_{win}(\mathbf{X}_r)}$, а за неправильної – $D_r^0 = \dfrac{\mu_{win}(\mathbf{X}_r) - \mu_{y_r}(\mathbf{X}_r)}{\mu_{win}(\mathbf{X}_r)}$. Крім того, аналогічно до (11), за помилкової класифікації зважимо різницю штрафним коефіцієнтом $p \geq 1$. Математично критерій навчання запишемо таким чином:

$$Crit_4 = p \cdot \sum_{\substack{y_r \neq F(\mathbf{K}, \mathbf{X}_r) \\ r=\overline{1, M}}} D_r^0(\mathbf{K}) - \sum_{\substack{y_r = F(\mathbf{K}, \mathbf{X}_r) \\ r=\overline{1, M}}} D_r^1(\mathbf{K}). \qquad (12)$$

Критерій 5 – квадратична відстань між головними конкурентами з штрафом за помилкове рішення. Цей

критерій є модифікацією попереднього. Відмінність полягає у використанні не відносних різниць, а їх квадратів:

$$Crit_5 = p \cdot \sum_{\substack{y_r \neq F(\mathbf{K}, \mathbf{X}_r) \\ r=\overline{1,M}}} D_r^0(\mathbf{K})^2 - \sum_{\substack{y_r = F(\mathbf{K}, \mathbf{X}_r) \\ r=\overline{1,M}}} D_r^1(\mathbf{K})^2 .$$

Піднесення до квадрату в $Crit_5$ дозволяє, як і в методі найменших квадратів, збільшити важливість великих різниць – тобто грубих промахів.

4 ЕКСПЕРИМЕНТИ

Метою експериментів є виявлення критерію навчання, який забезпечує найкращу безпомилковість. Розглядається тестова задача Wine Dataset з UCI Machine Learning Repository. Вона полягає у виявленні сорту винограду (y), з якого виготовлено вино. База даних містить результати лабораторних аналізів по 13-ти показникам 178 зразків італійських вин, виготовлених в одному регіоні. Для кожного зразка вказано, з якого із з трьох сортів винограду виготовлено вино.

Навчальну вибірку сформуємо з рядків бази даних з граничними значеннями кожного із 13 атрибутів. Додатково в навчальну вибірку включимо всі непарні рядки бази даних. Всі інші дані занесемо в тестову вибірку. В результаті отримаємо навчальну вибірку з 100 рядків і тестову – з 78. Побудуємо нечіткий класифікатор вин за трьома ознаками: x_7 – flavanoids, x_{10} – color intensity та x_{13} – proline. Експерименти проведемо для нечіткого класифікатора з базою знань з табл. 1. Нечіткі терми задамо гаусовою функцією належності:

$$\mu(x) = \exp\left(-\frac{(x-b)^2}{2c^2}\right),$$

де $c > 0$.

Параметри функцій належності початкового нечіткого класифікатора наведено в табл. 2.

Таблиця 1 – Нечітка база знань класифікатора вин [6]

№	x_7	x_{10}	x_{13}	y
1	–	–	Високий	Сорт 1
2	Високий	Високий	Середній	Сорт 1
3	–	Низький	Низький	Сорт 2
4	Низький	Низький	Середній	Сорт 2
5	Низький	Високий	–	Сорт 3

Таблиця 2 – Початкові параметри функцій належності термів нечіткого класифікатора

Вхідна змінна	Терм	Параметри	
		b	c
x_7	Низький	2	0,34
	Високий	2	5,08
x_{10}	Низький	6	1,28
	Високий	6	13
x_{13}	Низький	3	2,78
	Середній	3	10
	Високий	3	16,8

Для кожного критерію проведемо 1000 експериментів із навчання нечіткої бази знань на основі квазіньютоніського алгоритму. Після навчання кожний класифікатор перевіримо на тестовій вибірці за частотою помилок (критерій $Crit_1$). Під час навчання налаштуємо вагові коефіцієнти перших чотирьох правил. Достовірність п'ятого правила не викликає сумнівів, тому його ваговий коефіцієнт залишимо рівним 1. Налаштуємо також коефіцієнти концентрації (c) функції належності кожного нечіткого терма. Для збереження інтерпретуємості бази знань згідно до [11] налаштуємо координати максимумів (b) функцій належності лише не крайніх термів. В базі знань 1 некрайній терм – «Середній», координату максимуму якого і будемо змінювати. Таким чином, загальна кількість параметрів для налаштування становить $4+7+1=12$. Початкові точки для навчання оберемо випадково – для вагових коефіцієнтів правил з діапазону $[0, 1]$, а для параметрів функцій належності в межах $\pm 30\%$ від значень з табл. 2.

Проведемо 2 серії експериментів. Першу серію для нечіткого класифікатора з реалізацією t-норми мінімумом (min), а другу – з реалізацією t-норми добутком (prod). В експериментах з критеріями $Crit_3$, $Crit_4$ та $Crit_5$ спочатку визначимо прийнятний рівень штрафного коефіцієнту. Для цього проведемо по 200 оціночних експериментів для $p = 1, 3, ..., 9$. Результати цих експериментів наведено в табл. 3. В ній напівжирним шрифтом виділено за яких значень штрафного коефіцієнту навчання відбувається краще. Саме за таких значень штрафного коефіцієнту проведемо решту експериментів.

Таблиця 3 – Вплив штрафного коефіцієнта в критерії навчання на безпомилковість класифікатора на тестовій вибірці (статистика 200 експериментів)

t-норма	Критерій	Середня безпомилковість				
		$p=1$	$p=3$	$p=5$	$p=7$	$p=9$
min	$Crit_3$	**0,0856**	0,0915	0,0912	0,0910	0,0960
	$Crit_4$	0,0732	**0,0640**	0,0662	0,0736	0,0770
	$Crit_5$	0,0690	0,0604	**0,0602**	0,0688	0,0657
prod	$Crit_3$	0,0727	0,0656	0,0665	0,0681	**0,0649**
	$Crit_4$	0,0565	0,0512	**0,0493**	0,0499	0,0527
	$Crit_5$	0,0448	**0,0443**	0,0456	0,0486	0,0446

5 РЕЗУЛЬТАТИ

Результати експериментів наведено на рис. 1–3 та в табл. 4.

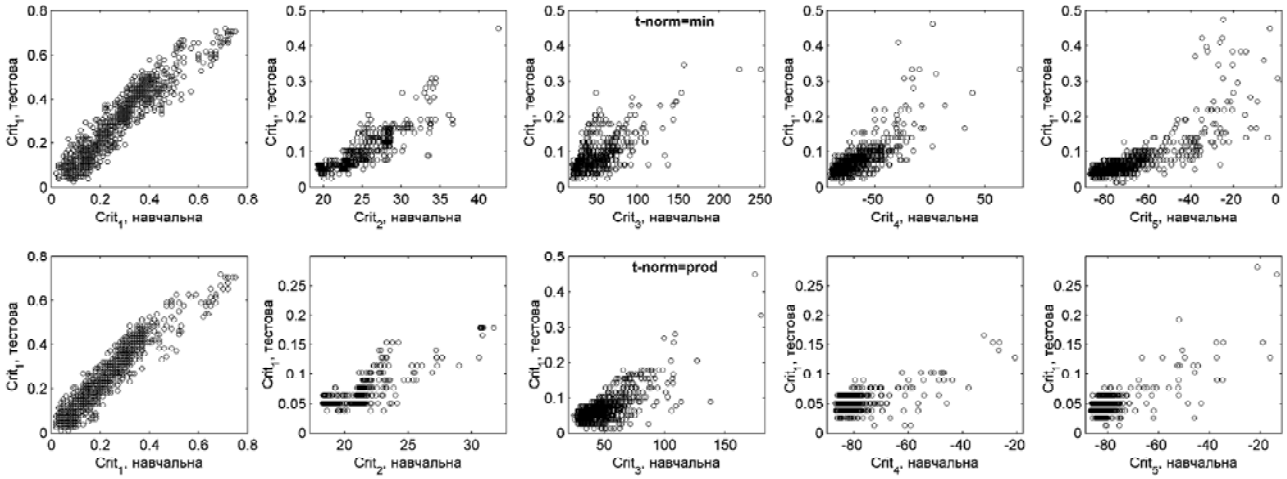


Рисунок 1 – Розподіл результатів навчання нечіткого класифікатора (статистика 1000 експериментів)

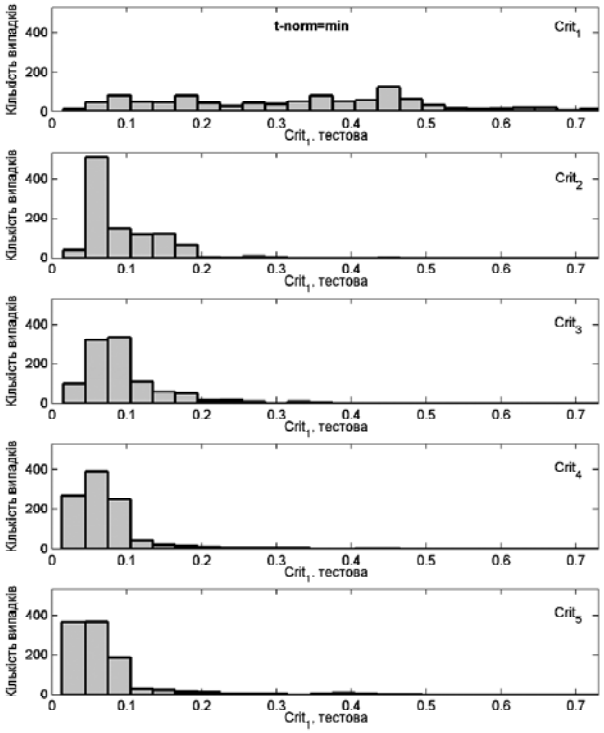


Рисунок 2 – Розподіл якості навчання класифікатора, *t*-норму якого реалізовано мінімумом

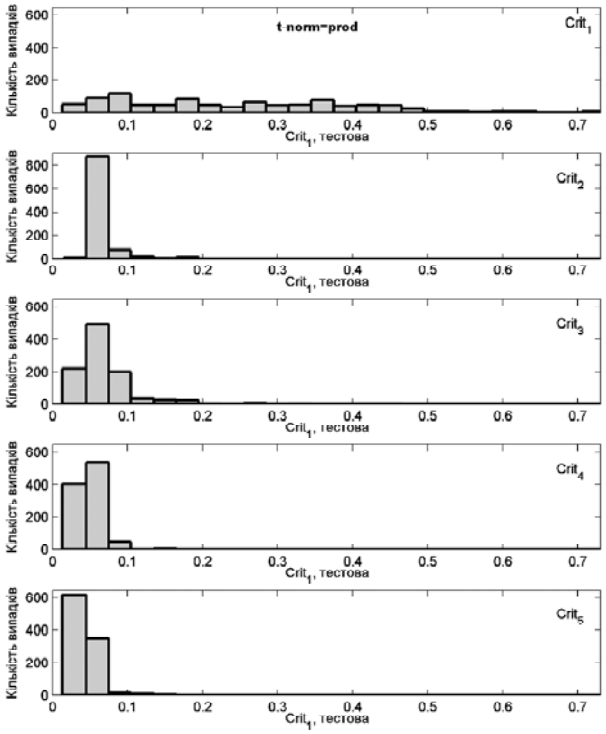


Рисунок 3 – Розподіл якості навчання класифікатора, *t*-норму якого реалізовано добутком

Таблиця 4 – Статистика навчання нечітких класифікаторів (напівжирним виділено найкращі результати)

t-норма	Критерій навчання	Частота помилок ($Crit_1$) на тестовій вибірці			
		мінімальне	середнє	медіанне	максимальне
min	$Crit_1$	0,0256	0,3194	0,3333	0,7179
	$Crit_2$	0,0256	0,0900	0,0641	0,4487
	$Crit_3$	0,0256	0,0921	0,0769	0,3462
	$Crit_4$	0,0128	**0,0680**	0,0513	0,4615
	$Crit_5$	0,0128	**0,0680**	0,0513	0,4744
prod	$Crit_1$	0,0128	0,2459	0,2308	0,7179
	$Crit_2$	0,0385	0,0631	0,0641	0,1795
	$Crit_3$	0,0128	0,0663	0,0641	0,4487
	$Crit_4$	0,0128	0,0503	0,0513	0,1667
	$Crit_5$	0,0128	**0,0454**	0,0385	0,2821

6 ОБГОВОРЕННЯ

Результати експериментів на рис. 1 вказують на корельованість значень критеріїв $Crit_1$ – $Crit_5$ на навчальній вибірці з частотою помилок на тестовій вибірці. Відповідно ці критерії можна застосовувати для навчання нечіткого класифікатора. Щодо результативності навчання (див. табл. 4 та рис. 2, 3), то вона суттєво краща при використанні нових критеріїв $Crit_4$ та $Crit_5$. Нові критерії забезпечують кращу безпомилковість як в середньому, так і за кількістю найкращих випадків навчання. Серед нових критеріїв помірну перевагу має $Crit_5$.

ВИСНОВКИ

У роботі запропоновано нові критерії навчання нечіткого класифікатора, які враховують різницю належностей нечіткого висновку лише до головних конкурентів. За правильної класифікації головним конкурентом прийнятого рішення є клас з другим за величиною ступенем належності. У випадку неправильної класифікації помилково прийняте рішення є головним конкурентом правильного класу.

Проведені нами експерименти із навчання нечіткого класифікатора для UCI-задачі із розпізнавання італійських вин засвідчили суттєву перевагу нових критеріїв. Серед нових критеріїв помірну перевагу має критерій на основі квадратичної відстані між головними конкурентами з штрафом за помилкове рішення. Відповідно, мету дослідження досягнуто – встановлено, що для налаштування нечітких класифікаторів доцільно використовувати нові критерії навчання на основі різниці належності рішення до головних конкурентів. Нові критерії можуть застосовуватися не лише для настроювання нечітких класифікаторів, але і для навчання деяких інших моделей, наприклад, нейронних мереж.

ПОДЯКИ

Публікація містить результати досліджень, проведених при грантовій підтримці Державного фонду фундаментальних досліджень за конкурсним проектом №62/201–2015.

СПИСОК ЛІТЕРАТУРИ

1. Kuncheva L. I. Fuzzy classifier design / L. I. Kuncheva //Studies in Fuzziness and Soft Computing. – Berlin – Heidelberg: Springer-Verlag, 2000. – Vol. 49. – 314 p.
2. Shtovba S. Analyzing the criteria for fuzzy classifier learning / S. Shtovba, O. Pankevich, A. Nagorna // Automatic Control and Computer Sciences. – 2015. – Vol. 49, № 3. – P. 123–132.
3. Madala H. R. Inductive learning algorithms for complex systems modeling / H. R. Madala, A. G. Ivakhnenko. – Boca Raton : CRC Press, 1994. – 368 p.
4. Bellman R. Abstraction and pattern classification / R. Bellman, R. Kalaba, L. Zadeh // Journal of Mathematical Analysis and Applications. – 1966. – Vol. 13, № 1. – P. 1–7.
5. Construction of fuzzy classification systems with rectangular fuzzy rules using genetic algorithms / [Ishibuchi H., Nozaki K., Yamomoto N., Tanaka H.] // Fuzzy sets and systems. – 1994. – Vol. 65, № 2. – P. 237–253.
6. Ishibuchi H. Classification and modeling with linguistic information granules: advanced approaches advanced approaches to linguistic data mining / H. Ishibuchi, T. Nakashima, M. Nii. – Berlin-Heidelberg : Springer-Verlag, 2005. – 307 p.
7. Штовба С. Д. Порівняння критеріїв навчання нечіткого класифікатора / С. Д. Штовба // Вісник Вінницького політехнічного інституту. – 2007. – № 6. – С. 84–91.
8. Abe S. Tuning of a fuzzy classifier derived from data / S. Abe, M. S. Lan, R. Thawonmas // International Journal of Approximate Reasoning. – 1996. – Vol. 14. – P. 1–24.
9. Nauck D. A neuro-fuzzy method to learn fuzzy classification rules from data / D. Nauck, R. Kruse // Fuzzy Sets and Systems. – 1997. – Vol. 89, № 3. – P. 277–288.
10. Rotshtein A. P. Design and Tuning of Fuzzy If – Then Rules for Automatic Classification / A. P. Rotshtein, D. I. Katelnikov // Proc. of NAFIPS'98 – International Conf. «Annual Meeting of North American Fuzzy Information Processing Society», Tampa, USA, 1998. – P. 50–55.
11. Rudzicski F. A multi-objective genetic optimization of interpretability-oriented fuzzy rule-based classifiers / F. Rudzicski // Applied Soft Computing. – 2016. – Vol. 38. – P. 118–133.
12. Shtovba S. Ensuring accuracy and transparency of Mamdani fuzzy model in learning by experimental data / S. Shtovba // Journal of Automation and Information Sciences. – 2007. – Vol. 39, № 8. – P. 39–52.

Стаття надійшла до редакції 21.12.2015.
Після доробки 26.01.2016.

Штовба С. Д.[1], Галущак А. В.[2]

[1]Д-р техн. наук, профессор, профессор кафедры компьютерных систем управления, Винницкий национальный технический университет, Винница, Украина

[2]Ассистент кафедры компьютерных систем управления, Винницкий национальный технический университет, Винница, Украина

ОБУЧЕНИЕ НЕЧЕТКОГО КЛАССИФИКАТОРА НА ОСНОВЕ РАССТОЯНИЯ МЕЖДУ ГЛАВНЫМИ КОНКУРЕНТАМИ

Классификация – это отнесение объекта по некоторым признакам к одному из классов. К классификации сводятся разнообразные задачи принятия решений в инженерии, экономике, медицине, социологии и других областях. В нечетких классификаторах зависимость «входы – выход» описываются с помощью лингвистических правил <Если – то>, антецеденты которых содержат нечеткие термы «низкий», «средний», «высокий» и т. п. Для повышения безошибочности нечеткий классификатор обучают по экспериментальным данным. В данной работе предложены новые критерии обучения нечеткого классификатора, которые учитывают разницы принадлежностей нечеткого вывода только к главным конкурентам. При правильной классификации главным конкурентом принятого решения является класс со второй по величине степенью принадлежности. В случае неправильной классификации ошибочно принятое решение является главным конкурентом правильного класса. Компьютерные эксперименты по обучению нечеткого классификатора для распознавания трех сортов итальянских вин показали существенное преимущество новых критериев. Среди новых критериев обучения небольшое преимущество имеет критерий в форме квадратичного расстояния между главными конкурентами со штрафом за ошибочное решение. Новые критерии могут применяться не только для обучения нечетких классификаторов, но и для обучения других моделей, например, нейронных сетей.

Ключевые слова: классификация, нечеткая база знаний, обучение, критерии обучения, главные конкуренты.

Shtovba S. D.[1], Galushchak A. V.[2]
[1]Prof., Dr.Sc., Vinnytsia National Technical University, Vinnytsia, Ukraine
[2]Assistant, Vinnytsia National Technical University, Vinnytsia, Ukraine

FUZZY CLASSIFIER LEARNING BASED ON DISTANCE BETWEEN THE MAIN COMPETITORS

The classification problem is the assignment an object with certain features to one of classes. Various engineering, management, economic, political, medical, sport, and other problems are reduced to classification. In fuzzy classifiers «inputs – output» relation is described by linguistic <If – then> rules. Antecedents of these rules contain fuzzy terms «low», «average», «high» etc. To increase the correctness it is necessary to tune the fuzzy classifier on experimental data. The new criteria for fuzzy classifier learning that take into account the difference of membership degrees to the main competitors only are proposed. When the classification is correct, the main competitor of the decision is the class with the second largest membership degree. In cases of misclassification the wrong decision is the main competitor to the correct class. Computer experiments with learning the fuzzy classifier of 3 kinds of Italian wines recognition showed a significant advantage of the new criteria. Among new learning criteria the criterion in the form of squared distance between main competitors with the penalty for wrong decision has minor advantage. New criteria can be used not only for tuning fuzzy classifiers but for tuning some other models, such as neural networks.

Keywords: classification, fuzzy knowledge base, tuning, learning criteria, main competitors.

REFERENCES

1. Kuncheva L. I. Fuzzy classifier design, *Studies in Fuzziness and Soft Computing*. Berlin, Heidelberg, Springer-Verlag, 2000, Vol. 49, 314 p.
2. Shtovba S. Pankevich O., Nagorna A. Analyzing the criteria for fuzzy classifier learning, *Automatic Control and Computer Sciences*, 2015, Vol. 49, No. 3, pp. 123–132.
3. Madala H. R., Ivakhnenko A. G. Inductive learning algorithms for complex systems modeling. Boca Raton, CRC Press, 1994, 368 p.
4. Bellman R., Kalaba R., Zadeh L. Abstraction and pattern classification, *Journal of Mathematical Analysis and Applications*, 1966, Vol. 13, No. 1, pp. 1–7.
5. Ishibuchi H., Nozaki K., Yamomoto N., Tanaka H. Construction of fuzzy classification systems with rectangular fuzzy rules using genetic algorithms, *Fuzzy sets and systems,* 1994, Vol. 65, No. 2, pp. 237–253.
6. Ishibuchi H., Nakashima T., Nii M. Classification and modeling with linguistic information granules: advanced approaches advanced approaches to linguistic data mining. Ishibuchi, Berlin-Heidelberg, Springer-Verlag, 2005, 307 p.
7. Shtovba S. D. Porivnjannja kriteriïv navchannja nechitkogo klasifikatora, *Visnik Vinnic'kogo politehnichnogo institutu*, 2007, No. 6, pp. 84–91.
8. Abe S., Lan M. S., Thawonmas R. Tuning of a fuzzy classifier derived from data, *International Journal of Approximate Reasoning*, 1996, Vol. 14, pp. 1–24.
9. Nauck D., Kruse R. A neuro-fuzzy method to learn fuzzy classification rules from data, *Fuzzy Sets and Systems*, 1997, Vol. 89, No. 3, pp. 277–288.
10. Rotshtein A. P., Katelnikov D. I. Design and Tuning of Fuzzy If – Then Rules for Automatic Classification, *Proc. of NAFIPS'98 – International Conf. «Annual Meeting of North American Fuzzy Information Processing Society»*. Tampa, USA, 1998, pp. 50–55.
11. Rudzicski F. A multi-objective genetic optimization of interpretability-oriented fuzzy rule-based classifiers, *Applied Soft Computing*, 2016, Vol. 38, pp. 118–133.
12. Shtovba S. Ensuring accuracy and transparency of Mamdani fuzzy model in learning by experimental data, *Journal of Automation and Information Sciences*, 2007, Vol. 39, No. 8, pp. 39–52.

ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

ПРОГРЕССИВНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

PROGRESSIVE INFORMATION TECHNOLOGIES

УДК 004.056.55

Гальченко А. В.[1], Козіна Г. Л.[2]

[1]Ст. гр. РТ – 710М, магістр кафедри Захисту інформації Запорізького національного технічного університету, Запоріжжя, Україна

[2]Канд. фіз.-мат. наук, доцент кафедри Захисту інформації Запорізького національного технічного університету, Запоріжжя, Україна

МОДИФІКАЦІЯ АЛГОРИТМУ ЗАПЕРЕЧУВАНОГО ШИФРУВАННЯ МЕНГА

В статті обговорюється проблема стійкості сучасних криптографічних систем до атак на основі примушування стосовно абонентів криптографічних систем. У зв'язку зі стрімким розвитком галузі інформаційних технологій ця проблема є актуальною в сфері інформаційної безпеки. Для вирішення проблеми стійкості сучасних криптографічних систем авторами запропоновано використання алгоритмів заперечуваного шифрування, які гарантують, що зловмисник не має змоги отримати будь-яку цінну інформацію від абонентів. Вирішення проблеми полягає у використанні алгоритму заперечуваного шифрування Менга, який гарантує захист не лише інформації, а й самих учасників обміну.

Основна мета статті полягає у виконанні модифікації первісного алгоритму заперечуваного шифрування Менга [1] шляхом використання протоколу «непомітної» передачі OT_n^1, використання якого запропоновано Моні Наором [2]. Застосування протоколу «непомітної» передачі OT_n^1 дозволяє суттєво скоротити час, необхідний для виконання алгоритму заперечуваного шифрування Менга, та спрощує його реалізацію для вирішення прикладних завдань в сфері інформаційної безпеки.

За результатами експериментів авторами статті було підтверджено, що використання протоколу «непомітної» передачі OT_n^1 є ефективним вирішенням проблеми генерації і розподілу ключів в алгоритмі заперечуваного шифрування Менга.

Ключові слова: протокол, заперечуване шифрування, неоднозначність, розподіл ключів, фіктивне повідомлення, зловмисник, розширена схема Рабіна, непомітна передача, факторизація, дискретне логарифмування.

НОМЕНКЛАТУРА

AMD – марка виробника центральних процесорів;

BCP – протокол шифрування розроблений E. Bresson, D. Catalano і D. Pointcheval;

RD – PKE – Receiver – Deniable Public Key Encryption Protocol;

RSA – протокол шифрування розроблений Rivest, Shamir і Adleman;

Гб – одиниця вимірювання об'єму пам'яті;

ГГц – одиниця вимірювання частоти (10^{-9} секунди);

КЗІ – криптографічний захист інформації;

мс – одиниця вимірювання часу ($0,001$ секунди);

ОЗУ – оперативний запам'ятовуючий пристрій (внутрішня пам'ять);

ОС – операційна система;

ПЗ – програмне забезпечення;

ЦП – центральний процесор.

$\left(A_1, A_2\right)$ – пара числових значень A_1 і A_2;

a – секретний ключ відправника;

α – елемент мультиплікативної групи;

$C\left(r_2, m\right)$ – контрольна функція від числових значень r_2 і m;

D – проміжне числове значення;

$\left\langle\left(\partial, \wp\right), B\right\rangle$ – криптограма в алгоритмі заперечуваного шифрування Менга;

$\left(e; n\right)$ і $\left(d; n\right)$ – публічний та секретний ключ в алгоритмі RSA;

DOI 10.15588/1607-3274-2016-2-10

$\langle (e,n),x \rangle$ – трійка публічний параметрів в протоколі OT_n^1;

$f_{\text{ДЛ}}(\ldots)$ – функція дискретного логарифмування;

g – генератор мультиплікативної групи;

$hash(r_2)$ – значення хеш – функції від r_2;

$k < n$ – випадкове числове значення;

k' – проміжне числове значення;

m – секретне повідомлення, яке належить до групи Z_N;

m' – фіктивне повідомлення, яке належить до групи Z_N;

(m_1, m_2) – пара числових значень m_1 і m_2;

(m,r) – пара числових значень, еквівалентна m і r_2;

N – модуль утворений добутком чисел p і q;

$\langle (N,g,h),(p,q) \rangle$ – пара публічного та секретного ключів одержувача;

$ord(\ldots)$ – порядок будь-якого числового елемента групи;

OT_n^1 – Oblivious Transfer Protocol;

p – модуль перетворень еквівалентний N^2;

p і q – великі прості сильні числа (p' і q');

r – випадковий елемент мультиплікативної групи, еквівалентний r_2;

r_1 – випадковий елемент мультиплікативної групи;

$r_2 \in \langle g \rangle$ – випадковий елемент мультиплікативної групи, побудований на g;

$\{T_i, M_i\}$ – пара секретного та фіктивного повідомлень в алгоритмі Рабіна;

t_p – час виконання будь – якої операції при даному розмірі модуля p;

$t_{\text{гк}}$ – числове значення часу генерації ключів в алгоритмі;

ν – проміжне числове значення;

$X(t)$ – числове значення псевдовипадкової послідовності;

$x < n$ – випадкове числове значення;

$\langle (y,g,p),a \rangle$ – пара публічного та секретного ключів відправника;

γ_1 – проміжне числове значення;

$Z^*_{N^2}$ – мультиплікативна група для числа N^2.

ВСТУП

Сучасні алгоритми шифрування не можуть забезпечувати абсолютну стійкість до атак зловмисників як зовні, так і з середини системи. Атаки на основі примушування можна застосовувати для отримання будь-якої секретної інформації. Вирішення цієї проблеми полягає у застосуванні алгоритмів «неоднозначного» або заперечуваного шифрування.

Наукова новизна полягає у функціональній заміні задачі дискретного логарифмування, яка використовується в процедурі розподілу ключів, на протокол «непомітної» передачі OT_n^1 в алгоритмі заперечуваного шифрування Менга.

Актуальність теми статті полягає у зниженні ефективності захисту інформації сучасними криптографічними системами та необхідність пошуку нових рішень для виправлення цієї проблеми.

Об'єкт дослідження – процедура розподілу ключів між абонентами в алгоритмі заперечуваного шифрування Менга.

Актуальність теми статті полягає у зниженні ефективності захисту інформації сучасними криптографічними системами та необхідність пошуку нових рішень для виправлення цієї проблеми.

Об'єкт дослідження – процедура розподілу ключів між абонентами в алгоритмі заперечуваного шифрування Менга.

Предмет дослідження – це властивостей алгоритму заперечуваного шифрування Менга та особливостей протоколу «непомітної» передачі OT_n^1.

Мета роботи полягає в модифікації алгоритму заперечуваного шифрування Менга шляхом огляду та аналізу існуючих алгоритмів і протоколів безпечної передачі інформації. Для досягнення цієї мети необхідно виконати наступні завдання:

– зробити огляд аналогічних рішень у подібних алгоритмах шифрування персональних даних;

– виконати аналіз структури та особливостей захисту інформації подібними алгоритмами заперечуваного шифрування;

– виконати модифікацію процедури розподілу ключів у алгоритмі заперечуваного шифрування Менга;

– виконати порівняння швидкодії процедур розподілу ключів в первісному та модифікованому алгоритмах заперечуваного шифрування Менга, зробити висновки щодо результів модифікацій алгоритму та їх вплив на ефективність захисту інформації.

1 ПОСТАНОВКА ЗАДАЧІ

Основна проблема при використанні алгоритму заперечуваного шифрування Менга полягає в умовності його структури, що унеможливлює його практичне застосування для захисту інформації в реальних комп'ютерних мережах. Оскільки вона виникає на етапі генерації ключів, то необхідно виконати модифікацію процедури розподілу ключів, але не робити значних змін у структурі самого алгоритму шифрування.

Суть проблеми полягає у неможливості виконання розподілу ключів між абонентами через використання задачі дискретного логарифмування для передачі секретного ключа a відправника одержувачеві. Оскільки задача дискретного логарифмування за складністю обчислень еквівалентна до задачі з факторизації, то при обчисленні одержувачем секретного ключа a відправника з публічного параметру h і модуля $p \sim N^2$ час виконання процедури розподілу ключів можна описати виразом $t_{p \sim N^2} \to \infty$.

Для забезпечення оптимального часу виконання процедури генерації та розподілу ключів використати протокол «непомітної» передачі OT_n^1, який описано в [2] і застосовано для вирішення подібної проблеми в алгоритмі заперечуваного шифрування Ібраxіма [3]. Оскільки стійкість протоколу «непомітної» передачі ґрунтується на вирішенні задачі з факторизації, то необхідно виконати аналіз швидкодії даного протоколу із швидкістю вирішення еквівалентної задачі – дискретного логарифмування, в оригінальному алгоритмі заперечуваного шифрування Менга.

2 ОГЛЯД ЛІТЕРАТУРИ

Для надійного захисту інформації, як на локальних комп'ютерах так і тієї, що передається по мережі, використовуються криптографічні засоби захисту інформації. Одним із напрямів криптографічного захисту інформації є шифрування. В зв'язку з цим проводиться безперервні дослідження існуючих криптографічних схем шифрування та створення нових криптографічних схем, на їх базі. Як новий напрям, в криптографії виділяють розробку та дослідження алгоритмів заперечуваного шифрування. В сучасній літературі для цього напряму використовують такі ж поняття, як «неоднозначне», «суперечливе» або «заперечуване».

Існує не так багато джерел, які описують алгоритми заперечуваного шифрування, на відміну від інших. Така тенденція пов'язана з одним із основним аспектів заперечуваного шифрування, який гарантує надійність захисту – таємність алгоритмів заперечуваного шифрування. В результаті досліджень даного напряму було знайдено деякі публікації і статті, які описують перспективність алгоритмів заперечуваного шифрування для застосування в сфері криптографічного захисту інформації.

Досліджені алгоритми заперечуваного шифрування, здебільшого, побудовані на основі криптографічних систем з відкритим ключем, оскільки останні набули популярності в останні роки. До таких алгоритмів можна віднести алгоритми заперечуваного шифрування розроблені: Раном Канетті, Хамадою Ібрахімом, Джин – Квін Вангом і Бо Менгом, М. А. Молдов'яном і О.О. Горячевим, Шаффі Голвасером і С. Мікалі [4] та ін.

Алгоритм заперечуваного шифрування Рана Канетті [4] є криптографічною системою з відкритим ключем, яка дозволяю користувачам не виконувати попередній обмін секретними параметрами для шифрування/дешифрування інформації. Алгоритм заперечуваного шифрування Канетті передбачає реалізацію механізму побітового шифрування даних, тому він є менш продуктивним, але має досить надійний рівень захисту.

Алгоритм заперечуваного шифрування Хамади Ібрахіма [3] побудовано на базі протоколу RD – PKE, який дозволяє повністю захистити одержувача секретної інформації від застосування примушування. Така змога надається шляхом поділу секретного ключа між одержувачем та довіреною стороною безпеки за допомогою протоколу [3]. Оскільки одержувач не має достатньої кількості інформації для самостійного розшифрування криптограми, то застосування примушування стосовно нього не є ефективним.

Алгоритм заперечуваного шифрування М. А. Молдов'яна і О. О. Горячева[5] побудовано на базі розширеної криптографічної схеми Рабіна з відкритим ключем. Даний алгоритм заперечуваного шифрування на достатньому рівні вирішує завдання пов'язані зі стійкістю криптографічної системи до атак заснованих на базі примушування. Модифікація алгоритму шифрування Рабіна передбачає відновлення чотирьох різних пар секретних/фіктивних повідомлень $\{T_i, M_i\}$, три з яких мають випадковий характер. Оскільки даний алгоритм відноситься до криптографічних систем з відкритим ключем, то основою захисту є обчислювальна складність, яка базується на вирішенні задачі факторизації. Таким чином, секретність власне алгоритму не є критичною.

Алгоритм заперечуваного шифрування Шафі Голдвасера і Сильвіо Мікалі [4] побудовано на базі комбінованого алгоритму ймовірнісного шифрування [6]. В літературі даний алгоритм отримав назву – асоційований алгоритм ймовірнісного шифрування. Основним критерієм захисту є неоднозначність дешифрованих повідомлень, тобто ймовірність того, що зловмисник має змогу дешифрувати криптограму та відрізнити результати від випадкових значень досить низька. Проте його недоліком є можливість дешифрування лише одного повідомлення.

Алгоритм заперечуваного шифрування Джин-Квін Ванга і Бо Менга [1] побудовано на базі протоколу BCP [7] та ідей запропонованих Клоновскі [8]. В даному алгоритмі комбіновано кращі сторони алгоритмів RSA та Ель-Гамаля, саме тому він має більш кращі показники захищеності та продуктивності роботи. Представлений алгоритм заперечуваного шифрування передбачає використання двох алгоритмів розшифрування захищеного повідомлення, який катастрофічно знижував ефективність захисту. Але використання двох пар секретних ключів і обчислювальна складність криптографічних систем з відкритим ключем забезпечують досить надійний захист секретної інформації та досить високий рівень продуктивності.

3 МАТЕРІАЛИ І МЕТОДИ

Алгоритм заперечуваного шифрування Менга вирішує проблему щодо атак на основі примушування, яке застосовується як до відправника та одержувача, так і до обох одночасно. Таким чином виключається будь – яка можливість отримання секретної інформації зловмисником. Структурна схема алгоритму заперечуваного шифрування Менга приведена в Додаток А.

Оскільки на етапі генерації ключів виникає проблема пов'язана з розподілом ключів між абонентами, то її вирішення полягає у модифікації первісного алгоритму (Додаток А) шляхом функціональної заміни задачі дискретного логарифмування на протокол «непомітної» передачі OT_n^1 (рис. 1).

Згідно з результатами експериментів проведених у пункті 4 можна зробити висновок, що застосування протоколу «непомітної» передачі OT_n^1 в процедурі розподілу ключів алгоритму заперечуваного шифрування Мен-

га має більш оптимальний час і спрощує його технічну реалізацію. Таким чином модифікований варіант алгоритму заперечуваного шифрування Менга приведено в Додаток Б, а приклад його роботи приведено нижче (в якості тестових повідомлень використані «Security» і «United Ukraine» відповідно секретне та фіктивне тестові повідомлення).

На етапі підготовки було виконано генерацію секретних і публічних ключів одержувача $\langle (N,g,h),(p,q) \rangle$ та відправника $\langle (y,g,p),a \rangle$, які приведено в табл. 1.

На етапі шифрування було виконано обчислення криптограми $\langle (\partial,\wp),B \rangle$, яку відправник передає по відкритому каналу одержувачеві. Результати обчислень приведено в табл. 2.

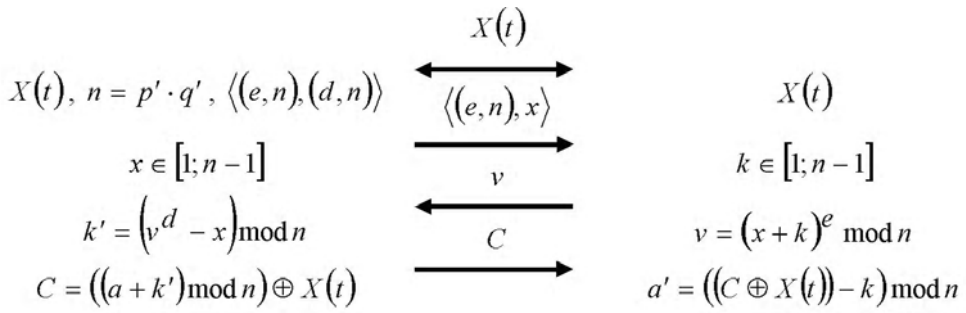


Рисунок 1 – Структурна схема протоколу «непомітної передачі» OT_n^1

Таблиця 1 – Секретний та відкритий ключі відправника та одержувача

Абоненти	Позначення	Значення
Відправник	N	*340282393626195344062954022172595310137*
	g	*45769203547515261684864 4*
	h	*109066888395335831186793921949405725400121626796378019323902565298338926914653*
	p	*18446744847862403543*
	q	*18446744747251547759*
Одержувач	$X(t)$	*454517681881*
	(e,d,n)	*(325514516927,13740242815079940142790 3,663108052854589966200253)*
	x	*233168199764*
	k	*365161775673*
	v	*210858578369542822886041*
	a	*775224541535*
	C	*1513289100609*
	a'	*775224541535*
	y	*109066888395335831186793921949405725400121626796378019323902565298338926914653*
	g	*45769203547515261684864 4*
	p	*11579210741197294971852128355385294945305813979775842618871218310928721495876 9*

Таблиця 2 – Обчислення криптограми $\langle (\partial,\wp),B \rangle$

Позначення параметрів	Значення
r_2	*976111453564*
m	*8751735916204352851*
$C(r_2,m)$	*4944128327931454803528618175015452521801373092820406824902092812272307796148 8*
m'	*2057271081577553248481314125475413*
r_1	*2744558201036439643830698319078052477901324357140583800248748001864070337832 6*
$C(r_1,m')$	*4944128327931454803528618175015452521801373092820406824902092812272307796148 8*
$hash(r_2)$	*976111453573*
B	*4944128327931454803528618175015452521801373092820406824902092812272307796148 8*
∂	*7120757359751600539380457168433360239215956805097502870744764841488557454648*
$\wp$	*8983929684275778172654129352885340170739411186409128932028046571941013467619 3*

На етапі розшифрування одержувачем було відновлено секретне повідомлення m з криптограми $\langle(\partial,\wp),B\rangle$, результати обчислень якого приведено в табл. 3.

В результаті виконаних обчислень одержувач отримав секретне повідомлення «Security», яке було адресоване йому.

На етапі дешифрування, в результаті застосування примусу, одержувач виконав відновлення фіктивного повідомлення m' з криптограми $\langle(\partial,\wp),B\rangle$, яке призначене власне для зловмисника, результати обчислень якого приведено в табл. 4.

В результаті виконаних обчислень одержувач відновив фіктивне повідомлення «United Ukraine» та передав його зловмисникові.

Таким чином було виконано моделювання роботи модифікованого алгоритму заперечуваного шифрування Менга за допомогою 128-бітного ключа та продемонстровано його особливості, які забезпечують надійний захист інформації від атак на основі примусу. Для оцінки ефективності виконаних модифікацій алгоритму авторами проведено експерименти щодо визначення оптимальної швидкодії його роботи шляхом порівняння швидкодії первісного та модифікованого алгоритмів.

Оскільки модифікація полягала в покращенні часових характеристик процедури генерації та розподілу ключів, то для спрощення експериментів, у пункті 4, було досліджено швидкодії процедур генерації ключів $t_{\text{ГК}}$: на основі задачі дискретного логарифмування та протоколу «непомітної» передачі OT_n^1.

4 ЕКСПЕРИМЕНТИ

Для покращення часових характеристик процедури генерації та розподілу ключів авторами було проведено експерименти з дослідження швидкодії процедур генерації ключів $t_{\text{ГК}}$: на основі задачі дискретного логарифмування та протоколу «непомітної» передачі OT_n^1 (рис. 2).

Для виконання експериментів було використано структурні схеми первісного та модифікованого алгоритмів заперечуваного шифрування Менга (Додаток А і Додаток Б) і персональний комп'ютер з технічними характеристиками:

– ОС: Windows 7;

– ЦП: AMD Athlon(tm) II P360 Dual – Core Processor 2.3 ГГц;

– ОЗУ: 3 Гб;

– ПЗ: пакет математичних обчислень Maple 14.

Основний показник, який досліджувався в ході експериментів – це швидкодія виконання процедури генерації ключів $t_{\text{ГК}}$. За результатами проведених експериментів автори виконали порівняльну характеристику швидкодії процедур генерації ключів і визначили оптимальний час $t_{\text{ГК}}$ для виконання процедури генерації ключів, а також метод за допомогою якого вона виконана.

5 РЕЗУЛЬТАТИ

В ході проведених експериментів (рис. 2) автори отримали експериментальні дані щодо генерації ключів із використанням задачі дискретного логарифмування (табл. 5).

При досяжності розрядності модуля p у 256 біт і більше, в першому експерименті спостерігається експоненціальне зростання часу відведеного на генерацію ключів. Оскільки за вимогами сучасних криптографічних систем безпечна розрядність ключа повинна складати 1024–2048 біт, то час необхідний для генерації ключів можна описати виразом $t_p \to \infty$.

В ході проведення другого експерименту (рис. 2) експериментальні дані, отримані авторами, носять практично константний характер, тобто одне й те саме значення для ключів, яке складає 15–16 мс для ключів розрядністю до 2048 біт і більше (коливання часу на рівні 1 мс).

Таблиця 3 – Відновлення секретного повідомлення m

Позначення параметрів	Значення
D	*29605105317476900925245466022643727530484548476548504967891351479412026975 42*
π	*58422416232404436471085961576632679733*
m	*8751735916204352851*

Таблиця 4 – Відновлення секретного повідомлення m'

Позначення параметрів	Значення
r_2	*976111453564*
r_1	*27445582010364396438306983190780524779013243571405838002487480018640703378326*
A_1	*86120330409262410220532378382346931244890834587769244209422058121590708003399*
A_2	*42368364285961378337324940846616483686887069999828052556155055445614458883516*
m_1	*20572710815775532484813141254754 13*
m_2	*8751735916204352851*
m	*8751735916204352851*
r	*976111453564*
m'	*20572710815775532484813141254754 13*

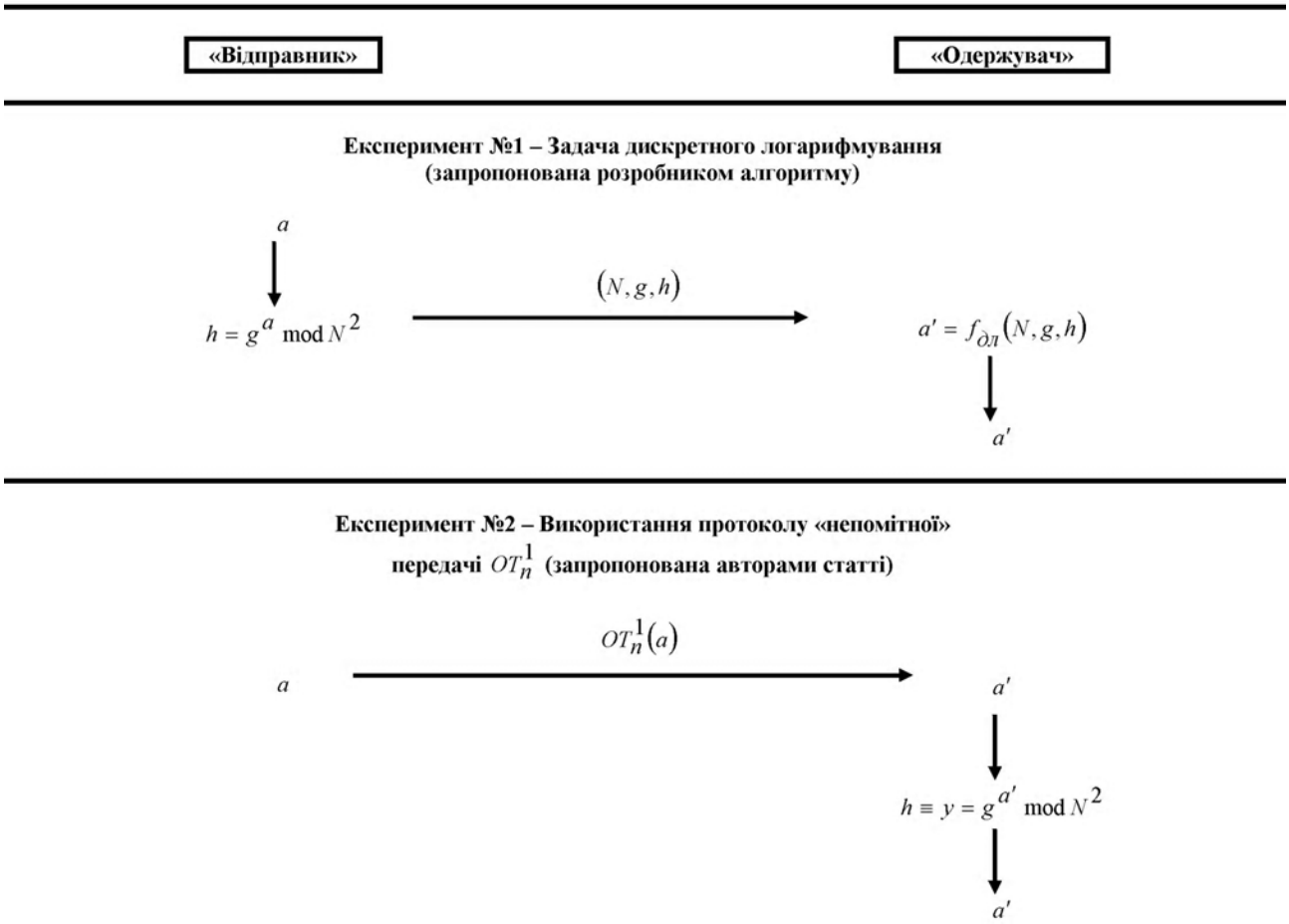


Рисунок 2 – Схема експериментів

Таблиця 5 – Дані отримані з першого експерименту

Модуль p, біт	8	16	32	64	128	256
Час $t_{ГК}$, секунд	0,051	0,46	0,61	6,19	598,92	8329,3

6 ОБГОВОРЕННЯ

Проблема практичної реалізації алгоритму заперечуваного шифрування Менга на початку статті полягала обчислювальній стійкості задачі дискретного логарифмування, при якій час генерації виконання усього алгоритму в цілому описується виразом $t_p \to \infty$. Використовуючи експериментальні дані з табл. 5 автори зробили теоретичний прогноз щодо зростання часу генерації ключів, який відображає графік на рис. 3.

Згідно із прогнозом авторів у першому експерименті, для генерації ключів в алгоритмі заперечуваного шифрування Менга розрядністю до 4096 біт, комп'ютеру із запропонованими тестовими характеристиками необхідно витратити до 596 днів машинного часу. Що згідно зі стандартами криптографічних систем, які декларують безпечну розрядність ключа в 1024–2048 біт, є не оптимальним застосуванням алгоритму для вирішення прикладних завдань.

За результатами другого експерименту авторами було встановлено, що оптимальний час для виконання всього алгоритму заперечуваного шифрування Менга складає 15–16 мс для ключів розрядністю до 4096 біт і

Рисунок 3 – Прогнозована оцінка зростання часу виконання процедури генерації ключів в першому експерименті

більше. Таким чином, дані отримані в результаті проведення другого експерименту (рис. 2), із використанням протоколу «непомітної» передачі OT_n^1, свідчать про ефективність виконаних модифікацій в алгоритмі заперечуваного шифрування Менга. Отже підхід запропонований авторами для вирішення проблеми у первісного алгоритму є прийнятним з точки зору інформаційної безпеки та практичної реалізації власне алгоритму, а отже застосування протоколу «непомітної» передачі в достатній мірі вирішує проблему пов'язану зі структурою первісного алгоритму заперечуваного шифрування Менга [7].

ВИСНОВКИ

Згідно з метою, поставленою на початку статті, авторами вирішено проблему практичного застосування алгоритму заперечуваного шифрування Менга. В результаті проведених досліджень та експериментів було вирішено наступні завдання:

– виконано огляд рішень проблеми, яку автори визначили структурі алгоритму заперечуваного шифрування Менга, на основі аналогічних алгоритмів заперечуваного шифрування персональних даних;

– виконано аналіз структури та особливостей подібних алгоритмів заперечуваного шифрування та визначено методи для вирішення проблеми;

– виконано модифікацію алгоритму заперечуваного шифрування Менга за допомогою використання протоколу «непомітної» передачі OT_n^1;

– на основі проведених експериментів було проведено аналіз ефективності виконаних модифікацій і доведено ефективність використання протоколу «непомітної» передачі OT_n^1 для усунення недоліків первісного алгоритму заперечуваного шифрування Менга.

За результатами досліджень автори виконали модифікацію алгоритму заперечуваного шифрування Менга (Додаток Б). Для вирішення проблеми було використано новий підхід заснований на використанні протоколу «непомітної» передачі OT_n^1, що дозволило суттєво зменшити час роботи алгоритму Менга та зробити можливим його використання для вирішення практичних завдань в сфері інформаційної безпеки.

Наукова новизна полягає у використанні протоколу «непомітної» передачі для розподілу ключів між абонентами, що є принципово новим підходом до захисту інформації в алгоритмі заперечуваного шифрування Менга.

Робота виконана на кафедрі захисту інформації в рамках НДКР №04515 «Дослідження і розробка криптографічних та технічних засобів захисту інформації».

СПИСОК ЛІТЕРАТУРИ

1. Wang J. A Receiver Deniable Encryption Scheme / J. Wang, Bo Meng // Proceedings of the 2009 International Symposium on Information Processing (ISIP'09), 21–23 August 2009: proceedings. – Huangshan : P. R. China, 2009. – P. 254–257.
2. Naor M. Efficient oblivious transfer protocols / M. Naor, B. Pinkas // Proceedings of SIAM Symposium on Discrete Algorithms (SODA '01), 2001: proceedings. – Society for Industrial and Applied Mathematics, 2001. – P. 448–457.
3. Ibrahim H. Receiver–deniable Public–Key Encryption / H. Ibrahim // International Journal of Internet Security. – 2009. – Vol. 8, № 2. – P. 159–165.
4. Козіна Г. Л. Заперечуване шифрування / Г. Л. Козіна, А. В. Гальченко // Тиждень науки – 2015: Тези доповідей щорічної наук. – практ. конф. викладачів, науковців, молодих учених, аспірантів, студентів ЗНТУ, Запоріжжя, 13–17 квітня 2015 р. – Запоріжжя : ЗНТУ, 2015.
5. Canetti R. Deniable Encryption / [R. Canetti, C. Dwork, M. Naor, R. Ostronsky] // Advances in Cryptology. – CRYPTO, 1997, Proceedings. – P. 90–104.
6. Молдовян Н.А. Расширение криптосхемы Рабина: алгоритм отрицаемого шифрования по открытому ключу / Н. А. Молдовян, А. А. Горячев, М. А. Вайчикаускас // ВЗИ. Журнал по вопросам защиты информации. – ФГУП «ВИМИ», 2014. – № 2. – С. 12–16.
7. Фисун С.Н. Комбинированный алгоритм вероятностного шифрования / С. Н. Фисун, О. И. Куржиевская // Изд-во СевНТУ, 2010. – № 101. – С. 37–40.
8. Bresson E. A Simple Public – Key Cryptosystem with a Double Trapdoor Decryption Mechanism and its Applications / E. Bresson, D. Catalano, D. Pointcheval // Advances in Cryptology – ASIACRYPT, 2003. LNCS, Vol. 2894. Springer, Heidelberg, 2003. – P. 37–54.
9. Klonowski M. Practical Deniable Encryption // M. Klonowski, P. Kubiak, and M. Kutyiowsk // SOFSEM 2008: Theory and Practice of Computer Science, 34th Conference on Current Trends in Theory and Practice of Computer Science, Slovakia : Novy Smokovec, 19–25 January 2008: proceedings. – 2008. – P. 599–609.
10. Николенко С. Поиск дискретного логарифма [Электронный ресурс] / С. Николенко. – Режим доступа: https://compscicenter.ru/media/slides/cryptoprotocols2014_2015_spring/2015_03_11_cryptoprotocols2014_2015_spring_lGKdY5s.pdf.
11. Горбенко И. Д. Методы распараллеливания алгоритма Полларда решения задачи дискретного логарифмирования для систем с общей памятью / И. Д. Горбенко, Е. Г. Качко, К. А. Погребняк // Изд-во ХНУРЭ, 2012. – С. 1–6.

Стаття надійшла до редакції 16.12.2015.
Після доробки 20.12.2015.

Гальченко А. В.[1], Козина Л.[2]

[1]Ст. гр. РТ – 710М, магистр кафедры защиты информации Запорожского национального технического университета, Запорожье, Украина

[2]Канд. физ.-мат. наук, доцент кафедры защиты информации Запорожского национального технического университета, Запорожье, Украина

МОДИФИКАЦИЯ АЛГОРИТМА ОТРИЦАЕМОГО ШИФРОВАНИЯ МЕНГА

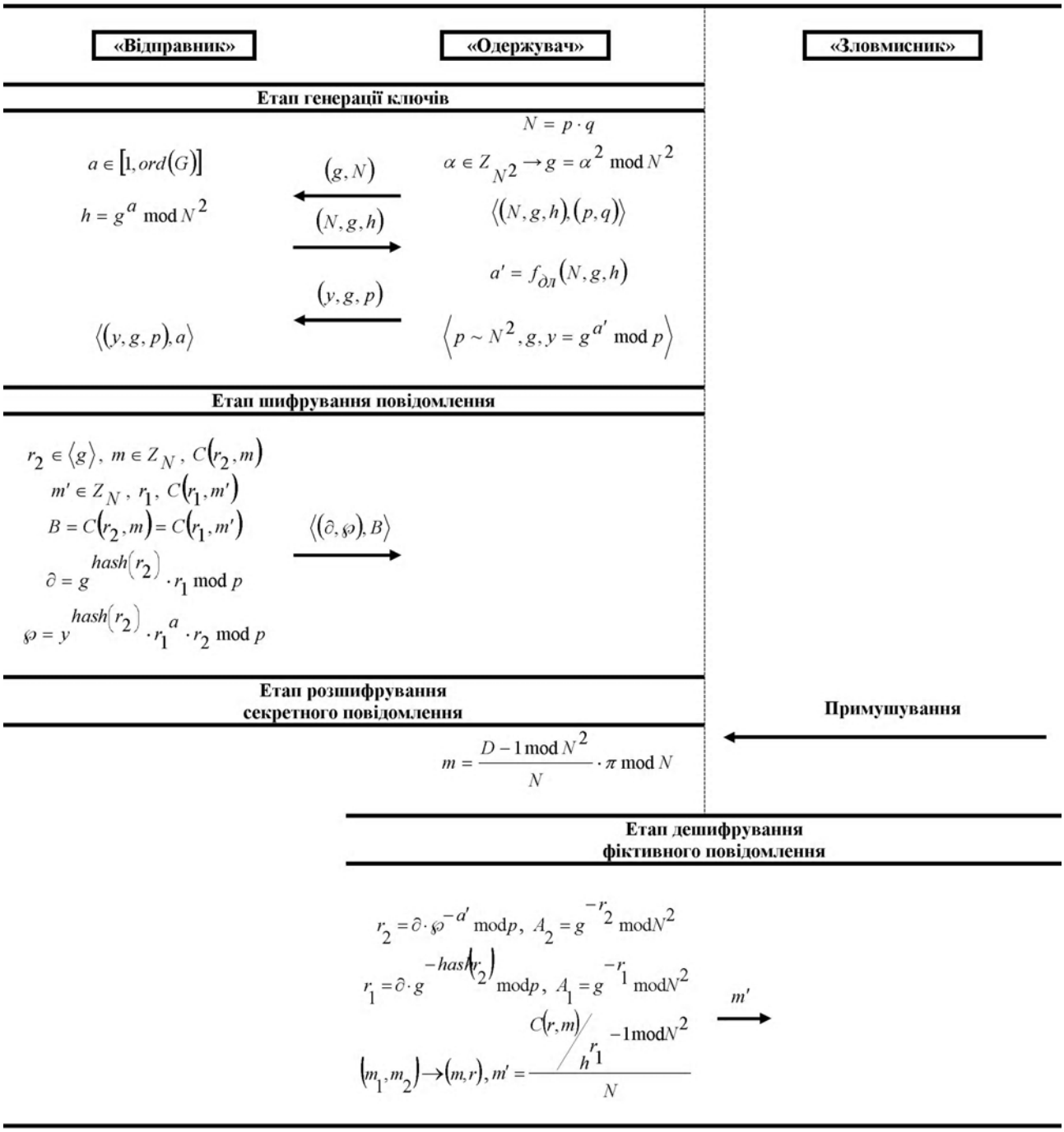
В статье обсуждается проблема устойчивости современных криптографических систем к атакам на основе принуждения в отношении абонентов криптографических систем. В связи со стремительным развитием отрасли информационных технологий эта проблема актуальна в сфере информационной безопасности. Для решения проблемы устойчивости современных криптографических систем авторы предлагают использовать алгоритмы отрицаемого шифрования, которые гарантируют, что злоумышленник не имеет возможности получить какую – либо ценную информацию от абонентов. Решение проблемы заключается в использовании алгоритма отрицаемого шифрования Менга, который гарантирует защиту не только информации, но и самих участников обмена.

Основная цель статьи состоит в модификации первоначального алгоритма отрицаемого шифрования Менга [1] с помощью протокола «незаметной» передачи OT_n^1, использование которого предложено Мони Наором [2]. Применение протокола «незаметной» передачи позволяет существенно сократить время, необходимое для выполнения алгоритма отрицаемого шифрования Менга, и упрощает его реализацию для решения прикладных задач в области информационной безопасности.

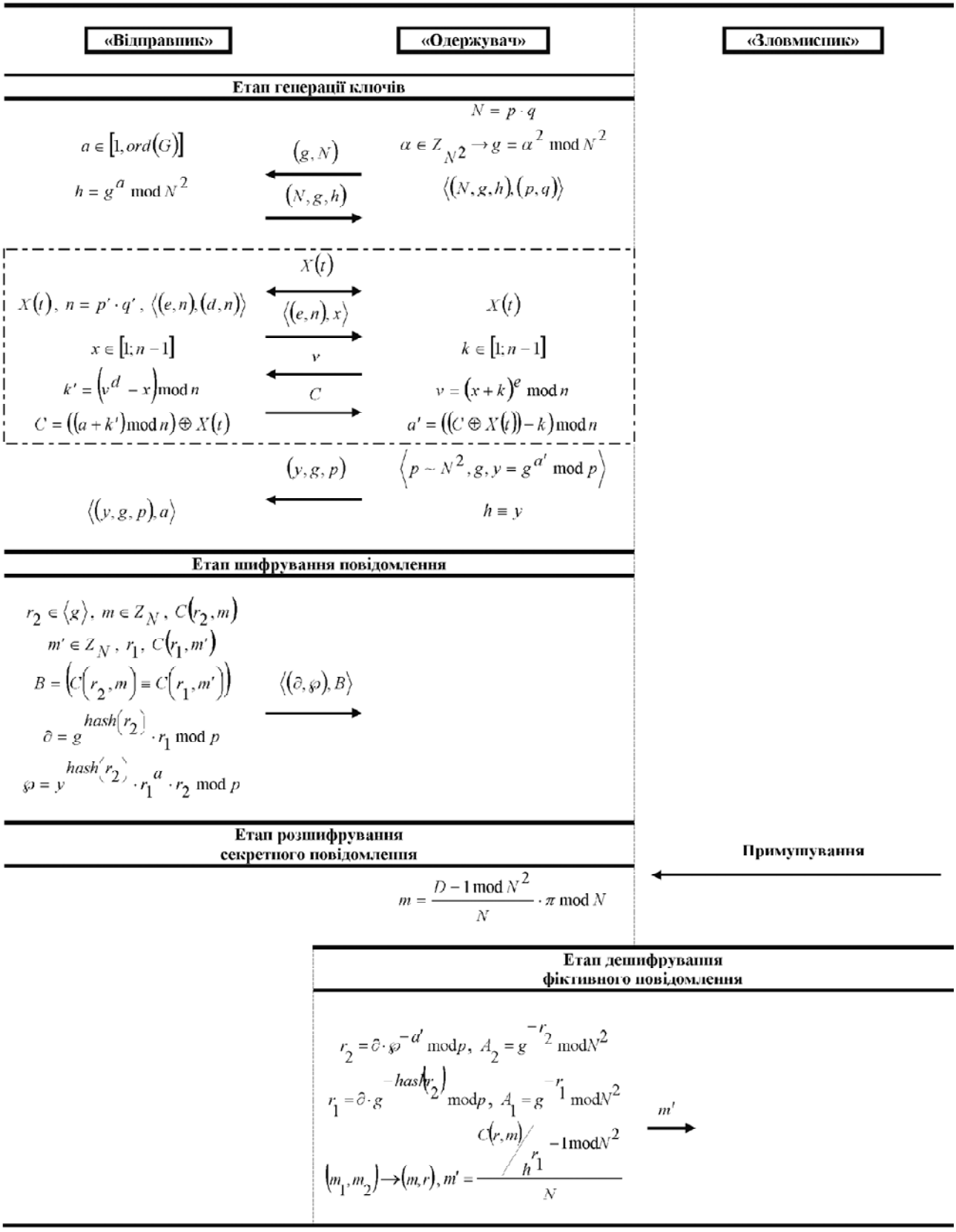
По результатам проведенных экспериментов авторами статьи было подтверждено, что использование протокола «незаметной» передачи является эффективным решением проблемы генерации и распределения ключей в алгоритме отрицаемого шифрования Менга.

Ключевые слова: протокол, отрицаемое шифрование, неоднозначность, распределение ключей, фиктивное сообщение, злоумышленник, расширенная схема Рабина, незаметная передача, факторизация, дискретное логарифмирования.

Додаток А – Структурна схема первісного алгоритму заперечуваного шифрування Менга



Додаток Б – Структурна схема модифікованому алгоритму заперечуваного шифрування Менга



Galchenko A. V.[1], Kozina G. L.[2]

[1]Gr. RT – 710M, Master of the Department of Information Protection of Zaporozhzhya National Technical University, Zaporizhzhya, Ukraine

[2]Candidate of Phys.-Math. sciences, Associate Professor of the Department of Information Protection of Zaporozhzhya National Technical University, Zaporizhzhya, Ukraine

MODIFICATION OF MENG'S DENIABLE ENCRYPTION ALGORITHM

The article discusses the stability of modern cryptographic systems to attack from coercion in respect of subscribers cryptographic systems. Due to the rapid development of information technology, this problem is relevant in the field of information security. To address the sustainability of modern cryptographic systems use algorithms offered by the authors deniable encryption ensures that the attacker is unable to get any valuable information from subscribers. Solving the problem is to use Meng's deniable encryption algorithm, which guarantees protection not only information, but also the participants of the exchange.

The main purpose of the article is performed by modifying the initial Meng's deniable encryption algorithm [1] with using Oblivious Transfer Protocol , which prompted by Moni Naor [2]. Oblivious Transfer Protocol to significantly reduce the time required to perform the Meng's deniable encryption algorithm, and facilitates its implementation to solve applied problems in the field of information security.

As a result of experiments, the authors confirmed that Oblivious Transfer Protocol using is an effective solution to the problem of generation and distribution of keys in the Meng's deniable encryption algorithm.

Keywords: protocol, deniable encryption, ambiguity, distribution keys, fake messages, an extended Rabin's scheme, oblivious transfer, factorization, discrete logarithm.

REFERENCES

1. Wang J., Meng Bo A Receiver Deniable Encryption Scheme, *Proceedings of the 2009 International Symposium on Information Processing (ISIP'09), 21–23 August 2009: proceedings*, Huangshan, P. R. China, 2009, pp. 254–257.
2. Naor M., Pinkas B. Efficient oblivious transfer protocols, *Proceedings of SIAM Symposium on Discrete Algorithms (SODA '01), 2001: proceedings. – Society for Industrial and Applied Mathematics,* 2001, pp. 448–457.
3. Ibrahim H. Receiver-deniable Public-Key Encryption, International Journal of Internet Security, 2009, Vol. 8, No. 2, P. 159–165.
4. Kozina G. L., Galchenko A. V. Deniable encryption, *Weeak of Science – 2015: Abstracts annual scientific – practical conference of teachers, scientists, young scientists, graduate students ZNTU, Zaporizhzhya , 13–17 April 2015*. Zaporizhzhya, ZNTU, 2015.
5. Canetti R., Dwork C., Naor M., Ostronsky R. Deniable Encryption. Advances in Cryptology – CRYPTO, 1997, Proceedings, pp. 90–104.
6. Moldovyan N. A., Goryachew A. A., Wichikaukas M. A. Extened Rabins's cryptographic scheme: deniable encryption by public key, *VZI. Journal of information security*. FSUE «VIMI», 2014, No. 2, pp. 12–16.
7. Fisun S. N., Kurzhievskaya O. I. Combined probabilistic encryption algorithm, *SevNTU*, 2010, No. 101, pp. 37–40.
8. Bresson E., Catalano D., Pointcheval D. A Simple Public – Key Cryptosystem with a Double Trapdoor Decryption Mechanism and its Applications, *Advances in Cryptology – ASIACRYPT, 2003. LNCS, Vol. 2894*. Springer, Heidelberg, 2003, pp. 37–54.
9. Klonowski M., Kubiak P., and Kutyiowsk M. Practical Deniable Encryption, *SOFSEM 2008: Theory and Practice of Computer Science, 34th Conference on Current Trends in Theory and Practice of Computer Science, Slovakia: Novy Smokovec, 19–25 January 2008: proceedings*. – 2008. P. 599–609.
10. Nikolenko S. Search the discrete logarithm [Electronic resource]. Access: https://compscicenter.ru/media/slides/cryptoprotocols2014_2015_spring / 2015_03_11_cryptoprotocols2014_2015_spring_lGKdY5s.pdf.
11. Gorbenko I. D., Muscled E. G., Pogrebnyak K. A. Methods of algorithm parallelization Pollard solutions discrete logarithm problem for systems with shared memory, *KNURE Publishing House*, 2012, pp. 1–6.

УДК 004.932.2:004.93'1

Гороховатский В. А.[1], Берестовский А. Е.[2], Передрий Е. О.[3]

[1]Д-р. техн. наук, профессор, профессор кафедры информационных технологий, Харьковский учебно-научный институт ГВУЗ «Университет банковского дела», Харьков, Украина
[2]Аспирант кафедры информатики, Харьковский национальный университет радиоэлектроники, Харьков, Украина
[3]Канд. техн. наук, преподаватель кафедры информатики и компьютерной техники, Харьковский национальный экономический университет имени Семена Кузнеца, Харьков, Украина

СИСТЕМАТИЗАЦИЯ ПРОСТРАНСТВА СТРУКТУРНЫХ ПРИЗНАКОВ НА ОСНОВЕ МЕТОДОВ САМООБУЧЕНИЯ В ЦЕЛЯХ РЕЗУЛЬТАТИВНОГО РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ

Работа посвящена исследованию вопросов кластеризации для множеств характерных признаков изображений. Для построения массива характерных признаков использован метод Speeded Up Robust Features. Реализованы алгоритмы кластеризации структурных описаний изображений на основе самоорганизующейся нейронной сети Кохонена и метода разностного группирования. Объектом исследования есть методы кластеризации применительно к множествам структурных признаков. Целью работы является построение векторных представлений описаний на основе кластеризации, что повышает быстродействие распознавания. Предметом исследования является систематизация множеств структурных признаков визуальных объектов.

Обсуждаются результаты применения методов кластеризации для структурных описаний изображений в виде множеств характерных признаков с целью повышения быстродействия распознавания визуальных объектов. Для систематизации и сжатия пространства признаков предложено осуществить самообучение с применением методов разностного группирования и сетей Кохонена.

Проведено моделирование и экспериментальные исследования методов кластеризации на примерах конкретных множеств характерных признаков. Результаты исследований доказывают возможность эффективного представления описаний в виде вектора с целочисленными элементами. Данный подход может использоваться для решения задач распознавания и поиска изображений.

В результате построено компактное векторное описание эталонов, получены количественные оценки ошибки кластеризации, подтверждена работоспособность методов для прикладной базы изображений.

Ключевые слова: компьютерное зрение, распознавание изображений, характерные признаки, структурное описание изображения, метод SURF, кластеризация, нейронная сеть, метод разностного группирования, сеть Кохонена, ошибка квантования.

НОМЕНКЛАТУРА

SOM – Self-Organizing Map (самоорганизующиеся карты);

SURF – Speeded up robust features (ускоренное устойчивое выделение особенностей);

WTA – Winner takes all (победитель получает все);

ПФ – пиковая функция;

РГ – разностное группирование;

СК – сеть Кохонена;

ХП – характерный признак;

$\rho(z, c_i)$ – расстояние между z и c_i;

r – константа, которая определяет «сферу соседства»;

b – показатель степени многомерной функции Гаусса;

s – мощность множества Z;

ε – доля от значения пика $D(c_1)$;

$z \in C_i$ – точка данных кластера C_i с центром c_i;

k – число кластеров;

S – мощность обучающего множества Z;

$0 < \eta < 1$ – коэффициент обучения;

$z(h)$ – входной вектор $z \in Z$ на шаге h.

ВВЕДЕНИЕ

Современные системы компьютерного зрения нацелены на автоматизацию решения актуальных прикладных задач искусственного интеллекта: распознавание человеческих лиц в целях идентификации; поиск визуальных объектов заданного вида в базах (коллекциях) изображений; распознавание объектов на изображениях сцен, включая условия влияния перекрытий объектов, фона и помех; идентификация и определение координат движущихся объектов в видео-потоке. Перспективным подходом к построению систем распознавания в таких сложных ситуациях есть структурный анализ [1].

Процесс извлечения знаний в системах интеллектуальной обработки условно трактуют как ряд этапов [2–5]:

1. Отбор и предварительная обработка данных.
2. Редукция/проекция данных.
3. Поиск, оценка и интерпретация закономерностей.
4. Применение приобретенных знаний.

Способность к обучению и обобщению накопленных знаний (этап 3) считают одним из критериев, определяющих уровень интеллектуальности системы в рамках современной теории интеллектуальных вычислений. Обучение позволяет не только адаптировать распознавание к имеющимся данным и конкретным условиям, но и выявляет новые закономерности в целях дальнейшего обобщения знаний и улучшения результативности. В ходе обучения часто определяются ключевые признаки, в наибольшей степени отличающие образы объектов, что в целом систематизирует пакет признаков и дает возможность образовать надежный фундамент для принятия качественных решений о классификации образов. Система распознавания совершенствует свою эффективность через процесс обучения на основе данных из окружающей среды [2].

1 ПОСТАНОВКА ЗАДАЧИ

Целью статьи есть изучение возможности и оценивание эффективности применения методов разностного группирования и аппарата сетей Кохонена для осуществления результативного самообучения системы струк-

DOI 10.15588/1607-3274-2016-2-11

турного распознавания изображений в плане построения кластерного компрессионного представления данных в пространстве признаков прикладной базы изображений. За счет кластеризации-классификации на множестве структурных элементов обеспечивается переход к векторному представлению эталонов, что значительно сокращает объем вычислительных затрат и способствует улучшению быстродействия распознавания.

Задачи исследования – изучение особенностей и усовершенствование информационных технологий самообучения применительно к описаниям в виде множеств дескрипторов структурных признаков изображений, а также оценивание качества кластеризации на прикладных образцах.

2 ОБЗОР ЛИТЕРАТУРЫ

Обучение и самообучение, которые находят применение в структурном распознавании изображений, удачно осуществляют систематизацию признаков путем кластер-анализа (этап 2), что способствует снижению размерности признакового пространства [5–8]. Данные, сгруппированные в рамках кластера, представляются его центром и размером, что пропорционально сокращает как объем памяти, так и вычислительные затраты на распознавание [1].

Для успешной кластеризации требуется инициализация – тщательная предварительная обработка данных с целью подготовки начальных условий функционирования методов обучения. Например, для группирования данных необходимо установить число кластеров и начальные значения их центров в пространстве данных. Затем метод обучения, например, сеть Кохонена, отталкиваясь от этих значений, уточняет и завершает группировку объектов. Заметим что, k-степени зависят от начальных установок. Задачи инициализации могут быть решены методами пикового и разностного группирования [3], которые можно считать самостоятельными средствами получения знаний. В сравнении с возможностями аппарата сетей Кохонена эти методы позволяют автономно осуществить самоорганизацию центров образовавшихся кластеров с оцениваемой точностью.

В литературе описан ряд теоретических и практических аспектов построения универсальных методов самоорганизации данных, таких как разностное группирование [3] и сети Кохонена [3, 5, 9]. Основы теории кластеризации изложены в работах [6, 7]. В то же время теория и практика применения этого аппарата к множествам структурных признаков в целях распознавания визуальных объектов на изображениях только начинают свое развитие [8]. Особый интерес представляет изучение специфики технологий обучения и выбор наиболее подходящих среди них. Важно также оценить влияние ошибки обучения на результат распознавания с применением модифицированного пространства признаков.

3 МАТЕРИАЛЫ И МЕТОДЫ

Разностное группирование (РГ) наиболее эффективно в применении к векторам большой размерности, которая для структурных признаков SURF равна 64 и выше [4]. Рассмотрим особенности РГ применительно к пакетам векторов-дескрипторов SURF.

Структурное описание изображения – это конечное множество $Z \subset R_1^n$, $R_1^n = \{z \mid z \in R^n, \|z\| \approx 1\}$, где $R_1^n \subset R^n$ – пространство n-мерных векторов с евклидовой нормой: $\|z\| = \sqrt{\sum_{k=1}^{n} z_k^2} \approx 1$ [1]. Выполнение условия нормировки к 1 позволяет напрямую применять вектора описания в процедурах обучения без дополнительной обработки.

В методе РГ кластеры строятся последовательным урезанием исходного множества способом агломерации. Вначале для каждого $z_i \in Z$ осуществляют вычисление значения пиковой функции (ПФ)

$$D(z_i) = \sum_{j=1}^{s} \exp\left\{\frac{-\rho^{2b}(z_i, z_j)}{(r/2)^2}\right\}, \qquad (1)$$

Идея РГ лежит в русле развития метода потенциальных функций как одного из наиболее общих подходов к классификации образов [6,7]. Значение (1) пропорционально числу векторов из окрестности центра z_i. Малое значение $D(z_i)$ свидетельствует, что центр располагается в зоне сосредоточения незначительной группы векторов z_j. Считается, что коэффициент r практически не оказывает влияния на итоговые пропорции между $D(z_i)$, поэтому его подбор не критичен [3].

После расчета значений ПФ $\forall z_i \in Z$ отбирается вектор z с наибольшей мерой $D(z)$. Эта точка – первый центр: $c_1 = \arg\max_{z \in Z} D(z)$. Перед поиском следующего центра исключаем c_1 из Z и все точки в пределах сферы r. Все они образуют первый кластер.

Отбор элементов сферы соседства для кластера C_1 можно осуществить, например, путем анализа близости значений ПФ:

$$C_1 = \{z \in Z \mid D(c_1) - D(z) \le \varepsilon D(c_1)\}, \qquad (2)$$

Далее переопределяем ПФ для оставшихся точек:

$$D_{new}(z_i) = D(z_i) - D(c_1)\exp\left\{\frac{-\rho^{2b}(z_i, c_1)}{(r_1/2)^2}\right\}, \qquad (3)$$

ПФ $D_{new}(z_i)$ принимает нулевое значение при $z_i = c_1$ и близка к нулю для элементов C_1.

После модификации (3) определяется следующая точка z с максимальным $D_{new}(z)$, $z \in Z$, $z \notin C_1$. Она образует центр c_2. Поиск следующего центра возобновляется после исключения компонентов, включенных в уже отобранные кластеры. Кластеризация завершается при фиксации всех центров, предусмотренных начальными условиями.

Метод РГ реализует самоорганизацию множества векторов, суть которой – нахождение центров, представляющих множество данных с минимальной погрешностью. Заметим, что метод РГ инвариантен к нумерации списка входных точек, что важно для задач обработки изображений.

Критерием качества кластеризации (погрешность квантования) выступает функционал усредненной по

числу записей суммы квадратов расстояний между центрами кластеров и включенными в них данными [3]

$$E = \frac{1}{s}\sum_{i=1}^{k}\sum_{z\in C_i}\rho^2(z,c_i), \qquad (4)$$

Значение (4) – это усредненная ошибка в течение s шагов обучения. Выражение (4) можно применить как ко всей базе, так и к отдельному эталону. Если в (4) убрать усреднение – получим величину мгновенной ошибки на очередном шаге обучения.

Важно не только значение ошибки, но и ее динамика изменения в процессе обучения. В частности, представляет интерес событие стабилизации значений центров, что соответствует незначительным колебаниям (4). Методы, минимизирующие (4), называют группировкой с минимальной дисперсией [6]. Оптимизация критерия (4) в процессе кластеризации вскрывает внутреннюю структуру пространства данных в виде концентрированных сгущений точек.

Самоорганизующиеся карты (SOM – Self-Organizing Map) Кохонена (СК) реализуют конкурентное обучение на основе нейробиологического принципа по схеме WTA (Winner Takes All – победитель получает все) [9]. СК – однослойная искусственная нейронная сеть с прямой передачей информации. Преимуществами СК, кроме объединения операций кластеризации и проецирования, считаются простота архитектуры и независимость самоорганизации от размерности задачи [9]. В результате самообучения СК пространство Z разбивается на k подобластей $Z_v \subset Z$ так, что вектор-образу $z \in Z_v \subset Z$ соответствует точка-нейрон пространства $Y : Z = \cup_{v=1}^{k} Z_v$, $Z_v \cap Z_\tau = \varnothing, v \neq \tau$.

Схема соединения и трансформации нейронов в пространстве признаков SURF показана на рис. 1.

Результаты классификации-кластеризации посредством СК в значительной степени зависят от начальных весов нейронов, которые зачастую инициализируют случайным образом. Одним из путей уменьшения степени эвристичности есть применение непосредственно векторов из обучающей выборки [3, 8].

Отметим существенное отличие принципов построения методов РГ и СК: РГ выбирает центры кластеров из имеющегося списка, в то время как СК создает и подстраивает сеть центров под имеющиеся данные. Один из вариантов коллективного использования СК и РГ предполагает применение на первом этапе РГ для инициализации и определения таких параметров сети, как центры и число кластеров. На втором этапе употребляется СК для завершения самоорганизации нейронов. Обучающее множество используется повторно на каждом шаге обработки.

Формализуем обучение СК для распознавания. Имеем конечное множество $Z = \{Z^i\}_{i=1}^{J}$ структурных описаний базы из J эталонов. Эталон Z^i – это конечное множество дескрипторов SURF. Поставим задачу осуществить кластеризацию Z на k кластеров. В качестве обучающей выборки используем множество Z, а нейроны $W = (W_1, W_2, ..., W_k)$ (центры кластеров) представим пакетом из k векторов $\{W_i = (W_{i,1}, W_{i,2}, ..., W_{i,64})\}, i = 1, k$.

По принципу обработки WTA на очередном шаге для обучающего вектора $z \in Z$ определим номер q нейрона-победителя:

$$q = \arg\underset{i}{opt}\,\rho(W_i, z). \qquad (5)$$

Подмножество $W^* \subseteq W$ соседей W_q определяется как $W^* = \{W_i \mid \rho(W_i, z) \le r\}$, где r – порог. В обучении по Кохонену применяется линейная схема подстройки подмножества W^* в направлении вектора z [3]

$$W_i(h+1) = W_i(h) + \eta(h)[z(h) - W_i(h)]. \qquad (6)$$

Выражение (6) отвечает градиентному методу оптимизации, а значение $[z(h) - W_i(h)]$ определяет направление в многомерном пространстве, в соответствии с которым осуществляется уточнение вектора весов. Для случая нормализованных к единице входных векторов определение нейрона-победителя по максимуму скалярного произведения равнозначно критерию наименьшего евклидова расстояния. Следствием конкуренции становится самоорганизация в ходе обучения, а нейроны-победители приобретают свойство различения «своей» категории входных данных.

Метод самообучения Кохонена включает следующие шаги [3, 5].

1. Инициализация. Для k нейронов сети устанавливаются начальные нормализованные веса, скорость η и радиус r обучения.

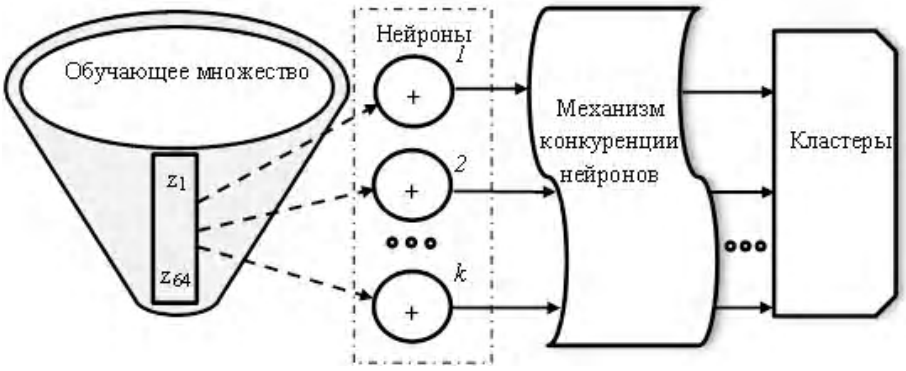


Рисунок 1 – Схема обучения СК на множестве Z

2. Возбуждение. На входной слой подается вектор z обучающей выборки.

3. Конкуренция. Для каждого нейрона W_v вычисляется расстояние $\rho[W_v, z]$, определяется нейрон-победитель.

4. Определяется подмножество нейронов в пределах радиуса обучения, их веса подстраиваются в соответствии с (6).

5. Коррекция параметров.

Натренированная сеть функционирует как векторный классификатор, основанный на сопоставлении входного вектора с набором сформированных векторов-нейронов [3]. Если обучение осуществить на группе похожих векторов с коэффициентом $\eta < 1$, то веса нейронов примут значения, усредненные по этим векторам. В режиме классификации СК лучше всего реагирует на векторы, близкие к средним значениям векторов обучающей группы. Сеть реализует функцию векторного квантования: за счет самоорганизации произвольное количество многомерных данных, образующих кластер, отображается вектором весов центра. Пространственное размещение нейронов определяет зоны концентрации данных. Представление структурного описания базы в виде конечного множества центров кластеров – это компрессия данных с потерями, которая сопровождается определенной погрешностью квантования.

Результат кластеризации для множества ХП прикладной базы изображений всецело определяется используемыми технологиями обучения, которые в настоящее время строго не формализованы и существенно зависят от опыта исследователя. Считается, что независимо от способа обучения самоорганизующихся сетей значима избыточность обучающих данных, без которой качественное обучение невозможно [3].

Процесс обучения сводится к формированию кластеров с фиксированным механизмом конкуренции нейронов при подборе параметров k, η, r. Например, для предотвращения расходимости обучения вблизи локального минимума отслеживают значение целевой функции $E(h)$ на шаге h с тем, чтобы не допустить ее возрастания сверх фиксированного ограничения, например в 5% [3]. Если выполнено условие $E(h+1) < 1{,}05E(h)$, то изменения игнорируют, приращение считается несущественным, веса не изменяются. Если же $E(h+1) \geq 1{,}05E(h)$, очередной шаг считается целесообразным, и уточнение проводится. Подбор коэффициентов требует значительного числа экспериментов и часто зависит от специфики проблемы.

Значимым моментом для обеспечения высокой статистической точности есть возможность многократного использования данных с изменяющимися параметрами обучения. Множество обучающих векторов предъявляют несколько раз, причем в случайной последовательности (бутстреп-обучение [9]). Технология многократного предъявления выборки способствует стабилизации весов и ошибки сети. Обучение завершают, если изменение вектора весов становится меньше принятого значения толерантности.

Грамотный подбор коэффициента η, как правило, изменяющегося в ходе обучения, оказывает огромное влияние на сходимость к минимуму целевой функции. Слишком малое значение не позволяет быстро минимизировать целевую функцию и требует многократных итераций. Большой шаг может привести к «перепрыгиванию» через минимум и повторных возвращений к нему. Практически приемлемые результаты в плане сходимости достигаются при с течением обучения [9]. Фиксация на весь период обучения упрощает обработку. В то же время более эффективный метод основан на адаптивном подборе коэффициента с учетом фактической динамики величины целевой функции.

Если обозначить погрешности соответственно e_{i-1}, e_i на последующих шагах обучения, а η_{i-1}, η_i – коэффициенты обучения, то в случае $e_i > \gamma e_{i-1}$ (γ – коэффициент допустимого роста погрешности) значение η уменьшают по формуле $\eta_{i+1} = \alpha\eta_i$, где α – коэффициент уменьшения. Если же $e_i \leq \gamma e_{i-1}$, то осуществляют увеличение $\eta_{i+1} = \beta\eta_i$, где β – коэффициент увеличения. Такой адаптивный метод подбора η сильно зависит от вида целевой функции и значений α, β, γ. Оптимальные для функций одного вида значения могут замедлить процесс обучения для других функций. В одной из задач квадратичной аппроксимации удачно использованы значения коэффициентов $\gamma = 1{,}4$, $\alpha = 0{,}7$, $\beta = 1{,}05$ [3]. Другие виды адаптации предполагают задание прямой зависимости снижения η от шага. Один из вариантов изменения η в течение N шагов может иметь вид $\eta(t) = 0{,}9(1 - t/N)$. Другой распространенный вариант – $\eta(t) = A/(t+B)$, где A, B – константы [3].

Иногда применяют разновидность обработки, где нейрон-победитель уточняется по формуле (6), а ближайшие к нему центры – в противоположном направлении с коэффициентом $-\eta$, что позволяет отдалить близкие центры и тщательно обследовать пространство данных. Исследователи отмечают, что ни один из алгоритмов не гарантирует абсолютную сходимость к глобальному оптимуму, обеспечивая лишь локальную оптимизацию, зависящую от начальных условий и параметров [3]. В итоге нейросетевые методы, включая SOM, не дают однозначно определяемых результатов [9]. Адаптивные процессы могут повести себя совершенно неожиданным образом в зависимости от выбранных значений их параметров и хода обучения. Следует использовать проверенные рекомендации и программные средства, чтобы обеспечить контроль над ходом самоорганизации и качество результатов в приложениях.

Обратим также внимание на тот факт, что ошибка обучения (4), вычисляемая пошагово в процессе обучения, не совпадает со значением, вычисленным после завершения обучения (пост-ошибка), т.к. нейроны-центры изменяются. Если для обучения значима динамика ошибки, то после обучения уже важна ее абсолютная величина как итог обучения. Для получения абсолютного значения необходимо пересчитывать (4) либо при каждом изменении нейрона-центра, либо по результату фиксированного числа шагов.

С точки зрения результативности классификации-кластеризации интересен экспериментальный анализ следующих стратегий обучения.

1. Сравнение качества кластеризации раздельно и совместно для двух обсуждаемых методов (РГ и СК) при фиксированном коэффициенте обучения.

2. Анализ качества методов при изменяющемся по некоторому закону коэффициенте обучения.

3. Изучение функционирования методов с применением адаптации и многократного употребления множеств входных данных.

Эксперименты проведены для структурных описаний, полученных методом SURF для базы из 25 изображений гербов городов Украины [8]. Моделирование метода РГ показало, что при постоянном $r = 1$ значение максимума ПФ монотонно снижается (рис. 2).

Моделирование для изображения герба Харькова (342 ХП) показало, что погрешность квантования (4) для метода РГ находится в диапазоне 0,4…0,5, в то время как для метода СК она составила значение 0,24. Рис. 3 содержит изображение герба Харькова и координаты 30 ХП из его описания SURF. Число образованных кластеров изменяется в пределах от 1 до 26 с изменением порога ε в соотношении (2) от 0,1 до 0,9. При малых ε наблюдается сосредоточение в 2-3-х кластерах, что впоследствии может сказаться на значении вероятности распознавания. По этой причине для данного набора ХП можно рекомендовать $\varepsilon = 0{,}7$, при котором образовано $q = 11$ кластеров, где элементы по кластерам распределены достаточно равномерно, а ошибка квантования составила 0,41.

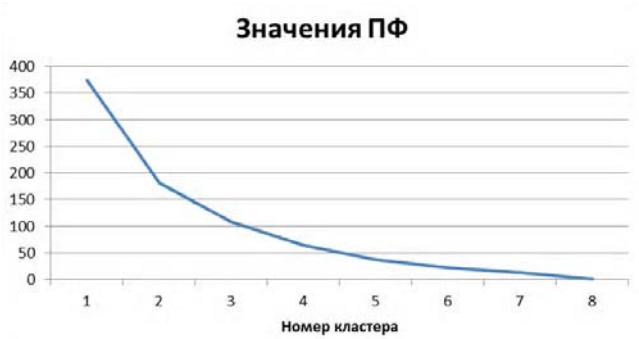


Рисунок 2 – Значения максимума ПФ в процессе кластеризации

Рисунок 3 – Изображение герба Харькова и координаты 30 ХП его описания

Исследования для пакета из 5 изображений гербов городов Украины [8], содержащей 1543 дескриптора SURF, практически подтверждают эти пост-экспериментальные рекомендации: при пороге $\varepsilon = 0{,}6$ получено 7 кластеров, при $\varepsilon = 0{,}7$ – 14 кластеров с более равномерным распределением элементов, ошибка квантования $E = 0{,}50$. Дальнейшее увеличение ε приводит еще к более Дальнейшее увеличение ε приводит еще к более равномерному распределению по кластерам, однако, увеличивается их число и ошибка: при $\varepsilon = 0{,}9$ значения $q = 42$, $E = 0{,}77$. Выбор оптимальных значений параметров переносится в область применений. Рисунок 4 демонстрирует зависимость числа кластеров от порога ε.

Исследования показали возможность управления параметром ε для достижения нужного числа кластеров методом РГ, которые впоследствии обеспечат результативное распознавание для произвольной базы изображений. Так для рассматриваемого пакета из 5 гербов при пороге $\varepsilon = 0{,}61$ имеем $q = 8$ кластеров с ошибкой $E = 0{,}47$.

Путем применения СК для базы из 5 гербов городов при пороге $\varepsilon = 0{,}61$ получено распределение по 8-ми кластерам с ошибкой $E = 0{,}236$.

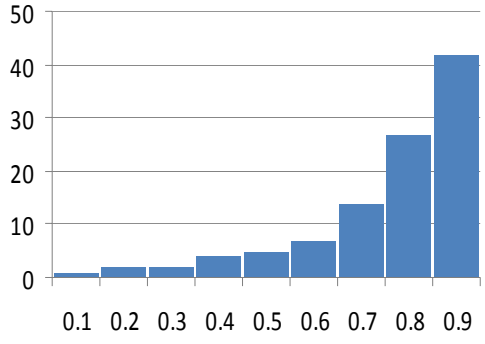


Рисунок 4 – Зависимость числа кластеров от порога кластеризации

Таблица 1 – Количества ХП в кластерном представлении эталонов

Изображения гербов	Номер кластера								Ошибка E
	1	2	3	4	5	6	7	8	
Днепропетровск	40	75	24	9	5	61	5	34	0,271
Львов	14	151	21	14	0	7	1	121	0,253
Киев	10	91	38	27	5	66	4	63	0,229
Харьков	102	113	20	17	4	14	3	69	0,256
Кременчуг	83	51	57	10	5	43	5	61	0,173
База	249	481	160	77	19	191	18	348	0,236

Из табл. 1 (матрица данных) наблюдаем практически равномерное распределение ошибки (4) по множеству эталонов и неравномерное распределение числа элементов по кластерам. Например, кластер 2 содержит 481 ХП (31,2% от общего числа элементов), а кластер 7 – всего 18 ХП, что соответствует 1,2%. В то же время есть мнение, что элементы малочисленных кластеров 5, 7 также влияют на результат распознавания. Альтернативой есть исключение элементов этих кластеров из эталонного описания базы.

Заметим, что в плане результативности распознавания, на наш взгляд, более значимую роль играет уровень существенного преобладания значений отдельных элементов в столбцах матрицы табл. 1. Это наблюдается, например, для кластеров 1–3, 8. Такое преобладание дает возможность с большей степенью уверенности или даже однозначно относить распознаваемый элемент к определенному эталону.

График экспериментально полученного значения ошибки (4) для метода СК при фиксированном $\eta = 0{,}5$ приведен на рис 5. Как видим, при $s = 300$ величина (4) равна 0,205, а в дальнейшем колеблется в среднем на уровне 0,23. Это характерно для разных вариантов начальных условий. Делаем вывод, что при числе элементов обучения более 300 процесс кластеризации дескрипторов стабилизируется. Ошибка в процессе и после кластеризации находится в пределах интервала 0,2–0,3.

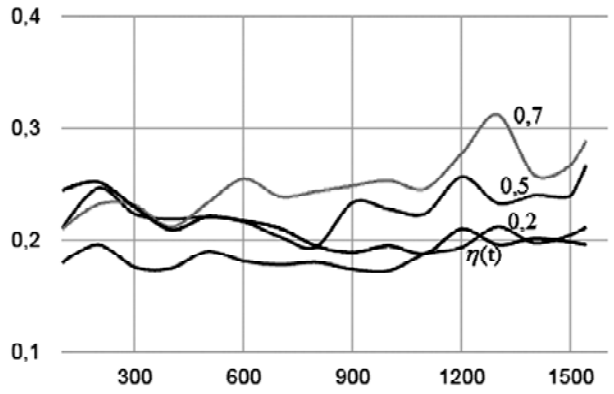


Рисунок 5 – Значения ошибки кластеризации при разных η

5 РЕЗУЛЬТАТЫ

Основным результатом работы есть формализация, моделирование и сравнение полученных экспериментальных оценок для современных методов кластеризации применительно к множествам характерных признаков для прикладной базы изображений.

Самоорганизация на основе сети Кохонена дает примерно в два раза меньшую ошибку квантования на множестве признаков, чем метод разностного группирования, который можно рекомендовать на этапе инициализации. Комбинирование методов разностного группирования и сети Кохонена снижает ошибку квантования в пределах 40%. Восемь кластеров вполне достаточно для осуществления уверенного распознавания в прикладной базе изображений гербов.

6 ОБСУЖДЕНИЕ

Наши эксперименты показали, что с изменением η от 0 до 1 ошибка возрастает. Конечное значение (4) при $\eta = 0{,}2$ составило 0,211, при $\eta = 0{,}5$ – 0,267, при $\eta = 0{,}7$–0,289. Кроме того, при постоянном η ошибка начинает медленно возрастать, примерно начиная с шага обучения с номером 1000.

Из графика рис. 5 видим, что изменение $\eta(t) = 0{,}9(1 - t/N)$ приводит к процессу уменьшения ошибки до значения 0,196. Это несколько меньше, чем при постоянном η, где наименьшее из значений равно 0,211.

Применение повторного обучения на материале базы ХП показало существенное снижение итоговой ошибки обучения от числа повторений. В соответствии с числом повторений экспериментальная ошибка составила: 1) 0,236; 2) 0,045; 3) 0,013; 4) 0,001; 5) 0,0002. Этот факт говорит об имеющихся резервных возможностях при необходимости уменьшать ошибку обучения, особенно для баз мало различающихся изображений, например, человеческих лиц.

Проведены эксперименты по адаптации процесса обучения путем отмены подстройки нейронов при условии $E(h+1) < 1{,}05E(h)$. Это реализовано для двух вариантов обучения: $\eta = 0{,}5$ и $\eta(t) = 0{,}9(1 - t/N)$. Моделирование показало, что в первом случае итоговая ошибка несколько уменьшилась (0,254 вместо 0,267), а во втором – возросла (0,243 вместо 0,196). Как видим, управление процессом обучения путем анализа значений текущей ошибки и игнорирования СК незначимых ее изменений в некоторых ситуациях может привести к возрастанию конечной ошибки.

Применение метода РГ для инициализации СК в сравнении со случайной инициализацией привело к снижению итоговой ошибки квантования (4) в методе СК до уровня 0,142, что соответствует около 40% улучшению. Это подтверждает необходимость и эффективность совместного применения разноплановых подходов кластеризации.

ВЫВОДЫ

Систематизация пространства характерных признаков в задаче распознавания изображений сводится к переходу к кластерному представлению с последующим применением трансформированной меры подобия в новом пространстве. Результирующее распределение признаков эталонов по кластерам определяет качество распознавания. Обучение дает возможность адаптировать структурный анализ к данным эталонного множества признаков прикладной базы изображений, что улучшает показатели распознавания.

Управление обучением путем изменения функцией коэффициента обучения снижает конечную ошибку. Применение адаптации нейронной сети путем анализа значений текущей ошибки не приводит к значимому уменьшению итоговой ошибки квантования. В целом процессу обучения для прикладной базы изображений свойственна стабилизация ошибки квантования при числе шагов, превышающем 65% общего числа признаков.

Ключевым критерием эффективности с использованием самообучения остается вероятность правильного распознавания, а ошибка обучения (векторного квантования) непосредственно отражает лишь качество и свойства самого процесса обучения.

Научная новизна исследования состоит в эффективном применении самообучения системы структурного распознавания изображений путем построения кластерного сжатого представления в пространстве признаков. Это позволяет перейти к векторному описанию пространства эталонов, и как результат, существенно увеличивается быстродействие распознавания.

Практическая ценность работы – получение экспериментальных оценок результативности кластеризации-классификации множества признаков для прикладных примеров баз изображений.

Перспективой обучения на множестве структурных описаний из характерных признаков может быть обучение с учителем, т.к. в рассматриваемой постановке класс характерного признака в составе эталона считается известным. Дальнейшее снижение ошибки квантования может быть получено путем построения нечеткой или гибридной сети [3].

СПИСОК ЛИТЕРАТУРЫ

1. Гороховатский В. А. Структурный анализ и интеллектуальная обработка данных в компьютерном зрении / В. А. Гороховатский. – Харьков : Компания СМИТ, 2014. – 316 с.
2. Контурная обработка динамических изображений / [Л. И. Тимченко, А. А. Поплавский, Н. И. Кокряцкая и др.]. – Киев : Наукова думка, 2013. – 239 с.
3. Осовский С. Нейронные сети для обработки информации / С. Осовский. – М. : Финансы и статистика, 2002. – 344 с.
4. Bay H. Surf: Speeded up robust features / H. Bay, T. Tuytelaars, L.Van Gool // Computer Vision : Ninth European Conference on Computer Vision, Graz, 7–13 May, 2006: proceedings. – Berlin : Springer, 2006. – P.404–417.
5. Паклин Н. Б. Бизнес-аналитика: от данных к знаниям / Н. Б. Паклин, В. И. Орешков. – СПб. : Питер, 2013. – 704 с.
6. Duda R. O. Pattern classification. Second edition / R. O. Duda, P. E. Hart, D. G. Stork. – New York : Wiley, 2000. – 738 p.
7. Прикладная статистика: Классификация и снижение размерности / [С. А. Айвазян, В. М. Бухштабер, И. С. Енюков, Л. Д. Мешалкин; под ред. С. А. Айвазяна.]. – М. : Финансы и статистика, 1989. – 607 с.
8. Берестовский А. Е. Нейросетевые технологии самообучения в системах структурного распознавания визуальных объектов / А. Е. Берестовский, А. Н. Власенко, В. А. Гороховатский // Реєстрація, зберігання і обробка даних. – 2015. – № 1. – С. 108–120.
9. Кохонен Т. Самоорганизующиеся карты / Т. Кохонен. – М. : БИНОМ, Лаборатория знаний, 2013. – 655 с.

Статья поступила в редакцию 02.12.2015.
После доработки 14.12.2015.

Гороховатський В. О.[1], Берестовський А. Е.[2], Передрій О. О.[3]

[1]Д-р техн. наук, професор, професор кафедри інформаційних технологій, Харківський навчально-науковий інститут ДВНЗ «Університет банківської справи», Харків, Україна

[2]Аспірант кафедри інформатики, Харківський національний університет радіоеелектроніки, Харків, Україна

[3]Канд. техн. наук, викладач кафедри інформатики і комп'ютерної техніки, Харківський національний економічний університет ім. Семена Кузнеця, Харків, Україна

СИСТЕМАТИЗАЦІЯ ПРОСТОРУ СТРУКТУРНИХ ОЗНАК НА ОСНОВІ МЕТОДІВ САМОНАВЧАННЯ З МЕТОЮ РЕЗУЛЬТАТИВНОГО РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ

Робота присвячена дослідженню питань кластеризації для множин характерних ознак зображень. Для побудови масиву характерних ознак використаний метод Speeded Up Robust Features. Реалізовані алгоритми кластеризації структурних описів зображень на основі самоорганізуючої нейронної мережі Кохонена та методу різницевого групування. Об'єктом дослідження є методи кластеризації стосовно до множин структурних ознак. Метою роботи є побудова векторних уявлень описів на основі кластеризації, що підвищує швидкодію розпізнавання. Предметом дослідження є систематизація множин структурних ознак візуальних об'єктів.

Обговорюються результати застосування методів кластеризації для структурних описів зображень у вигляді множин характерних ознак з метою підвищення швидкодії розпізнавання візуальних об'єктів. Для систематизації та стиснення простору ознак запропоновано здійснити самонавчання із застосуванням методів різницевого групування і мереж Кохонена.

Проведено моделювання та експериментальні дослідження методів кластеризації на прикладах конкретних множин характерних ознак. Результати досліджень доводять можливість ефективного представлення описів у вигляді вектора з цілочисельними елементами. Даний підхід може використовуватися для вирішення задач розпізнавання і пошуку зображень.

У результаті побудовано компактний векторний опис еталонів, отримані кількісні оцінки помилки кластеризації, підтверджена працездатність методів для прикладної бази зображень.

Ключові слова: комп'ютерний зір, розпізнавання зображень, характерні ознаки, структурний опис зображення, метод SURF, кластеризація, нейронна мережа, метод різницевого групування, мережа Кохонена, помилка квантування.

Gorokhovatsky V. A.[1], Berestovskyi A. E.[2], Peredrii E. O.[3]

[1]Dr.Sc., Professor, Professor of the information technologies department, Kharkiv Educational and Scientific Institute SHEI "The University of banking", Kharkiv, Ukraine

[2]Post-graduate student of the informatics department, Kharkiv National University of Radioelectronics, Kharkiv, Ukraine

[3]Ph.D., teacher of the department of computer science and computer engineering, Simon Kuznets Kharkiv National University of Economics, Kharkiv, Ukraine

SYSTEMATIZATION OF SPACE OF STRUCTURAL FEATURES BASED ON SELF-LEARNING METHODS FOR EFFECTIVE IMAGE RECOGNITION

The work deals with issues of clustering sets of characteristic features of images. For the construction of array of the characteristic features is used method Speeded Up Robust Features. Implemented algorithms for clustering structural descriptions of images on the

basis of a self-organizing Kohonen neural network and method of grouping the difference. The object of the research are clustering methods which applied to the set of structural features. The aim is to construct a vector representations of descriptions based on clustering, which increases the speed of recognition. The subject of research is systematization a set of structural features of visual objects.

Discussing the results of the application of clustering methods for structural descriptions of images in the form of sets of characteristic features to improve the performance of visual recognition of objects. For systematization and compression the feature space proposed to carry out self-study using the methods of differential grouping and Kohonen networks.

The simulation and experimental study of clustering methods on examples of specific sets of characteristic features were done. The research results proves the possibility of effective representation of the descriptions in the form of a vector with integer elements. This approach can be used to solve problems of recognition and retrieval of images.

As a result compact vector description of etalon images is built, quantitative estimates of clustering error are estimated, efficiency of proposed method during processing of real image database is confirmed.

Keywords: computer vision, image recognition, characteristic signs, structural description of image, method SURF, clusterization, neural network, differential grouping method, Kohonen's neural network, quantization error.

REFERENCES

1. Gorokhovatsky V. Strukturny'j analiz i intellektual'naya obrabotka danny'x v komp'yuternom zrenii: monografiya. Kharkiv, Kompaniya SMIT, 2014, 316 p.
2. Timchenko L. I., Poplavskij A.A., Kokryackaya N. I. i dr.Konturnaya obrabotka dinamicheskix izobrazhenij : monogr. Kyiv, Naukova dumka, 2013, 239 p.
3. Osovskij S. [per. s pol'skogo] Nejronny'e seti dlya obrabotki informacii. Moscow, Finansy' i statistika, 2002, 344 p.
4. Bay H., Tuytelaars T., Van L. Gool Surf: Speeded up robust features, *Computer Vision : Ninth European Conference on Computer Vision, Graz, 7–13 May, 2006: proceedings*. Berlin, Springer, 2006, pp.404–417. DOI: 10.1007/11744023.
5. Paklin N. B., Oreshkov V. I. Biznes-analitika: ot danny'x k znaniyam: uch posob. Sankt-Peterburg, Piter, 2013, 704 p.
6. Duda R. O., Hart P. E., Stork D. G. Pattern classification. Second edtition. N'ju-Jork, Wiley-Interscience, 2000, 738 p.
7. Ajvazyan S. A., Buxshtaber V. M., Enyukov I. S., Meshalkin L. D.; pod red. S. A. Ajvazyana. Prikladnaya statistika: Klassifikaciya i snizhenie razmernosti: sprav. izd. Moscow, Finansy' i statistika, 1989, 607 p.
8. Berestovskyi A., Vlasenko A. N., Goroxovatskij V. A. Nejrosetevy'e texnologii samoobucheniya v sistemax strukturnogo raspoznavaniya vizual'ny'x ob"ektov, *Reestraciya, zberigannya i obrobka danix*, 2015, Vol. 17, No. 1, pp. 108–120.
9. Koxonen T. Samoorganizuyushhiesya karty' per. 3-go angl. izd. Moscow, BINOM. Laboratoriya znanij, 2013, 655 p.

УДК 004.021

Дичка І. А.[1], Онай М. В.[2], Дрозда Т. П.[3]

[1]Професор, доктор технічних наук, декан факультету прикладної математики НТУУ «КПІ», Київ, Україна
[2]Старший викладач кафедри програмного забезпечення комп'ютерних систем факультету прикладної математики НТУУ «КПІ», Київ, Україна
[3]Магістрант кафедри програмного забезпечення комп'ютерних систем факультету прикладної математики НТУУ «КПІ», Київ, Україна

МОДИФІКОВАНИЙ ВІКОННИЙ МЕТОД ОДНОКРАТНОГО МНОЖЕННЯ ТОЧКИ ЕЛІПТИЧНОЇ КРИВОЇ НА СКАЛЯР У ПОЛІ *GF(P)*

При реалізації багатьох криптографічних додатків виникає потреба у швидких алгоритмах множення точки еліптичної кривої на число. У даній статті запропоновано модифікований віконний метод однократного множення точки еліптичної кривої на скаляр у полі GF(p). Об'єктом дослідження є процеси виконання операцій у еліптичних криптосистемах. Предметом дослідження є методи та алгоритми виконання операцій однократного множення точки еліптичної кривої на число у полі GF(p). Метою даного дослідження є розроблення та оптимізація методів і алгоритмів виконання операції множення точки еліптичної кривої на скаляр у полі GF(p) для поліпшення часових характеристик. Існуючі та запропоновані алгоритми реалізовані на мові програмування C# у середовищі розробки Visual Studio 2013. У даній статті проведено дослідження існуючих алгоритмів скалярного множення точки еліптичної кривої та розроблено три модифікації LR-алгоритму віконного методу і узагальнену модифікацію. Експериментальні дослідження реалізованих алгоритмів проводились згідно запропонованої нами методики, яка дозволяє нівелювати вплив на результати дослідження множника та точки еліптичної кривої. Проведене експериментальне дослідження віконних методів та їх модифікацій показало збільшення швидкодії роботи модифікованих алгоритмів у порівнянні з існуючими в середньому на 13%.

Ключові слова: ЕОМ, еліптична криптографія, скалярне множення, таблиця передобчислень, еліптична крива, скінченне поле.

НОМЕНКЛАТУРА

ЕОМ – електронно-обчислювальна машина;

ЕК – еліптична крива;

GF(*p*) – Galois field, поле Галуа або скінченне поле, де p – просте число, що є кількістю елементів поля;

LR – left-to-right;

RL – right-to-left;

NAF – a non-adjacent form;

wNAF – a window non-adjacent form;

p – модуль;

a, b – параметри еліптичної кривої;

x, y – змінні еліптичної кривої;

k – множник, скаляр;

P, Q – точка еліптичної кривої;

w – довжина вікна;

n – бітова довжина скаляра k;

Table – таблиця передобчислень;

for – початок циклу з лічильником;

end for – кінець циклу з лічильником;

while – початок циклу з передуомовою;

end while – кінець циклу з передуомовою;

do – виконати;

to – до;

downto – вниз до;

if – умовний оператор;

then – гілка умовного оператору за якою виконується перехід, якщо умова справедлива;

else – гілка умовного оператору за якою відбувається перехід, якщо умова не виконується;

return – видати результат на вихід алгоритму;

$(k_{l-1},...,k_1,k_0)_r$ – l розрядів подання числа k у системі числення з основою r.

ВСТУП

Задача забезпечення конфіденційності інформації та захист її від зловмисників стає дедалі складнішою і в той же час найбільш актуальною. Більшість систем захисту інформації будується на основі асиметричних криптосистем. Одним з розділів криптографії, який вивчає асиметричні криптосистеми, що засновані на еліптичних кривих над скінченними полями, є еліптична криптографія. Вона бере свій початок ще з 80-х років XX ст. (запропонована Віктором Міллером і Нілом Кобліцем) [1, 2]. Перевага використання еліптичних кривих в криптографічних цілях базується на складності розв'язання задач дискретного логарифмування у групі точок еліптичною кривою. Еліптична криптографія забезпечує набагато вищий рівень криптостійкості використовуючи ключі, що мають меншу довжину у порівнянні з іншими популярними криптосистемами, що засновані на факторизації цілих чисел та проблемі дискретного логарифмування у мультиплікативній групі кільця лишків за певним модулем. У даному дослідженні буде розглянута еліптична крива над полем *GF*(*p*). Елементами цього поля є цілі додатні числа від 0 до $p-1$, де p – модуль. У випадках коли p не дорівнює 2 або 3 буде-яке рівняння еліптичної кривої можна звести до форми Веєрштраса [3]. Тому дана робота ґрунтуватиметься на несингулярній еліптичній кривій у формі Веєрштраса: $y^2 = x^3 + ax + b$, де $p \neq 2,3$.

Найбільш обчислювально витратною операцією в еліптичних криптосистемах є операція множення точки еліптичної кривої на скаляр. У зв'язку з цим актуальною є задача прискорення роботи існуючих алгоритмів скалярного множення точки еліптичної кривої.

1 ПОСТАНОВКА ЗАДАЧІ

Прискорення операції скалярного множення на еліптичній кривій викликає інтерес багатьох дослідників у галузі криптографії [3–9]. З цією метою була запропонована низка методів виконання даної операції [4–9], що в основному полягають в представленні скаляра у деякій з відповідних форм та виконанні різних операцій на еліп-

DOI 10.15588/1607-3274-2016-2-12

тичній кривій (додавання, подвоєння, зменшення у два або три рази).

Множення цілого числа k на точку P еліптичної кривої можна представити наступним чином: $[k]P = \underbrace{P+P+...+P}_{k}$.

Зазвичай число k представляють у двійковому вигляді та за допомогою методів додавання і подвоєння знаходять $[k]P$. Для чисел великих порядків здійснення множення таким чином буде виконуватися занадто довго, тому актуальним є дослідження та модифікація віконних методів, які аналізують одночасно кілька розрядів подання множника у певній системі числення.

Таким чином, метою даного дослідження є модифікація віконного алгоритму виконання операції множення точки еліптичної кривої на скаляр у полі $GF(p)$ для поліпшення часових характеристик алгоритму.

2 ОГЛЯД ЛІТЕРАТУРИ

Вперше ідея подання скаляра у певній системі числення була використана Дональдом Кнутом [3], для мультиплікативної групи, тобто для піднесення до степеня відносно операції множення. Пізніше Хенкерсон та Менезес [4] використали цю ідею для побудови алгоритмів піднесення до степеня в адитивній групі. Скалярне множення точки еліптичної кривої є піднесенням до степеня в адитивній групі точок еліптичної кривої.

Загалом всі алгоритми множення точки еліптичної кривої на число ґрунтуються на поданні множника у певній системі числення та розгляді розрядів цього подання зліва направо, тобто від старшого до молодшого – *LR* або справа наліво, від молодшого до старшого – *RL*.

В літературі [4–8] розглядаються різні модифікації віконних алгоритмів скалярного множення, які завдяки побудові таблиць передобчислень (precomputation table) на початкових стадіях алгоритмів дають хороші показники швидкодії.

Віконні методи отримали таку назву через те, що в них розглядається не по одному двійковому розряду, а по w розрядів, де w – довжина вікна. В роботі [6] Метью Рівайн розглядає два види віконних алгоритмів: *LR*- та *RL*-алгоритми.

Нехай k – деякий скаляр на який виконується множення точки еліптичної кривої, тоді його подання у системі числення за основою 2^w матиме вигляд

$$k = \sum_{i=0}^{l-1} k_i 2^{iw},$$

де $k_i \in \{0,1,...2^w - 1\}$, $k_{l-1} \neq 0$, $l = \left\lfloor \frac{n}{w} \right\rfloor$ та n – бітова довжина скаляра k.

Віконний *LR*-алгоритм полягає у виконанні обчислень за наступними формулами:

$$T_{l-1} = [k_{l-1}]P$$

$$T_i = \left[2^w\right] T_{i+1} + [k_i]P, i = l-2..0.$$

Після проведення обчислень за цими формулами значення $[k]P$ буде знаходитися у T_0.

Віконний *RL*-алгоритм передбачає виконання обчислень за такою формулою:

$$[k]P = \sum_{i=0}^{l-1} \left[k_i \cdot 2^{iw}\right] P.$$

Для обох алгоритмів на початковій стадії будується таблиця передобчислень, яка складається з $2^w - 1$ елементів.

Алгоритм 1 – Бінарний віконний *LR*-алгоритм

Вхід: $P \in E(GF(p)), k = (k_{l-1},...,k_1,k_0)_{2^w} \in \mathbb{N}$

Вихід: $Q = [k]P$

1. *for* $i = 1$ *to* $2^w - 1$ *do*
1.1. $Table[i] \leftarrow i \cdot P$
2. *end for*
3. $Q \leftarrow 0$
4. *for* $i = l-1$ *downto* 0 *do*
4.1. $Q \leftarrow 2^w \cdot Q$
4.2. *if* $k_i > 0$ *then* $Q \leftarrow Q + Table[k_i]$
5. *end for*
6. *return* Q

Алгоритм 2 – Бінарний віконний *RL*-алгоритм

Вхід: $P \in E(GF(p)), k = (k_{l-1},...,k_1,k_0)_{2^w} \in \mathbb{N}$

Вихід: $Q = [k]P$

1. *for* $i = 1$ *to* $2^w - 1$ *do*
1.1. $Table[i] \leftarrow i \cdot P$
2. *end for*
3. $Q \leftarrow 0$
4. *for* $i = 0$ *to* $l-1$ *do*
4.1. *if* $k_i > 0$ *then* $Q \leftarrow Q + Table[k_i]$
4.2. $Table \leftarrow 2^w \cdot Table$
5. *end for*
6. *return* Q

Наприклад, таблиця передобчислень для вікна довжиною 3, тобто $w = 3$ матиме вигляд (P – точка еліптичної кривої):

№ елементу	Значення точки
1	$001 \cdot P$
2	$010 \cdot P$
3	$011 \cdot P$
4	$100 \cdot P$
5	$101 \cdot P$
6	$110 \cdot P$
7	$111 \cdot P$

При аналітичному аналізі алгоритмів 1 та 2 стає зрозумілим, що *RL*-алгоритм буде повільнішим за *LR*-алгоритм, оскільки у *RL*-алгоритмі на кожній ітерації циклу виконується переобчислення значень, що записані у таблицю, тому далі ми будемо розглядати лише *LR*-алгоритми реалізації віконних методів.

Наступним методом скалярного множення точки ЕК, що розглядається в роботі [8] є метод з пересувним вікном. Даний метод дістав таку назву через те, що згідно цього методу необхідно виділяти вікно довжиною w біт тільки якщо старший біт дорівнює 1, в іншому випадку виконують дії передбачені відповідним бінарним алгоритмом.

При розробці віконного методу з пересувним вікном ставилось за мету досягти компромісу між кількістю додавань і подвоєнь точки. На початковій стадії цього методу будується таблиця передобчислень, що містить елементи $[t]P$ для $t = \{2^{w-1}, 2^{w-1}+1, ..., 2^w - 1\}$. *LR*-представлення для скаляра k буде мати наступний вигляд: $k = k_0 + 2k_1 + 2^2 k_2 + ... + 2^m k_m$.

Алгоритм 3 – Бінарний *LR*-алгоритм з пересувним вікном

Вхід: $P \in E(GF(p)), k \in \mathbb{N}$

Вихід: $Q = [k]P$

1. $Q \leftarrow 0, i \leftarrow \log_2 k$
2. *while* $i \geq 0$ *do*
2.1. *if* $(k_i = 0)$ *then* $Q = 2 \cdot Q$
2.2. *else*
2.2.1. *if* $i \geq w - 1$ *then*
a) $t \leftarrow (k_i, ..., k_{i-w+1})$
b) $Q \leftarrow 2^w \cdot Q$
c) $Q \leftarrow Q + t \cdot P$
2.2.2. *else*
a) Викликати бінарний *LR*-алгоритм
2.2.3 $i \leftarrow i - w$
3. *end while*
4. *return* Q

Одним з напрямків прискорення методів скалярного множення точки ЕК є переведення множника у *NAF* представлення [4]. Подання множника k у формі *NAF* виражається формулою $k = \sum_{i=0}^{l-1} k_i 2^i$, де $k_i \in \{0; \pm 1\}$ та $k_{l-1} \neq 0$.

Метод з поданням множника у вигляді *NAF*, є ефективнішим (при наявності *NAF*-розкладення множника) ніж звичайні віконні методи через те, що два сусідні розряди не можуть бути одночасно не нульовим, а це скорочує кількість операцій додавання і віднімання точки. Віконна реалізація даного методу дістала назву метод з поданням множника у вигляді *wNAF*. Подання множника у *wNAF* формі виражається формулою $k = \sum_{i=0}^{n-1} t_i 2^i$, де кожне ненульове t_i є непарним та таким, що $|t_i| < 2^{w-1}$, $t_{n-1} \neq 0$ і хоча б один з w послідовних біт є не нульовим.

На початковій стадії реалізації *LR*-алгоритму цього методу множення потрібно крім побудови таблиці передобчислень перевести скаляр у *wNAF* представлення використовуючи алгоритм 4.

Алгоритм 4 – Переведення додатного цілого числа k у *wNAF* подання

Вхід: позитивне ціле k, ширина вікна w

Вихід: *wNAF(k)*

1. $i \leftarrow 0$
2. *while* $k \geq 1$ *do*
2.1. *if* k є непарним *then*
2.1.1. $t_i \leftarrow k \bmod 2^w$
2.1.2. $k \leftarrow k - t_i$
2.2. *else* $t_i \leftarrow 0$
2.3. $k \leftarrow \frac{k}{2}, i \leftarrow i + 1$
3. *end while*
4. *return* $\{t_{n-1}, t_{n-2}, ..., t_1, t_0\}$

Алгоритм 5 – *LR*-алгоритм з поданням множника у вигляді *wNAF*

Вхід: позитивне ціле k, $P \in E(GF(p))$, ширина вікна w

Вихід: kP

1. Використати алгоритм 4 для обчислення $wNAF(k) = \sum_{i=0}^{n-1} t_i 2^i$
2. *for* $i = 1$ *to* $2^{w-1} - 1$ з кроком 2 *do*
2.1. $Table[i] = i \cdot P$
3. *end for*
4. $Q \leftarrow 0$
5. *for* $i = n - 1$ *downto* 0 *do*
5.1. $Q \leftarrow 2 \cdot Q$
5.2. *if* $t_i \neq 0$ *then*
5.2.1. *if* $t_i > 0$ *then* $Q \leftarrow Q + Table[t_i]$
5.2.2. *else* $Q \leftarrow Q - Table[t_i]$
6. *end for*
7. *return* Q

Таким чином метод з поданням множника у вигляді *wNAF* є аналогом віконного методу, але оперуючи знаком *NAF* подання дозволяє скоротити об'єм необхідної для зберігання таблиці передобчислень пам'яті вдвічі порівняно з класичним віконним методом. Враховуючи особливості *wNAF* подання цілих додатних чисел таблиця передобчислень буде складатися з елементів $[i]P$, де $i = 1, 3, ..., 2^{w-1} - 1$, тобто порівняно з бінарним віконним методом отримуємо скорочення об'єму таблиці передобчислень у чотири рази.

Логічним вдосконаленням методу з поданням множника у вигляді *wNAF* є метод з пересувним вікном та поданням множника у вигляді *NAF*. На відміну від попереднього алгоритму з пересувним вікном, де ми переводимо число k у *wNAF* представлення, у даному алгоритмі потрібно перевести його у *NAF* форму, використовуючи алгоритм 6, а потім виконувати скалярне множення точки еліптичної кривої за алгоритмом 7.

Алгоритм 6 – Переведення додатного цілого числа k у *NAF* подання

Вхід: позитивне ціле k

Вихід: $NAF(k)$

1. $i \leftarrow 0$
2. $w$hile $k \geq 1$ do
2.1. if k is odd then
2.1.2. $k_i \leftarrow 2-(k \bmod 4)$
2.1.3. $k \leftarrow k-k_i$
2.2. $e$lse $k_i \leftarrow 0$
2.3. $k \leftarrow \frac{k}{2}$
2.4. $i \leftarrow i+1$
3. *end while*
4. *return* $\{k_{l-1}, k_{l-2}, ..., k_1, k_0\}$

Алгоритм 7 – *LR*-алгоритм з пересувним вікном та поданням множника у вигляді *NAF*

Вхід: $P \in E\left(GF(p)\right), k \in \mathbb{N}$

Вихід: $[k]P$

1. Використати алгоритм 6 для обчислення $NAF(k) = \sum_{i=0}^{l-1} k_i 2^i$
2. *for* $i = 1$ *to* $(2^w - (-1)^w)/3 - 1$ з кроком 2 *do*
2.1. $Table[i] = i \cdot P$
3. *end for*
4. $Q \leftarrow 0, i \leftarrow l-1$
5. *while* $i \geq 0$ *do*
5.1. *if* $k_i = 0$ *then* $t \leftarrow 1, u \leftarrow 0$
5.1.1. *else* знайти $\max(t) \leq w$ таке що $u \leftarrow (k_i, ..., k_{i-t+1})$ є непарним
5.2. $Q \leftarrow 2^t \cdot Q$
5.3. *if* $u > 0$ *then* $Q \leftarrow Q + Table[u]$
5.4. *else if* $u < 0$ *then* $Q \leftarrow Q - Table[u]$
5.5. $i \leftarrow i - t$
6. *end while*
7. *return* Q

Таблиця передобчислень для цього методу буде містити такі елементи $[i]P$, де $i = 1, 3, ..., \frac{2(2^w - (-1)^w)}{3} - 1$.

Як зазначалося вище, час роботи віконних алгоритмів прискорюється завдяки побудові на першій стадії їх роботи таблиць передобчислень. У таблиці 1 наведено перелік точок, які необхідно обчислити на початковій стадії кожного з розглянутих алгоритмів.

Таблиця 1 – Наперед обчисленні точки для методів скалярного множення

Методи	Значення множника	Кількість наперед обчислених точок
Бінарний віконний метод	$\{1, 2, ..., 2^w - 1\}$	$2^w - 1$
Бінарний метод з пересувним вікном	$\{2^{w-1}, 2^{w-1}+1, ..., 2^w - 1\}$	$2^w - 2^{w-1}$
Метод з поданням множника у вигляді *wNAF*	$\{1, 3, 5, ..., 2^{w-1} - 1\}$	2^{w-2}
Метод з пересувним вікном та поданням множника у вигляді *NAF*	$\{1, 3, ..., \frac{2(2^w - (-1)^w)}{3} - 1\}$	$\frac{1}{3}(2^w - (-1)^w)$

3 МАТЕРІАЛИ ТА МЕТОДИ

Статистично показано, що двійкові подання чисел довжиною понад 100 біт містять довгі послідовності нулів, тому нами запропонована модифікація віконного методу згідно якої буде виділятися вікно зі старшим одинарним бітом та іншими нульовими. Дану модифікацію назвемо модифікація №1. При такому підході таблиця передобчислень буде складатися з таких елементів $\left[2^i\right]P$, де $i = 0, 1, .., w-1$.

Алгоритм 8 – Модифікований віконний *LR*-алгоритм № 1

Вхід: $P \in E\left(GF(p)\right), k \in \mathbb{N}$

Вихід: $[k]P$

1. $Table[0] \leftarrow P$
2. *for* $i = 1$ *to* $w-1$
2.1. $Table[i] \leftarrow 2 \cdot Table[i-1]$
3. *end for*
4. $Q \leftarrow 0; i \leftarrow \log_2 k$
5. *while* $i \geq 0$ *do*
5.1. $Q = 2 \cdot Q$
5.2. *if* $k_i = 1$ *then*
5.2.1. *if* $k_{i-1} = 0$ *then*
a. знайти $\max(t) \leq w$ таке що $k_{i-1} = k_{i-2} = ... = k_{i-t+1} = 0$
b. $Q \leftarrow 2^{t-1} \cdot Q$
c. $Q \leftarrow Q + Table[t-1]$
d. $i \leftarrow i - t$
5.2.2. *else*
a. $Q \leftarrow Q + P$
b. $i \leftarrow i - 1$
6. *end while*
7. *return* Q

Наприклад, таблиця передобчислень для $w = 5$ матиме вигляд (P – точка еліптичної кривої):

№ елементу	Множник	Значення точки
0	2^0	$1 \cdot P$
1	2^1	$10 \cdot P$
2	2^2	$100 \cdot P$
3	2^3	$1000 \cdot P$
4	2^4	$10000 \cdot P$

У випадку якщо множник k складатиметься переважно з нулів, такий спосіб побудови таблиці передобчислень суттєво збільшить швидкодію віконного методу. Але в протилежному випадку, коли біти скаляра будуть тільки одиничні такий метод не дасть хороших результатів. Тому актуальним є побудова нової таблиці множники якої будуть складатись тільки з одиничних біт. Такий метод назвемо модифікація №2. Множник k таблиці передобчислень в даному методі буде рівним $2^i - 1$, де $i = 1, 2, .., w$.

Алгоритм 9 – Модифікований віконний *LR*-алгоритм № 2

Вхід: $P \in E\left(GF(p)\right), k \in \mathbb{N}$

Вихід: $[k]P$

1. *for* $i = 1$ *to* w
1.1. $Table[i] \leftarrow (2^i - 1) \cdot P$
2. *end for*
3. $Q \leftarrow 0; i \leftarrow \log_2 k$
4. *while* $i \geq 0$ *do*
4.1. $Q = 2 \cdot Q$
4.2. *if* $k_i = 1$ *then*
4.2.1. знайти $\max(t) \leq w$ таке що $k_{i-1} = k_{i-2} = ... = k_{i-t+1} = 1$
4.2.2. $Q \leftarrow 2^{t-1} \cdot Q$
4.2.3. $Q \leftarrow Q + Table[t-1]$
4.2.4. $i \leftarrow i - t + 1$
4.3. $i \leftarrow i - 1$
5. *end while*
6. *return Q*

Для вікна $w = 5$ таблиця пердобчислень матиме вигляд:

8	Множник	Значення точки
1	$2^1 - 1$	$1 \cdot P$
2	$2^2 - 1$	$11 \cdot P$
3	$2^3 - 1$	$111 \cdot P$
4	$2^4 - 1$	$1111 \cdot P$
5	$2^5 - 1$	$11111 \cdot P$

Також можна сподіватися на приріст швидкодії, коли з розрядів числа k буде виділятися частина біт, що починається і закінчується одиницею, а між ними міститиметься певна кількість нулів (модифікація №3). Множник k у таблиці передобчислень в модифікації №3 буде рівним $2^{i-1} + 1$, де $i = 2, 3, .., w$.

Алгоритм 10 – Модифікований віконний *LR*-алгоритм № 3

Вхід: $P \in E\left(GF(p)\right), k \in \mathbb{N}$

Вихід: $[k]P$

1. $Table[1] \leftarrow P$
2. *for* $i = 2$ *to* w
2.1. $Table[i] \leftarrow (2^{i-1} + 1) \cdot P$
3. *end for*
4. $Q \leftarrow 0; i \leftarrow \log_2 k$
5. *while* $i \geq 0$ *do*
5.1. $Q = 2 \cdot Q$
5.2. *if* $k_i = 1$ *then*
5.2.1. *if* $k_{i-1} = 0$ *then*
a. знайти $\max(t) \leq w$ таке що $k_i = k_{i-t+1} = 1$ та $k_{i-1} = k_{i-2} = ... = k_{i-t+2} = 0$
b. $Q \leftarrow 2^{t-1} \cdot Q$
c. $Q \leftarrow Q + Table[t]$
d. $i \leftarrow i - t + 1$
5.2.2. *else*
a. $Q \leftarrow 2 \cdot Q$
b. $Q \leftarrow Q + Table[2]$
c. $i \leftarrow i - 1$
5.2.3. $i \leftarrow i - 1$
6. *end while*
7. *return Q*

Оскільки за допомогою виразу $2^{i-1} + 1$ неможливо отримати 1, то до початку циклу побудови таблиці передобчислень у елемент таблиці з індексом 1, необхідно записати значення точки P. Приклад такої таблиці пердобчислень для вікна $w = 5$:

№ елементу	Множник	Значення точки
1	1	$1 \cdot P$
2	$2^1 + 1$	$11 \cdot P$
3	$2^2 + 1$	$101 \cdot P$
4	$2^3 + 1$	$1001 \cdot P$
5	$2^4 + 1$	$10001 \cdot P$

У таблиці 2 наведено значення множників для таблиць передобчислень запропонованих модифікацій. Зрозуміло, що кожна із запропонованих модифікацій в певному частковому випадку буде збільшувати швидкодію віконного алгоритму скалярного множення.

Модифікації № 1–3 дають приріст швидкодії тільки в окремих випадках, тому актуальним є побудувати узагальнений модифікований віконний метод скалярного множення точки ЕК та об'єднати переваги кожної із запропонованих модифікацій таким чином, щоб прискорити віконний алгоритм для будь-якого вигляду двійкового представлення множника k. Доцільним буде об'єднувати таблиці передобчислень з модифікації №2 та №3, оскільки модифікація №1 покривається модифікацією №3. Особливістю побудови таблиць передобчислень для другої та третьої модифікації є те, що номер елемента у таблиці відповідає бітовій довжині множника для якого обчислене відповідне значення.

Алгоритм 11 – Узагальнений модифікований віконний *LR*-алгоритм скалярного множення

Вхід: $P \in E\left(GF(p)\right), k \in \mathbb{N}$

Вихід: $[k]P$

1. $Table1[1] \leftarrow P$
2. $Table2[1] \leftarrow P$
3. *for* $i = 2$ *to* w *do*
3.1. $Table1[i] \leftarrow (2^i - 1) \cdot P$
3.2. $Table2[i] \leftarrow (2^{i-1} + 1) \cdot P$
4. *end for*
5. $Q \leftarrow 0, i \leftarrow \log_2 k$

Таблиця 2 – Наперед обчислені точки для модифікованих віконних методів

Метод	Значення множника	Кількість наперед обчислених точок
Модифікація №1	2^i, де $i \in [0; w-1]$	w
Модифікація №2	$2^i - 1$, де $i \in [1; w]$	w
Модифікація №3	1 та $2^{i-1} + 1$, де $i \in [2; w]$	w

6. *while* $i \geq 0$ *do*
6.1. *if* $k_i = 1$ *then*
6.1.1. *if* $k_{i-1} = 0$ *then*
a. знайти $\max(t) \leq w$ такий, що $k_i = k_{i-t+1} = 1$ *and* $k_{i-1} = k_{i-2} = ... = k_{i-t+2} = 0$
b. $Q \leftarrow 2^t \cdot Q$
c. $Q \leftarrow Q + Table1[t]$
6.1.2. *else*
a. знайти $\max(t) \leq w$ такий, що $k_{i-1} = k_{i-2} = ... = k_{i-t+1} = 1$
b. $Q \leftarrow 2^t \cdot Q$
c. $Q \leftarrow Q + Table2[t]$
6.1.3. $i \leftarrow i - t$
6.2. *else*
6.2.1. $Q = 2 \cdot Q$
6.2.2. $i = i - 1$
7. *end while*
8. *return Q*

На кроці 11 при реалізації алгоритму 11 потрібно передбачити перевірку чи є праворуч від *i*-го біта одиничні біти. Якщо всі біти молодші *i*-го є нульовими, то необхідно виконати такі дії: $Q = 2 \cdot Q$, $Q = Q + P$, $i = i - 1$ та перейти на наступну ітерацію циклу *while*.

4 ЕКСПЕРИМЕНТИ

З метою проведення експериментальних досліджень було розроблено програмний продукт на мові програмування C# у середовищі розробки «Visual Studio 2013». Експериментальне дослідження проводилося на ЕОМ з операційною системою Windows 8.1, об'ємом оперативної пам'яті 2Gb, процесором Pentium Dual-Core 2,30 Hz.

Даний програмний продукт дозволяє проводити тестування коректності роботи алгоритмів та проводити дослідження розглянутих віконних методів скалярного множення точки ЕК. Завдяки тестуванню віконних алгоритмів для різних значень довжини вікна *w* було отримано оптимальне значення *w*.

У розробленому програмному продукті, окрім відомих алгоритмів реалізовано запропоновані нами модифікації віконних *LR*-алгоритмів, оскільки, як показало дослідження *LR*-алгоритми показують кращі результати ніж *RL*-алгоритми. Для збільшення швидкодії алгоритмів на початковій стадії роботи кожного з них будується таблиця передобчислень, значення таблиць передобчислень наведені у таблиці 1 та таблиці 2. При аналізі літературних джерел [5, 7] було встановлено, що оптимальними параметрами еліптичної кривої, які забезпечують високу крипостійкість є: $a = 79, b = -3$. Тому ці значення параметрів було використано у дослідженні.

Замір швидкодії роботи алгоритмів проводився за наступною методикою:

1) обирається довжина модуля *p* (64, 128, 256 та 512 біт);

2) для обраного модуля *p* формується множина з 25 випадкових точок еліптичної кривої;

3) випадковим чином генерується 25 множників довжиною від 32 до 64 біт;

4) виконується множення кожної точки з п. 2 на кожен множник з п. 3, отримані часові показники усереднюються.

Пошук оптимальних значень довжини вікна (рис. 1) виконувався для модуля *p* що має довжину 128 біт та довжин вікна $w \in [2;9]$, оскільки після збільшення довжини вікна до 10 біт час виконання алгоритмів зростає. Значення множників змінювалось від 10 до 1000 з кроком 100, оскільки для проведення дослідження при більших значеннях множника було не достатньо обчислювальної потужності комп'ютера.

5 РЕЗУЛЬТАТИ

За наведеною методикою для існуючих віконних *LR*- та *RL*-алгоритмів було побудовано таблицю 3 та для запропонованих модифікацій і узагальненого модифікованого віконного методу таблицю 4.

Як можна помітити з таблиці 3, *LR*-алгоритми дають значно кращі результати ніж *RL*- алгоритми, що підтверджує результати проведеного аналітичного дослідження, тому доцільним є пошук оптимального значення довжини вікна *w* для існуючих *LR*-алгоритмів.

На рис. 1 наведено залежність часу роботи *LR*-алгоритмів від довжини вікна, де 1, 3, 5, 7 – номери відповідних алгоритмів у таблиці 3.

Як видно з рисунку 1 найкращі часові характеристики показують алгоритми 1 і 3, а саме бінарний віконний *LR*-алгоритм та бінарний *LR*-алгоритм з пересувним вікном при довжині вікна 9 біт, що підтверджує доцільність їх модифікацій. Швидкодія роботи модифікованих віконних *LR*-алгоритмів та узагальненого алгоритму наведена в таблиці 4.

Таблиця 3 – Часові характеристики методів скалярного множення, мс

№	Методи	Довжина модуля, біт			
		64	128	256	512
1	Бінарний віконний *LR*-алгоритм	5,5	9,0	32,1	84,4
2	Бінарний віконний *RL*-алгоритм	54,7	120,8	403,6	1125,7
3	Бінарний *LR*-алгоритм з пересувним вікном	6,8	10,3	32,5	84,6
4	Бінарний *RL*-алгоритм з пересувним вікном	33,8	73,7	257,0	696,2
5	*LR*-алгоритм з пересувним вікном та поданням множника у вигляді *NAF*	4,3	8,6	30,5	84,2
6	*RL*-алгоритм з пересувним вікном та поданням множника у вигляді *NAF*	19,1	42,1	144,7	390,8
7	*LR*-алгоритм з поданням множника у вигляді *wNAF*	4,1	8,9	32,1	87,0
8	*RL*-алгоритм з поданням множника у вигляді *wNAF*	14,7	33,1	110,8	311,1

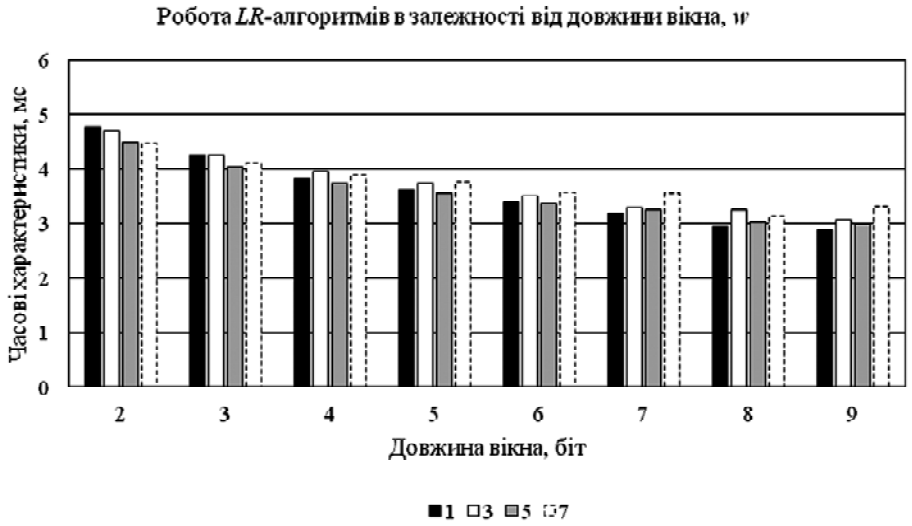


Рисунок 1 – Часові характеристики віконних *LR*-алгоритмів в залежності від довжини вікна

Таблиця 4 – Часові характеристики модифікованих методів в залежності від довжини модуля

№	Метод	Довжина модуля, біт			
		64	128	256	512
1	Модифікація №1	4,7	11,1	39,1	106,6
2	Модифікація №2	4,3	9,2	34,4	91,1
3	Модифікація №3	4,3	9,0	31,2	86,4
4	Узагальнений модифікований віконний метод	4,1	8,2	31,1	82,0

6 ОБГОВОРЕННЯ

За результатами наведеними у таблиці 3, найкращі часові характеристики для модулів будь-якої довжини показують *LR*-алгоритм з поданням множника у вигляді *wNAF* та бінарний *LR*-алгоритм з пересувним вікном і поданням множника у вигляді *NAF*. При правильно підібраній довжині вікна (рис. 1) та довжині модуля 128 біт найкращі часові характеристики показують віконний *LR*-алгоритм та *LR*-алгоритм з пересувним вікном. З таблиці 4 видно, що побудовані модифікації методу з пересувним вікном дають приріст швидкодії для модулів довжиною 64 біт на 20%. Модифікація №3 порівняно з попередніми модифікаціями покращує часові характеристики майже на 3% для модулів довжиною 128 і 256 біт.

Найефективнішою є побудова узагальненої модифікації віконного методу та використання таблиць передобчислень зі значеннями $[t]P$ для $t = 2^i + 1$, де $i = 1..w$ та $t = 2^i - 1$, де $i = 2..w$. Побудований *LR*-алгоритм на її основі дає приріст швидкодії в середньому на 13% для модулів 64, 128, 256 та 512 біт. Запропоновані таблиці передобчислень використовують порівняно з бінарними віконними методами у $\frac{2^w - 1}{w}$ разів менше оперативної пам'яті.

ВИСНОВКИ

У роботі вирішено актуальну задачу вдосконалення існуючих методів скалярного множення точки еліптичної кривої у полі *GF*(*p*).

Наукова новизна роботи полягає у тому, що дістав подальшого розвитку науковий підхід модифікації бінарних методів множення точки еліптичної кривої на число, що заснований на розгляді кількох розрядів двійкового подання множника одночасно на кожній ітерації циклу.

Проведений аналіз віконних методів множення точки еліптичної кривої на скаляр у полі *GF*(*p*) показав, що при правильному виборі довжини вікна та побудові на початковій стадії таблиць передобчислень можна суттєво збільшувати швидкодію алгоритмів. За допомогою розробленого програмного забезпечення для аналізу і тестування методів множення точки ЕК на скаляр, проведено експериментальне дослідження, яке довело практичну доцільність використання запропонованих методів, замість існуючих.

Розроблені три модифікації віконного *LR*-алгоритму з пересувним вікном, які відрізняються від існуючого методу способом побудови таблиць передобчислень, забезпечують приріст швидкодії для множників спеціальної структури. Побудована узагальнена модифікація віконного *LR*-алгоритму з пересувним вікном, що полягає у комбінації, на початковій стадії роботи алгоритму, двох таблиць передобчислень з модифікації №2 та №3, яка забезпечує приріст швидкодії порівняно з існуючими методами в середньому на 13%.

Практична цінність отриманих результатів полягає у тому, що розроблено програмне забезпечення, яке реалізує запропоновані модифікації та існуючі методи множення точки еліптичної кривої на число і дозволяє проводити аналіз алгоритмів, що реалізують зазначені методи. За допомогою даного програмного забезпечення може бути вирішена практична задача вибору найкращого алгоритму скалярного множення точки еліптичної кривої для використання у алгоритмі цифрового підпису на еліптичних кривих, який широко використовується.

Перспективою для подальшого дослідження є побудова модифікованого методу з пересувним вікном та поданням множника у вигляді *NAF* та модифікованого методу з поданням множника у вигляді *wNAF*.

ПОДЯКИ

Дослідження виконано у межах держбюджетної науково-дослідної теми «Розроблення та дослідження високоефективних архітектур спеціалізованих комп'ютерних систем для реалізації обчислень у скінченних полях» Національного технічного університету України «Київський політехнічний інститут» (номер державної реєстрації 0115U000319).

СПИСОК ЛІТЕРАТУРИ

1. Miller V. Use of elliptic curves in cryptography / V. Miller // Lecture Notes in Computer Science. Advances in cryptology – CRYPTO 85. – Springer, 1986. – P. 417–426. 10.1007/3-540-39799-X_31
2. Koblitz N. Introduction to Elliptic Curves and Modular Forms / Neal Koblitz. – New York : Springer, 1984. – 248 p. 10.1007/978-1-4684-0255-1
3. Knuth, D. The Art of Computer Programming. Volume 2 Seminumerical Algorithms, Third Edition / D. E. Knuth. – Massachusetts: Addison-Wesley, 1997. – 762 p.
4. Hankerson D. Guide to Elliptic Curve Cryptography / D. Hankerson, A. Menezes, S. Vanstone. – New York : Springer, 2004. – 311 p.
5. Crandall R. Prime Numbers. A Computational Perspective. Second Edition / Richard Crandall, Carl Pomerance. – New York : Springer, 2005. – 604 p. 10.1007/978-1-4684-9316-0
6. Rivain M. Fast and Regular Algorithms for Scalar Multiplication over Elliptic Curves / Matthieu Rivain. – IACR Cryptology ePrint Archive, 2011. – 338 p.
7. Болотов, А. А. Алгоритмические основы эллиптической криптографии / А. А. Болотов. – М. : Изд-во, 2004. – 499 с.
8. Elliptic Curve Point Multiplication [Electronic resource] // December 31, 2015: Proceedings. – Mode of access: https://en.wikipedia.org/wiki/Elliptic_curve_point_multiplication. – Last access: 2016. – Title from the screen.
9. Pathak H. Speeding Up Computation of Scalar Multiplication in Elliptic Curve Cryptosystem / H. Pathak, M. Sanghi // International Journal on Computer Science and Engineering. – 2010. – No. 04. – P. 1024–1028.

Стаття надійшла до редакції 15.02.2016.
Після доробки 01.03.2016.

Дичка И. А.[1], Онай Н. В.[2], Дрозда Т. П.[3]

[1]Профессор, доктор технических наук, декан факультета прикладной математики НТУУ «КПИ», Киев, Украина

[2]Старший преподаватель кафедры программного обеспечения компьютерных систем факультета прикладной математики НТУУ «КПИ», Киев, Украина

[3]Магистрант кафедры программного обеспечения компьютерных систем факультета прикладной математики НТУУ «КПИ», Киев, Украина

МОДИФИЦИРОВАННЫЙ ОКОННЫЙ МЕТОД ОДНОКРАТНОГО УМНОЖЕНИЯ ТОЧКИ ЭЛЛИПТИЧЕСКОЙ КРИВОЙ НА СКАЛЯР В ПОЛЕ GF(P)

При реализации многих криптографических приложений возникает потребность в быстрых алгоритмах умножения точки эллиптической кривой на число. В данной статье предложен модифицированный оконный метод однократного умножения точки эллиптической кривой на скаляр над полем GF(p). Объектом исследования являются процессы выполнения операций в эллиптических криптосистемах. Предметом исследования являются методы и алгоритмы выполнения операций однократного умножения точки эллиптической кривой на число над полем GF(p). Целью данного исследования является разработка и оптимизация методов и алгоритмов выполнения операции умножения точки эллиптической кривой на скаляр над полем GF(p) для улучшения временных характеристик. Существующие и предложенные алгоритмы реализованы на языке программирования C# в среде разработки Visual Studio 2013. В данной статье проведено исследование существующих алгоритмов скалярного умножения точки эллиптической кривой и разработаны три модификации LR-алгоритма оконного метода и обобщенная модификация. Экспериментальные исследования реализованных алгоритмов проводились согласно предложенной нами методики, которая позволяет нивелировать влияние на результаты исследования множителя и точки эллиптической кривой. Проведенное экспериментальное исследование оконных методов и их модификаций показало увеличение быстродействия работы модифицированных алгоритмов по сравнению с существующими в среднем на 13%.

Ключевые слова: ЭВМ, эллиптическая криптография, скалярное умножение, таблица предвычислений, эллиптическая кривая, конечное поле.

Dychka I. A., Onai M. V., Drozda T. P.

[1]Professor, Dc.Sc., Dean of the Faculty of Applied Mathematics NTUU «KPI», Kyiv, Ukraine

[2]Senior Lecturer, Department of Computer Systems Software of Faculty of Applied Mathematics, NTUU «KPI», Kyiv, Ukraine

[3]Master student, Department of Computer Systems Software of Faculty of Applied Mathematics, NTUU «KPI», Kyiv, Ukraine

MODIFIED METHOD FOR ELIPTIC CURVE SCALAR POINT MULTIPLICATION OVER GF(P)

During development of many cryptographic applications, we need to perform fast algorithms of scalar multiplication. In this paper we propose a modified window method of elliptic curve point multiplication over the GF(p). The object of the research are the processes of performing operations in elliptic cryptosystems. The subject of the research are the methods and the algorithms of elliptic curve point multiplication over the GF(p). The goal of the research is to develop and optimize the methods and the algorithms of performing elliptic curve point multiplication operation over the GF(p) for improving the time characteristics. Existing and proposed algorithms were implemented with C# programming language and integrated development environment – Visual Studio 2013. In this article we did an investigation of the existing algorithms of elliptic curve point multiplication and developed three versions of the window method LR-algorithm and generalized modification. Experimental studies of the implemented algorithms were performed according to the proposed methodology, which allows us to explore the impact of the multiplier and elliptical curve point on the results of the research. The experimental research of window methods and their modifications showed an increase speed of the modified algorithms compared to the existing algorithms in average of 13%.

Keywords: computers, elliptic curve cryptography, scalar multiplication, precomputation table, elliptic curve, finite field.

REFERENCES

1. Miller V. Use of elliptic curves in cryptography, *Lecture Notes in Computer Science. Advances in cryptology – CRYPTO 85*. Springer, 1986, pp. 417–426 10.1007/3-540-39799-X_31
2. Koblitz N. Introduction to Elliptic Curves and Modular Forms. New York, Springer, 1984, 248 p. 10.1007/978-1-4684-0255-1
3. Knuth D. The Art of Computer Programming. Volume 2 Seminumerical Algorithms, Third Edition. Massachusetts, Addison-Wesley, 1997, 762 p.
4. Hankerson D., Menezes A., Vanstone S. Guide to Elliptic Curve Cryptography. New York, Springer, 2004, 311 p.
5. Crandall R., Pomerance C. Prime Numbers. A Computational Perspective. Second Edition. New York : Springer, 2005, 604 p. 10.1007/978-1-4684-9316-0
6. Rivain M. Fast and Regular Algorithms for Scalar Multiplication over Elliptic Curves. IACR Cryptology ePrint Archive, 2011, 338 p.
7. Bolotov A. A. Algoritmicheskie osnovy e'llipticheskoj kriptografii. Moscow, Izd-vo, 2004, 499 p
8. Elliptic Curve Point Multiplication [Electronic resource]. December 31, 2015: Proceedings. Mode of access: https://en.wikipedia.org/wiki/Elliptic_curve_point_multiplication, Last access: 2016, Title from the screen.
9. Pathak H., Sanghi M. Speeding Up Computation of Scalar Multiplication in Elliptic Curve Cryptosystem, *International Journal on Computer Science and Engineering*, 2010, No. 04, pp. 1024–1028.

УДК 681.3.06.003.1:004.369.6

Нич Л. Я.[1], Шаховська Н. Б.[2], Камінський Р. М.[3]

[1]Асистент, кафедра інформаційних систем та мереж, Національний університет «Львівська політехніка», Львів, Україна
[2]Д-р техн. наук, професор, кафедра інформаційних систем та мереж, Національний університет «Львівська політехніка», Львів, Україна
[3]Д-р техн. наук, доцент, кафедра інформаційних систем та мереж, Національний університет «Львівська політехніка», Львів, Україна

ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ ІНФОРМАЦІЙНОГО ПОШУКУ В СИСТЕМАХ КОНСОЛІДОВАНОЇ ІНФОРМАЦІЇ

Для оцінювання ефективності інформаційного пошуку запропоновано поділяти знайдені документи на пертинентні, релевантні та нерелевантні. Ефективність пошуку визначати відношенням кількості пертинентних та релевантних документів до кількості нерелевантних документів, а властивості пошукової системи повинні бути подані відповідним коефіцієнтом. Основною метою цього дослідження є розробка інтегрального критерію оцінювання ефективності інформаційного пошуку за результатами видачі в системах консолідованої інформації. Введено поняття консолідованої інформації.

Розроблено метод оцінювання ефективності інформаційного пошуку. Він демонструє використання поділу знайдених і виданих документів на пертинентні, релевантні та нерелевантні. Введено інтегральний показник релевантності документу пошуковому запиту, який враховує негативну та позитивну оцінку. Оцінку ефективності подано як сумарне значення різних компонентів. Експеримент виконано на підставі проведення інформаційного пошуку в одному або в кількох інформаційних фондах і на різних пошукових системах за одного набору ключових слів.

Розроблений підхід до побудови оцінки інформаційного пошуку має практичне значення, оскільки отримані кількісні значення локальних оцінок дають підстави для оптимізації набору ключових слів, та визначення найбільш відповідних інформаційних фондів і пошукових систем.

Ключові слова: інформаційна система, інформаційний пошук, ефективність, пертинентність, релевантність.

НОМЕНКЛАТУРА

P – клас релевантних документів;

$П$ – клас пертинентних документів;

H – клас нерелевантних документів;

D – множина інформаційних джерел;

K – користувач, який формує запит для інформаційного пошуку в джерелах;

N – кількість документів у видачі;

$f_{П}$ – відносна частоту появи пертинентних документів;

$f_{Р}$ – відносна частоту появи релевантних документів;

$f_{Н}$ – відносна частоту появи нерелевантних документів;

$f_{Р+П}$ – відносна частоту появи релевантних і пертинентних документів;

γ – відношення загальної кількості релевантних і пертинентних документів до кількості нерелевантних документів.

ВСТУП

Розвиток комп'ютерної техніки та інформаційних технологій значною мірою стимулював створювання і наповнення різноманітною інформацією як загальні, так і спеціалізовані бази даних, забезпечуючи управління ними. Проте, з іншої сторони, величезні обсяги даних практично унеможливлюють безпосередню роботу користувача з ними, що у свою чергу стимулювало розвиток відповідних пошукових систем, основною метою яких є своєчасне і повне забезпечення користувача необхідними йому даними. Тому найкритичнішою проблемою, з якою зустрічаються користувачі, – це забезпечення надійного, постійного та повнофункціонального доступу до актуальних даних.

1 ПОСТАНОВКА ЗАДАЧІ

Проблема побудови критеріїв оцінювання функціональної ефективності інформаційного пошуку в системах консолідованої інформації полягає в тому, що шукана інформація зберігається в різних джерелах, створених в різний час і з різною метою; вона є складно структурованою, для різних задач має різну інформаційну цінність, і різними користувачами сприймається по-різному. Натомість, за високої надійності і стабільності апаратного та програмного забезпечення вся відповідальність за результати пошуку покладена на людський фактор в сенсі укладання пошукового запиту. В цьому плані, об'єктивна оцінка ефективності пошуку, а в даному випадку ще й консолідації знайдених і виданих документів може бути зроблена саме на підставі виданих документів.

Історично, а в певному сенсі і політично (з метою захисту інформації) різні джерела інформації (електронні бібліотеки, загальні та локальні бази, сховища, простори даних) мають свої особливості стосовно організації форм збереження, пошуку, виявлення, видачі потрібної інформації, які в основному полягають у видах і тонкощах мов запитів та кодування збережуваної інформації. Очевидно, що вихід з такої ситуації для користувача є і йому немає потреби вивчати премудрості мов різних запитів, оскільки потрібний пошук здійснюють спеціальні пошукові системи. Робота з однією чи навіть декількома базами даних практично полягає у правильному формуванні запиту і тут існуюча пошукова система допомагає знайти необхідну інформацію. Наприклад, локальні бази даних навіть великих підприємств досить швидко дають інформацію про виготовлені вироби, товари, зарплату працівників тощо. Проте, пошук даних в «чужих» базах

DOI 10.15588/1607-3274-2016-2-13

даних може стати складною проблемою [11]. Тут найкращим прикладом є пошукова система Google та аналогічні з нею, які видають десятки тисяч документів, з яких вибирають лише декілька, витрачаючи величезну кількість часу на пошук потрібних серед наданих пошуковою системою.

Не меншою є проблема інформаційного пошуку в системах консолідованої інформації. Термін консолідована інформація означає одержані з декількох інформативних джерел системно інтегровані різнотипні інформаційні ресурси в сукупності наділені ознаками повноти, цілісності та несуперечності. Вони фактично подаються у формі адекватної інформаційної моделі проблемної області для її аналізу, опрацювання та ефективного використання в процесах підтримки прийняття рішень. Як правило, такі системи є результатом інтеграції різноманітних джерел інформації, які були створені в різний час і за різними принципами та мовами запитів, а головне за різними фаховими ознаками та онтологіями. Досить часто основні техніко-економічні дані зосереджені в системах, які реалізують численні офісні, адміністративні і технологічні процеси, а в результаті такі дані не можуть спільно використовуватись в масштабах всього підприємства.

2 ЛІТЕРАТУРНИЙ ОГЛЯД

Поняття ефективності має широке тлумачення і переважно в економічному аспекті. В роботі [1] для оцінки якості роботи пошукової системи використовуються такі оцінки: точність, повнота, акуратність, помилка, *F*-міра, які визначаються як метрики на множинах документів і фактично дають кількісну характеристику самого пошуку. З результатів аналізу існуючих пошукових систем в [2] робиться висновок, що для пошуку документів гіпертекстових базах даних існуючі загальновизнані оцінки мають певні обмеження. Запропоновано використовувати додаткові характеристики, до яких відносять M-різновид вибірки та U-впорядкованість вибірки. На цій підставі наводиться коефіцієнт впорядкованості та коефіцієнт пошукового шуму. Виділена низка факторів, що впливають на успішність пошуку. Оцінці ефективності інформаційних систем, як одній з проблем інформаційного суспільства присвячена стаття [3], в якій на основі аналізу практичного застосування інформаційних систем показано, що в оцінці ефективності інформаційних систем можна виділити три типи ефектів: врахування додаткової інформації, нормування та врахування організаційних процесів та планування, оптимізації, управління процесами та ресурсами. Підкреслено роль врахування витрат, які ділять на дві складові: капітальні (бюджетні) або прямі витрати і позабюджетні, пов'язані з користувачами. Для оцінки трудовитрат приведена модифікована формула, яка враховує модель оцінки вартості розробки програмного забезпечення. Кількісні показники – оцінки функціональної ефективності інформаційно-пошукових систем приведені в [4]. До них віднесено такі: повноту, точність, акуратність, помилки. Для оцінювання функціональної ефективності інформаційно-пошукових систем запропоновано використовувати методи теорії статистичних рішень. Значна увага приділена модифікації відомого критерію зваженої комбінації, та показано його ефективність на прикладі експериментального пошуку в масиві патентів США. У роботі [5] розглянута проблема пошуку інформації в Інтернет, її зв'язок з традиційною проблемою пошуку інформації. Описано нові завдання, відрізняють проблему пошуку в Інтернет від традиційної проблеми пошуку інформації, даний огляд існуючих методів пошуку інформації в Інтернет. Модель розв'язку задачі інформаційного пошуку, яка включає математичний опис послідовного та бінарного пошуків приведена в [6]. Зміст послідовного пошуку полягає в проведенні порівнянь записів. Для бінарного пошуку використовується бінарне дерево. Показано, що ефективність пошуку визначається принаймні двома основними – точністю і повнотою, та чотирма додатковими – специфічністю, вибірковістю, коефіцієнтом втрати інформації та коефіцієнтом пошукового шуму – показниками. Зазначено, що для оцінки роботи пошукової системи потрібна репрезентативна кількість запитів. У [7] формулюються принципи оцінки ефективності функціонування сучасних інформаційно-пошукових систем Інтернету. Наводяться результати тестування шести інформаційно-пошукових систем на основі методу визначення глибини користувацького пошуку.

На підставі аналізу існуючих підходів до оцінювання ефективності інформаційного пошуку можна зробити такі висновки.

1. В теоретичному плані оцінювання ефективності проводиться на підставі математичних моделей інформаційного пошуку. Для цього використовують переважно теоретико-множинний апарат, рідше ймовірнісний, і розглядають відношення множин релевантних та нерелевантних документів у видачі та інколи у інформаційному фонді.

2. В практичному використанні використовують критерії точності і повноти, рідше включають і частку нерелевантних документів у видачі.

3. Відсутність інтегрального критерію ефективності інформаційного пошуку.

Основною метою дослідження є розробка інтегрального критерію оцінювання ефективності інформаційного пошуку за результатами видачі в системах консолідованої інформації.

Такий інтегральний показник повинен враховувати не лише позитивний результат пошуку, але і негативний – частку нерелевантних документів та частку релевантних, але не виданих документів. Релевантні невидані документи за одним запитом можуть бути знайдені і включені у видачу або за рахунок іншого (нового) запиту або за рахунок модифікації даного запиту. Проте в першу чергу базове оцінювання ефективності пошуку має здійснюватись виключно на підставі видачі першого запиту, а вже далі такий інтегральний показник можна уточнювати додатковими оцінками.

3 МАТЕРІАЛИ І МЕТОДИ

Пошук в системі консолідованої інформації. Розглянемо роботу системи консолідованої інформації як діяльність користувача, пов'язану з відбором відповідної інформації стосовно поставленої задачі. В результаті зроб-

леного запиту інформаційно-пошукова система здійснює видачу знайдених документів. Як правило, не всі видані документи відповідають зробленому запиту і потребам користувача. З точки зору його задачі видані документи можуть бути поділені принаймні на три класи: релевантні P, пертинентні Π та не релевантні H.

Позначимо, через D множину різноманітних інформаційних джерел $D = \{d_1, d_2, \dots, d_n,\ d_i \in D, d_i \cap d_j \neq \neq \varnothing,\ i, j = 1, 2, \dots, n\}$, які можуть мати спільні фонди інформаційного ресурсу; K – користувач, який формує запит для інформаційного пошуку в різнотипних джерелах.

Тоді, процес інформаційного пошуку в системах консолідованої інформації можна подати у вигляді схеми цілеорієнтованої роботи трьох блоків – «Опрацювання запитів», «Консолідації даних» та «Опрацювання даних» зображеної на рис. 1. Перший з цих блоків функціонально забезпечує переклад мови запиту користувача на мову запитів кожного з інформаційних джерел $d_i \in D$. В результаті, кожне таке джерело розуміє отриманий запит і процес пошуку може здійснюватися переважно його власною пошуковою системою.

Функціонально другий блок здійснює консолідацію знайдених даних, тобто приводить дані різних форматів у формат користувача, тобто вирішує обернену задачу – приведення різнотипних даних до типу запиту, сформованого користувачем. Консолідовані дані передаються в блок «Опрацювання даних», робота якого полягає у ранжуванні даних за частотою використання, часовими характеристиками, важливістю, доступністю, терміном використання тощо. Іншими словами, знайдена в результаті пошуку інформація має бути подана користувачеві у тій самій формі, у якій він сформував свій запит або в іншій, зрозумілій для нього формі. У такій ситуації користувач може отримати надзвичайно велику кількість, випадково перемішаних, як релевантних так і не релевантних документів. Для зменшення кількості нерелевантних документів у цьому блоці здійснюється відповідний логічний аналіз наявності збігів виданих документів з визначеними в запиті. Тут, фактично здійснюється фільтрація виданих документів, шляхом використання відповідних критеріїв, попередньо заданих користувачем. У цьому плані, інформаційний пошук в системах консолідованої інформації суттєво відрізняється від пошуку в звичайних базах чи сховищах даних. Тому, оцінювання ефективності інформаційного пошуку в системах консолідованої інформації має враховувати і особливості нормалізації різнотипних даних, тобто приведення їх до форми запиту користувача.

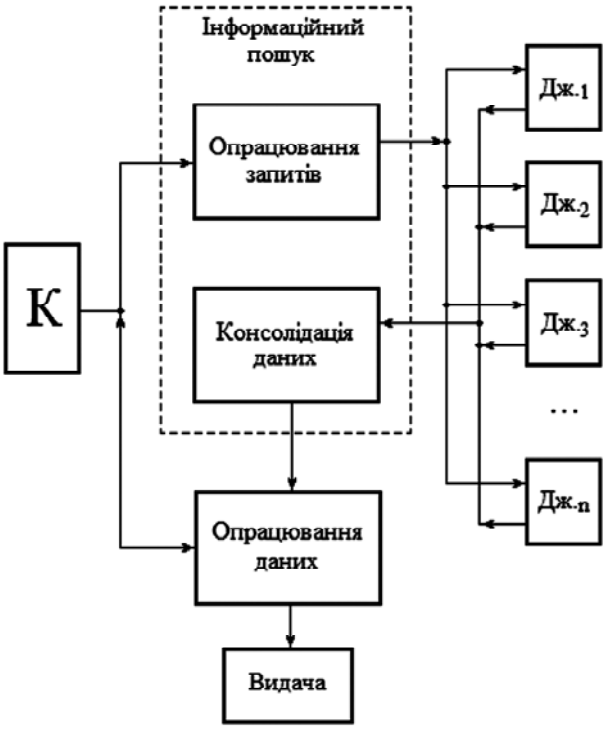


Рисунок 1 – Інформаційний пошук в різнотипних системах

Відповідність видачі запиту. Найскладнішим моментом в оцінюванні ефективності будь-якого інформаційного пошуку є встановлення відповідності між знайденими і виданими документами і документами, а точніше пошуковими ознаками документів, поданих у запиті. Справа в тому, що ступінь відповідності, тобто чи є релевантними видані документи чи ні, є вельми суб'єктивним. Крім того, якщо можна точно відповісти даний документ є релевантний або нерелевантний, то чітко вказати, чи даний документ є пертинентним чи ні, оскільки він може бути пертинентним різною мірою.

Зі змісту понять релевантності та пертинентності випливає, що оцінювання ефективності пошуку має принципові дві складові. Нагадаємо, що поняття релевантності означає відповідність інформаційного пошуку, зробленому користувачем запиту, а пертинентність – відповідність інформаційній потребі користувача.

Перша з них це оцінювання, а точніше розуміння пошуковою системою складеного користувачем запиту. В цьому плані інформаційно-пошукова система відбирає ті документи, ознаки яких вказані у запиті. Очевидно, що в такому разі семантичний аналіз виявлених документів в базі або сховищі даних, у файлах чи бібліотеках не проводиться, а лише здійснюється зіставлення ознак виявлених документів і, за умови повного чи часткового збігу, документи подаються у видачу.

Друга складова це оцінювання документів у видачі, отриманих користувачем, у результаті інформаційного пошуку. Тут користувач розділяє документи на три групи: релевантні (Р). пертинентні (П) та нерелевантні (Н).

Документи у видачі як правило сортуються інформаційною-пошуковою системою за певними критеріями: за датою (власна дата документа або остання дата звертання до нього), за рейтингом користування (скільки разів даний документ фігурував у запитах різних користувачів загалом чи за певний період). Можливі і інші критерії, наприклад за обсягом. Отримавши видачу, тобто перелік знайдених документів користувач послідовно або вибірково ознайомлюється з документами відбираючи релевантні та пертинентні і відкидаючи нерелевантні. Послідовність релевантних, пертинентних та нерелевантних документів у кожній конкретній видачі практично завжди є випадковою. Перевірка цього факту здійснена експериментально в такий спосіб.

Організація експериментів. В процесі пошуку необхідної інформації для проведення наукових досліджень поряд з відбором пертинентних документів також фіксувалися релевантні та нерелевантні. Зміст експерименту поданий планом дослідження.

4 ЕКСПЕРИМЕНТИ

Відбір ключових слів. Для цього були сформульовані такі ключові слова, а точніше словосполучення: інформаційний пошук; моделі інформаційного пошуку; інформаційно-пошукова система; ефективність інформаційно-пошукових систем; оцінювання ефективності інформаційного пошуку.

Уточнення понять. Пертинентність – документи, які за змістом максимально відповідають потребі користувача і мають усі реквізити для посилання на них у магістерській кваліфікаційній роботі (тобто документи, що є електронними копіями паперових монографій, статей у наукових журналах та збірниках праць, тезисах та працях наукових форумів та статті подані в енциклопедіях та довідниках).

Релевантність – документи, які за змістом цілком відповідають потребі користувача, але не мають реквізитів своїх паперових оригіналів і для посилання на них треба використовувати їхню електронну адресу, яка в деяких випадках є або громіздкою або неточною і для виявлення цього документу необхідно провести додатково ще й окремий спеціальний пошук, причому результат не гарантується.

Усі інші документи визнаються як нерелевантні.

Хід експерименту. Для експериментальних досліджень використано інформаційно-пошукові системи Google, Яndex, Мета, Rambler таYahoo, які за ключовими словами видають веб-сторінки знайдених документів. Налаштування пошуку забезпечило оптимальний варіант видачі результату – 10 електронних документів на кожній сторінці. На основі попередніх результатів пошуку і з власного досвіду, відомо, що потрібна інформація стосовно даного питання буде знаходитись на перших п'яти сторінках. Тому для експериментів вибрано обмеження 5 повних сторінок, тобто обсяг видачі становив 50 документів.

Для кожної сторінки, за результатами перегляду, кожному з десяти виданих документів присвоювалися індекси P, $П$ та H.

У табл. 1 приведені результати одноразової видачі знайдених документів для вказаних пошукових систем за ключовим словом «Оцінювання ефективності інформаційного пошуку».

Попередні результати. Задача користувача полягає в тому, щоб серед цієї множини вибрати саме ті, які йому потрібні. Очевидно за будь-якого пошуку перегляд отриманих документів буде аналогічним. Оскільки надана вибірка є скінчена, можемо оцінити ефективність пошукової системи відношенням сприятливих подій до всіх можливих, тобто відношенням, наприклад, кількості релевантних документів до кількості всіх наданих документів, отриманих за даним запитом. Якщо документи класифікувати як в даному прикладі, то можна отримати три частоти появи документів кожного класу:

$$f_P = \frac{\sum_{i=1}^{n} P_i}{N},\ f_H = \frac{\sum_{j=1}^{m} H_j}{N},\ f_П = \frac{\sum_{k=1}^{l} П_k}{N},$$

де N – кількість документів у видачі.

Таблиця 1 – Оцінювання ефективності інформаційного пошуку

Ключове слово «Оцінювання ефективності інформаційного пошуку»					
	Google	**Яndex**	**Мета**	**Rambler**	**Yahoo**
1	Р Р Р П Н	Р Р Н Р Н	Н Н Н Н Н	Н Р Н Р Н	Н Р Р Н Н
2	Р Р П Н Н	Р Р Н Н Н	Н Р Н Р Н	Н Р Р Н Р	Н Н Р Н Н
3	Р Н П Р Н	Р П Н Р Н	Н Р Р Р Н	Р Р Р Р Н	Р Р Р Р Н
4	Н Н Н Р Н	Н Р Н Р Н	Р Н Н Н Н	Р Р П Р Н	Р Р П Р Н
5	Н Н Р Н Н	Р Н Р Н Н	Р Н Н Н Н	Р Н Н Н Р	Р Н Н Н Н
6	Н Н Н Р Н	Р Р Р Н Н	Р Н Р Н Н	Р Н Н Н Н	Р Н Р Н Н
7	Н Н Н Н Н	Р Р Н Н Н	Н Н Р Н Н	Н Н Н Н Н	Н Н Н Р Н
8	Н Р Н Н Н	Р Н Н Н Н	Н Н Н Н Н	Н Н Н Н Н	Н Н Н Н Н
9	Н П Н Н Н	Р Н Н Н Н	Н Н Н Н Н	Н Н Р Р Н	Н Н Н Н Н
10	Н Р Н Р Н	Р Р Р Н Н	Р Н Н Н Н	Р Р Н Н Н	Р Н Н Н Н

Таблиця 2 – Зведена таблиця експериментів

Ключові слова	Google	Яndex	Мета	Rambler	Yahoo
Оцінка ефективності інформаційного пошуку	П – 4 Р – 13 Н – 33	П –1 Р –21 Н – 28	П –0 Р –11 Н – 39	П –1 Р –19 Н – 30	П –1 Р –15 Н – 34
Інформаційний пошук	П – 7 Р –11 Н – 32	П –2 Р –12 Н – 36	П –2 Р –14 Н – 34	П –3 Р –15 Н – 32	П –4 Р –10 Н – 36
Модель інформаційного пошуку	П –3 Р –11 Н – 36	П –2 Р –18 Н – 30	П –1 Р –9 Н – 40	П –9 Р –20 Н – 21	П –1 Р –21 Н – 28
Інформаційно-пошукова система	П – 1 Р –21 Н – 28	П –0 Р –22 Н – 28	П –0 Р –5 Н – 45	П –1 Р –18 Н – 31	П –1 Р –11 Н – 38
Ефективність інформаційно-пошукових систем	П –1 Р –16 Н – 33	П –2 Р –22 Н – 26	П –0 Р –17 Н – 33	П –1 Р –16 Н – 33	П –3 Р –21 Н – 26

На практиці як правило інформаційний пошук здійснюється за різними запитами, в залежності від поставлених задач.

У свою чергу, задачі можуть стосуватися різних предметних областей, обсягу їх онтологій, специфіки конкретних об'єктів, що потребують їхнього розв'язку. З другої сторони, не можна бути впевненому в тому, що інформаційні джерела мають усю необхідну інформацію з будь-якої області знань та діяльності людини. А тому кількості наданих користувачам документів є різними. Зазвичай пошук в джерелах інформації здійснюється пошуковою системою, яка працює за певним алгоритмом і визначеними формальними критеріями відповідності, а тому, можна припустити, що результати різних пошуків в одному і тому ж джерелі інформації будуть статистично однорідні, тобто матимуть певні статистичні закономірності, які можуть відбитися, принаймні, на співвідношенні частоти появи розглянутих вище класів.

5 РЕЗУЛЬТАТИ

Послідовність документів у видачі можна зобразити графічно, у вигляді діаграми приведеної на рис. 2.

В якості кількісної оцінки використано відносну частоту появи того чи іншого виду документів. Для поданого результату пошуку маємо такі співвідношення:

пертинентних $f_{\text{П}} = \dfrac{П}{50}$, релевантних $f_{\text{Р}} = \dfrac{P}{50}$, нерелевантних $f_{\text{Н}} = \dfrac{H}{50}$, релевантних і пертинентних (корисних) $f_{\text{Р+П}} = \dfrac{P+П}{50}$, а також відношення загальної кількості релевантних і пертинентних документів до кількості нерелевантних документів

$$\gamma = \frac{P+П}{H}.$$

Очевидно, що усі ці значення значною мірою залежать від обсягу документів в інформаційній системі (базі, сховищі даних, папках з файлами, бібліотеці), можливостей інформаційно-пошукової системи, форми запиту, а також від інформаційної потреби користувача – наскільки глибоко він розуміє завдання, для вирішення якого він здійснює даний пошук.

Оцінювання ефективності інформаційного пошуку. Сформовані, практично на відповідних пошукових мовах, властивих тому чи іншому інформаційному фонду, запити мають досить обмежену кількість пошукових ознак – ключових слів, певного типу розширень та пояснень чи обмежень. Алгоритми інформаційно-пошукових систем використовуючи ці дані в процесі сканування-пошуку існуючого каталогу переважно використовують в якості даних автора, назву та анотацію документів, хоча можливим є і сканування самого документа. Оскільки ключові слова в залежності від контексту можуть мати декілька значень у видачу потрапляють абсолютно нерелевантні документи.

У загальному оцінювання ефективності базується на визначенні, як було сказано вище, на оцінках точності і повноти. Спроба використати додаткові показники пошуку вимагає врахування не лише обсягу самого інформаційного фонду, але і обсягу релевантних та нерелевантних стосовно даного запиту документів. Отримати такі дані практично неможливо, оскільки для одної задачі документи можуть бути релевантними, а для другої вже ні. З другої сторони, якщо знати всі релевантні документи у фонді то можна здійснити пошук лише для них і тоді у видачі будуть лише релевантні документи, а це здійснити практично не можливо, принаймні з двох причин: ніхто не буде з багатотисячного інформаційного фонду відбирати релевантні для даної задачі окремого користувача документи, присутність конфіденціальної інформації та відсутність інформації про сам фонд, за винятком лише загальних його характеристик. Тому найбільш правомірним є оцінювання ефективності пошуку за його результатами, тобто на основі документів, які є у видачі.

За наявності трьох типів документів оцінити ефективність інформаційного пошуку можна в такий спосіб логічного виведення. Очевидно, що пертинентні документи мають найбільшу цінність для користувача.

Очевидним є те, що для своєї задачі користувач використовує лише релевантні та пертинентні документи, тому ефективність пошуку у загальному випадку є пропорційна кількості релевантних і пертинентних документів, що є сприятливою подією для користувача, тобто

$$E_{\text{пош}} = \frac{P+П}{П+P+H}. \qquad (1)$$

Наявність у видачі не релевантних документів є обернено-пропорційною подією до кількості пертинентних і релевантних документів у видачі, а тому, ефективність відносно не релевантних документів можна подати як

$$E_{\text{пош}} = \frac{P+П}{H}. \qquad (2)$$

Враховуючи особливості форми запиту, яка тісно пов'язана з даною конкретною інформаційною системою, тобто з її інформаційним фондом та його системою індексування необхідно ввести деякий коректую-

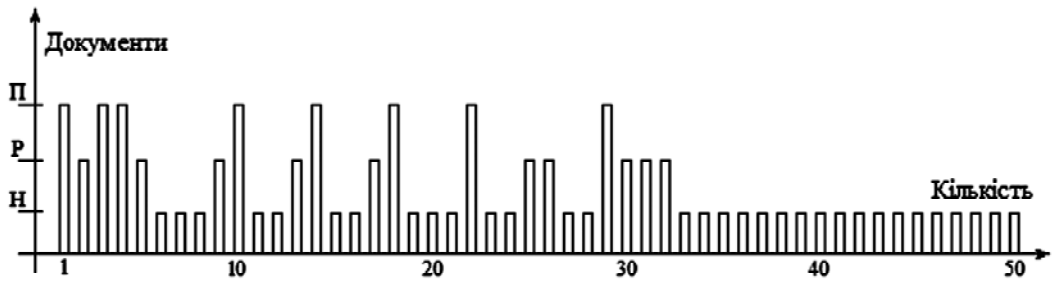


Рисунок 2 – Розподіл документів у видачі: П – пертинентні, Р – релевантні, Н – нерелевантні

чий множник – коефіцієнт пропорціональності β, в результаті чого отримаємо

$$E_{\text{пош}} = \beta \cdot \frac{П + Р}{П + Р + Н}. \quad (3)$$

Оцінка ефективності пошуку у вигляді (3) дає характеристику здійсненого інформаційного пошуку для даного конкретного ключового слова та конкретної пошукової системи за результатами отриманої видачі, обмеженої, наприклад, 50 документами.

За кількістю релевантних документів стосовно вибраних ключових слів та інформаційно-пошукових систем результати пошуку наведені в табл. 3.

У поданій оцінці ефективності, залишається невідомим показник β. Оцінити цей показник можна лише на підставі даних про пошуковий алгоритм і саму пошукову систему, яка використовується даним інформаційним фондом.

Якщо припустити, що пошуковий алгоритм інформаційної системи переглядає всі документи інформаційного фонду або, принаймні, усі документи конкретної рубрики згідно з інформацією поданою у запиті, тоді можна прийняти значення показника $\beta = 1$. Проте докладну інформацію про характеристики цього алгоритму отримати практично неможливо.

Очевидним є той факт, що чим більший обсяг інформаційного фонду, тим більше релевантних документів буде знайдено. Однак, тут треба мати на увазі і популярність чи розвиненість даної тематики, оскільки саме вона, тобто її популярність і затребуваність визначають обсяг документів у фонді [9–10].

6 ОБГОВОРЕННЯ

Вирази (2) і (3) дають об'єктивну оцінку ефективності інформаційного пошуку, але в перших (лівих) варіантах, лише за умови, що у видачі будуть присутніми і не релевантні документи – принаймні, хоча б один. Невиконання цієї умови означає ділення на нуль. Тобто, буде невірний результат оцінювання. Така ситуація може виникнути тоді, коли кількість релевантних документів в інформаційному фонді перевищує обсяг видачі.

У цьому випадку оцінюється ефективність за видачею для одного чи декількох запитів. Якщо такий показник використати для кожного з декількох запитів, але таких, що стосуються конкретної теми можна оцінити якість і самого запиту, точніше встановити, який з запитів чи які ключові слова є найбільш ефективними і вже за ними модифікувати наступні запити.

ВИСНОВКИ

Ефективність інформаційного пошуку в системах консолідованої інформації в сенсі побудови інтегрального показника практично не може бути визначена, оскільки крім двох показників – повноти і точності, усі інші вимагають знання кількості релевантних та нерелевантних документів у даному інформаційному фонді стосовно даної задачі. Отримати такі дані для великих за обсягами фондів є неможливо, оскільки: по-перше здійснити такий підрахунок означає перегляд кожного документа, по-друге, у великих базах даних перехід від нерелевантних документів до релевантних практично за будь-яким запитом є нечітким і розмитим, по-третє – для різних задач поняття релевантності документів різняться.

Найпростішим способом побудови оцінки ефективності пошуку є використання логічного підходу, який подається відношенням – кількості отриманих потрібних і замовлених документів до кількості документів у даній видачі. В цьому плані, на ефективність пошуку впливає не лише наявність в інформаційному фонді потрібних документів, але й правильність побудови самого запиту згідно з вимогами даної пошукової системи.

Наведений приклад оцінювання ефективності інформаційного пошуку демонструє використання поділу знайдених і виданих документів на пертинентні, релевантні та нерелевантні. В результаті якого, оцінку ефективності можна подати як усереднену, або сумарну, на підставі проведення інформаційного пошуку в одному або в кількох інформаційних фондах і на різних пошукових системах за одного набору ключових слів.

Розроблений підхід до побудови оцінки інформаційного пошуку має практичне значення, оскільки отримані кількісні значення локальних оцінок дають підстави для оптимізації набору ключових слів, та визначення найбільш відповідних інформаційних фондів і пошукових систем.

ПОДЯКИ

Роботу виконано в рамках держбюджетної науково-дослідної теми «Методи та засоби консолідації баз даних в інформаційних системах електронного урядування», тематика кафедри інформаційних систем та мереж Національного університету «Львівська політехніка», 2010/2012, № держреєстрації 0110U005022.

СПИСОК ЛІТЕРАТУРИ

1. Агеев М. Официальные метрики РОМИП 2010 / М. Агеев, И. Кураленок, И. Некрестьянов // Российский семинар по Оценке Методов Информационного Поиска. Труды РОМИП 2010, Казань, 15 октября 2010 г. – Казань, 2010. – С. 172–187.
2. Целых А.Н. Оценка эффективности информационного поиска / А. Н. Целых, Э. М Котов // Известия ТРТУ. Тематический выпуск «Управление в математических системах». – Таганрог : Изд-во ТРТУ. – 2006. – № 10 (65). – С. 43–45.
3. Яхина Е.П. Методы оценки информационных систем / Е. П. Яхина // В мире научных открытий. – 2010. – № 3 (09). – Часть 1. – С. 63–66.
4. Попов С. В. Оценка функциональной эффективности систем текстового поиска на примере поиска патентных документов / С. В. Попов // Патентная информация сегодня. – 2010. – № 1. – С. 22–25.
5. Козлов Д. Д. Информационно-поисковые системы в Internet: текущее состояние и пути развития / Д. Д. Козлов // Техноло-

Таблиця 3 – Оцінка ефективності інформаційно-пошукових систем за кількістю релевантних документів

Ключові слова	Google	Яndex	Мета	Rambler	Yahoo
Оцінка ефективності інформаційного пошуку	0,34	0,44	0,22	0,40	0,32
Інформаційний пошук	0,36	0,28	0,32	0,36	0,28
Модель інформаційного пошуку	0,28	0,40	0,20	0,58	0,44
Інформаційно-пошукова система	0,44	0,44	0,10	0,38	0,24
Ефективність інформаційно-пошукових систем	0,34	0,48	0,34	0,34	0,48
Усереднений показник ефективності	0,35	0,40	0,23	0,41	0,35

гический обзор [Электронный ресурс]. – Режим доступа: lvk.cs.msu.su/~ddk/ir_and_ia_review.pdf
6. Тявкин И. В. Математическая модель информационного поиска и оценка єффективности поисковой системы / И. В. Тявкин, В. М. Тютюнник // Вестник ТГТУ. – 2008. – Том 14. – № 3. – С. 478–481.
7. Козлов М. В. Метод оценки эффективности функционирования современных информационно-поисковых систем Интернета / М. В. Козлов, В. А. Яцко [Электронный ресурс]. – Режим доступа: http://www.dialog-21.ru/dialog2006/materials/html/Kozlov.htm.
8. Лекции по введению в информатику и информационные системы. – Лекция 13. Эффективность информационных систем. [Электронный ресурс]. – Режим доступа: http://informling.narod.ru/lectures.html
9. Kirchgassner G. Introduction to Modern Time Series Analysis / G.·Kirchgassner, J. Wolters. – Berlin : Springer, 2007. – 274 p.
10. Hegger R. Practical implementation of nonlinear time series methods: The TISEAN package / R. Hegger, H. Kantz, T. Schreiber // CHAOS 9. – 1999. –P. 413–435.
11. Kuhlthau, C. C. Seeking Meaning: A Process Approach to Library and Information Services. / C. C Kuhlthauю –. Westport, CT: Libraries Unlimited. – 2004. – 342 p.

Стаття надійшла до редакції 10.02.2016.
Після доробки 15.02.2016.

Ныч Л. Я.[1], Шаховска Н. Б.[2], Каминский Р. М.[3]
[1]Ассистент, кафедра информационных систем и сетей, Национальный университет «Львовская политехника», Львов, Украина
[2]Д-р техн. наук, профессор, кафедра информационных систем и сетей, Национальный университет «Львовская политехника», Львов, Украина
[3]Д-р техн. наук, доцент, кафедра информационных систем и сетей, Национальный университет «Львовская политехника», Львов, Украина

ОЦЕНКА ЭФФЕКТИВНОСТИ ИНФОРМАЦИОННОГО ПОИСКА В СИСТЕМАХ КОНСОЛИДИРОВАННОЙ ИНФОРМАЦИИ

Для оценки эффективности информационного поиска предложено разделять найденные документы на пертинентни, релевантные и нерелевантные. Эффективность поиска определяется отношением количества пертинентних и релевантных документов в количестве нежелательных документов, а свойством поисковой системы должно быть возможность учесть соответствующие коэффициенты. Основной целью данного исследования является разработка интегрального критерия оценки эффективности информационного поиска по результатам выдачи в системах консолидированной информации. Введено понятие консолидированной информации.

Разработан метод оценки эффективности информационного поиска. Он демонстрирует использование разделения найденных и выданных документов на пертинентные, релевантные и нерелевантные. Введено интегральный показатель релевантности документа поисковому запросу, который учитывает негативную и положительную оценку. Оценку эффективности определено как суммарное значение различных компонентов. Експеримент выполнен на основании проведения информационного поиска в одном или в нескольких информационных фондах и на разных поисковых системах с одним набором ключевых слов.

Разработанный подход к построению оценки информационного поиска имеет практическое значение, поскольку полученные количественные значения локальных оценок дают основания для оптимизации набора ключевых слов, и определение наиболее подходящих информационных фондов и поисковых систем.

Ключевые слова: информационная система, информационный поиск, эффективность, пертинентность, релевантность.

Nych L. Ya.[1], Kaminskyj R. M.[2], Shakhovska N. B.[3]
[1]Assistant professor, department of information systems and networks, Lviv Polytechnic National University, Lviv, Ukraine
[2]Dr.Sc., Professor, department of information systems and networks, Lviv Polytechnic National University, Lviv, Ukraine
[3]Dr.Sc., Professor, department of information systems and networks, Lviv Polytechnic National University, Lviv, Ukraine

EFFECTIVENESS EVALUATION OF SEARCH IN INFORMATION SYSTEMS WITH CONSOLIDATED INFORMATION

To evaluate the effectiveness of information retrieval there is proposed to share the found documents on pertinent, relevant and irrelevant. Search Performance is ratio to determine the number of pertinent and relevant documents to the number of irrelevant documents and search engine properties have been submitted by the coefficient. The goal of this paper is to develop integrated criterion of evaluating the effectiveness of information retrieval on the results of the issuance of consolidated information systems. The concept of consolidated informationis given.

The method of evaluating the effectiveness of information retrieval is built. It demonstrates the usage of the division found and published documents on pertinent, relevant and irrelevant. There is given integral indicator of the relevance of the document search query that takes into account the negative and positive features. Evaluation of effectiveness presented as the total value of the different components. The experiment was performed on the basis of information search in one or several search machines and information on the various search engines for one set of keywords.

The approach to building assessment information retrieval is of practical importance because quantitative values obtained local assessments give grounds to optimize the set of keywords and determine the most appropriate information collection and search engines.

Keywords: information system, information search, efficiency, pertinence, relevance, irrelevance.

REFERENCES

1. Aheev M., Kuralenok Y., Nekrestianov Y.Ofytsyalnye metryky ROMYP 2010, *Rossyiskyi semynar po Otsenke Metodov Informatsyonnoho Poiska. Trudy ROMYP 2010. (Kazan, 15 october 2010.)* Kazan, 2010, pp. 172–187.
2. Tselykh A. N., Kotov E. M. Otsenka эffektyvnosty ynformatsyonnoho poyska, *Yzvestyia TRTU. Tematycheskyi vypusk «Upravlenye v matematycheskykh systemakh»*. Tahanroh, Yzd-vo TRTU, 2006, No. 10 (65), pp. 43–45.
3. Yakhyna E.P. Metody otsenky ynformatsyonnykh system, *V myre nauchnykh otkrytyi,* 2010, No. 3 (09), Chast 1, pp. 63–66.
4. Popov S. V. Otsenka funktsyonalnoi effektyvnosty system tekstovoho poyska na prymere poyska patentnykh dokumentov, *Patentnaia informatsyia sehodnia,* 2010, No. 1, pp. 22–25.
5. Kozlov D. D. Ynformatsyonno-poyskovye systemy v Internet: tekushchee sostoianye i puty razvytyia, *Tekhnolohycheskyi obzor.* Access mode: lvk.cs.msu.su/~ddk/ir_and_ia_review.pdf
6. Tiavkyn Y. V., Tiutiunnyk V. M. Matematycheskaia model informatsyonnoho poiska i otsenka yeffektyvnosti poyskovoi sistemy, *Vestnyk THTU,* 2008, Vol 14, № 3, pp. 478–481.
7. Kozlov M. V., Yatsko V. A. Metod otsenky effektyvnosti funktsyonyrovanyia sovremennykh informatsyonno-poyskovykh system Interneta, Access mode: http://www.dialog-21.ru/dialog2006/materials/html/Kozlov.htm.
8. Lektsyy po vvedenyiu v informatyku i informatsyonnыe systemy. Lektsyia 13. Yeffektyvnost ynformatsyonnыkh system. Access mode: http://informling.narod.ru/lectures.html
9. Kirchgassner G., Wolters J. Introduction to Modern Time Series Analysis. Springer Berlin Heidelberg, New York , 2007, 274 p.
10. Hegger R., Kantz H., Schreiber T. Practical implementation of nonlinear time series methods: The TISEAN package. *CHAOS 9*, 1999, pp. 413–435.
11. Kuhlthau, C. C. Seeking Meaning: A Process Approach to Library and Information Services, *2nd. ed. Westport*, CT, Libraries Unlimited, 2004, 342 p.

УПРАВЛІННЯ У ТЕХНІЧНИХ СИСТЕМАХ

УПРАВЛЕНИЕ В ТЕХНИЧЕСКИХ СИСТЕМАХ

CONTROL IN TECHNICAL SYSTEMS

УДК 004.052: 656.61

Kołowrocki K.[1], Kwiatuszewska-Sarnecka B.[2], Soszyńska-Budny J.[3]

[1]Dr. hab. Sc., Professor, Professor of Department of Mathematics, Gdynia Maritime University, Gdynia, Poland
[2]PhD, Assistant Professor, Assistant Professor of Department of Mathematics, Gdynia Maritime University, Gdynia, Poland
[3]PhD., Associate Professor of Department of Mathematics, Gdynia Maritime University, Gdynia, Poland

RELIABILITY AND RISK OPTIMIZATION OF MULTISTATE SYSTEMS WITH APPLICATION TO PORT TRANSPORTATION SYSTEM

The complexity of technical systems' operation processes and its influence on the changing in time systems' structures and their components' reliability parameters posses a difficulty to first meet in real and then to fix and analyse those structures and reliability parameters. By constructing a joint model of reliability of complex technical systems at variable operation conditions, which links a semi-markov modelling of system operation processes with multi-state approach to system reliability analysis, we find the system's main reliability characteristics. Consequently, we use linear programming to build a model of complex technical systems reliability optimization. We investigate the model's application in marine transport, specifically in reliability and risk optimization of a bulk cargo transportation system. The tools we develop can be used in reliability evaluation and optimization of a very wide class of real technical systems operating at varying conditions that influence their reliability structures and the reliability parameters of their components. Consequently, the tools we developed can be implemented by reliability practitioners from both maritime transport industry and other industrial sectors.

Keywords: reliability function, risk function, operation process, optimization.

NOMENCLATURE

$Z(t), z_i$ is a system operation process; a system operation state; $i = 1,2,..., v$,

$\dot{p}_b$ is an optimal transient probability;

U is a particular reliability state of the system; $u = 1,2,..., z$;

$T(u)$ is a lifetime of a system in the reliability state subset $\{u, u+1,..., z\}$;

$[\boldsymbol{R}(t,u)]^{(b)}$ is a conditional reliability function of a system at the operational state z_b;

$\dot{\boldsymbol{R}}(t,u)$ is an optimal unconditional reliability function of a system;

$\mu(u)$ is a mean lifetime of the system in the reliability state subset $\{u, u+1,..., z\}$;

$\overline{\mu}(u)$ is a mean lifetime of the system in the reliability particular state u;

R is a critical state of the multi-state system;

$\dot{r}(t)$ is an optimal risk function of the multi-state system.

INTRODUCTION

Most real technical systems are very complex because they are composed of large numbers of components and subsystems and have high operating complexity. The complexity of the systems' operation processes and its influence on the changing in time systems' structures and their components' reliability parameters posses a difficulty to first meet in real and then to fix and analyse those structures and reliability parameters. A convenient tool to investigate this problem is a semi-markov [2] modelling of the system operation process linked with a multi-state approach for the system reliability analysis [1, 4, 9–10] and a linear programming for the system reliability optimization [3]. Using this approach, it is possible to find this complex system's main reliability characteristics including the system reliability function, the system mean lifetimes in the reliability states subsets and the system risk function [4, 6, 8]. Having those characteristics it is possible to optimize the system operation process to get optimal values [8]. To this end the linear programming [3] can be applied to maximize the mean value of the system lifetime in the subset of the system reliability states, which are not worse than the system critical reliability state.

DOI 10.15588/1607-3274-2016-2-14

1 SYSTEM RELIABILITY AT VARIABLE OPERATIONS PROCESS

We suppose that the system has v different operation states during its operation process. Thus, we can define the system operation process $Z(t)$, $t \in <0,+\infty>$, as the process with discrete operation states from the set $Z = \{z_1, z_2, \ldots, z_\upsilon\}$. Further, we assume that $Z(t)$ is a semi-markov process [2] with its conditional sojourn times θ_{bl} at the operation state z_b when its next operation state is z_l, $b, l = 1,2,...,v$, $b \neq l$. In this case the process $Z(t)$ may be described by:

– the vector of probabilities of the process initial operation states $[p_b(0)]_{1xv}$,

– the matrix of probabilities of the process transitions between the operation states $[p_{bl}]_{vxv}$,

– the matrix of conditional distribution functions $[H_{bl}(t)]_{vxv}$ of the process sojourn times θ_{bl}, $b \neq l$, in the operation state z_b when the next operation state is z_l.

Under these assumptions, the sojourn times θ_{bl} mean values are given by

$$M_{bl} = E[\theta_{bl}] = \int_0^\infty t dH_{bl}(t), \; b, l = 1,2,...,v, \; b \neq l. \qquad (1)$$

The mean values $E[\theta_b]$ of the unconditional sojourn times θ_b are given by

$$M_b = E[\theta_b] = \sum_{l=1}^{v} p_{bl} M_{bl}, \; b = 1,2,...,v, \qquad (2)$$

where M_{bl} are defined by (1).

Limit values of the transient probabilities at the operation states are given by

$$p_b = \lim_{t\to\infty} p_b(t) = \frac{\pi_b M_b}{\sum_{l=1}^{v} \pi_l M_l}, \; b = 1,2,...,v,$$

where the probabilities π_b of the vector $[\pi_b]_{1xv}$ satisfy the system of equations $[\pi_b] = [\pi_b][p_{bl}]$ and $\sum_{l=1}^{v} \pi_l = 1$ and $p_b(t) = P(Z(t) = z_b)$, $t \in <0,+\infty)$, $b = 1,2,...,v$.

We assume that the system is composed of n independent multistate components E_i, $i = 1,2,...,n$, and that the changes of the operation process $Z(t)$ states have an influence on both the system components E_i reliability and on the system reliability structure. Consequently, we denote the component E_i lifetime in the reliability state subset $\{u, u+1,...,z\}$ by $T_i^{(b)}(u)$ and by

$$[R_i(t,\cdot)]^{(b)} = [1, [R_i(t,1)]^{(b)}, [R_i(t,2)]^{(b)}, ..., [R_i(t,z)]^{(b)}],$$

where for $t \in <0,\infty)$, $b = 1,2,\ldots,v$, $u = 1,2,...,z$,

$$[R_i(t,u)]^{(b)} = P(T_i^{(b)}(u) > t \,|\, Z(t) = z_b),$$

is the conditional reliability function while the system is at the operational state z_b, $b = 1,2,\ldots,v$.

Next, we denote the system lifetime in the reliability state subset $\{u, u+1,...,z\}$ by $T^{(b)}(u)$ and by

$$[\boldsymbol{R}(t,\cdot)]^{(b)} = [1, [\boldsymbol{R}(t,1)]^{(b)}, [\boldsymbol{R}(t,2)]^{(b)}, ..., [\boldsymbol{R}(t,z)]^{(b)}],$$

where for $t \in <0,\infty)$, $b = 1,2,\ldots,v$, $u = 1,2,...,z$,

$$[\boldsymbol{R}(t,u)]^{(b)} = P(T^{(b)}(u) > t \,|\, Z(t) = z_b),$$

is the conditional reliability function of the system while the system is at the operational state z_b.

In the case when the system operation time is large enough, the unconditional reliability function of the system is given by

$$\boldsymbol{R}(t,\cdot) = [1, \boldsymbol{R}(t,1), \boldsymbol{R}(t,2), ..., \boldsymbol{R}(t,z)], \; t \geq 0,$$

where

$$\boldsymbol{R}(t,u) \cong \sum_{b=1}^{v} p_b [\boldsymbol{R}(t,u)]^{(b)}. \qquad (3)$$

The mean values of the system lifetimes in the reliability state subset $\{u, u+1,...,z\}$ are

$$\mu(u) = E[T(u)] \cong \sum_{b=1}^{v} p_b \mu_b(u), \; u = 1,2,...,z, \qquad (4)$$

and the mean values of the system lifetimes in the particular reliability state u, are [4]

$$\overline{\mu}(u) = \mu(u) - \mu(u+1), \; u = 1,2,...,z-1, \; \overline{\mu}(z) = \mu(z). \qquad (5)$$

A probability $\boldsymbol{r}(t) = P(s(t) < r \mid R(0) = z) = P(T^{(b)}(r) \leq t)$, $t \in (-\infty, \infty)$,
that the system is in the subset of reliability states worse than the critical state r, $r \in \{1,...,z\}$ while it was in the state z at the moment $t = 0$ is called a risk function of the multi-state system [4].

Under this definition, from (3), we have

$$\boldsymbol{r}(t) = 1 - \boldsymbol{R}(t,r), \; t \in (-\infty, \infty), \qquad (6)$$

and if τ is the moment when the risk exceeds a permitted level δ, then

$$\tau = \boldsymbol{r}^{-1}(\delta), \qquad (7)$$

where $\boldsymbol{r}^{-1}(t)$, if it exists, is the inverse function of the risk function $\boldsymbol{r}(t)$.

2 OPTIMAL TRANSIENT PROBABILITIES MAXIMIZING SYSTEM LIFETIME

Considering the equation (3), it is natural to assume that the system operation process has a significant influence on the system reliability. This influence is also clearly expressed in the equation (4) for the mean values of the system

unconditional lifetimes in the reliability state subsets. From linear equation (4), we can see that the mean value of the system unconditional lifetime $\mu(u)$, $u = 1,2,..., z$, is determined by the limit values of transient probabilities p_b, $b = 1,2,\ldots, \nu$, of the system operation states and the mean values $\mu_b(u)$, $b = 1,2,\ldots, \nu$, $u = 1,2,..., z$, of the system conditional lifetimes in the reliability state subsets $\{u, u+1,..., z\}$, $u = 1,2,..., z$. Therefore, the system lifetime optimization approach based on the linear programming can be proposed. Namely, we may look for the corresponding optimal values $\dot{p}_b$ of the transient probabilities p_b in the system operation states to maximize the mean value $\mu(u)$ of the unconditional system lifetimes in the reliability state subsets $\{u, u+1,..., z\}$ under the assumption that the mean values $\mu_b(u)$ of the system conditional lifetimes in the reliability state subsets are fixed. As a special case of the above formulated system lifetime optimization problem: if r, $r = 1,2,..., z$, is a system critical reliability state, then we want to find the optimal values $\dot{p}_b$ of the transient probabilities p_b in the system operation states to maximize the mean value $\mu(r)$ of the unconditional system lifetime in the reliability state subset $\{r, r+1,,..., z\}$ under the assumption that the mean values $\mu_b(r)$, $b = 1,2,\ldots, \nu$, of the system conditional lifetimes in this reliability state subset are fixed. More exactly, we formulate the optimization problem as a linear programming model with the objective function of the following linear form

$$\mu(r) = \sum_{b=1}^{\nu} p_b \mu_b(r) \tag{8}$$

for a fixed $r \in \{1,2,..., z\}$ and with the following bound constraints

$$\sum_{b=1}^{\nu} p_b = 1, \quad \breve{p}_b \le p_b \le \hat{p}_b, \quad b = 1,2,\ldots, \nu, \tag{9}$$

where $\mu_b(r)$, $\mu_b(r) \ge 0$, $b = 1,2,..., \nu$, are fixed mean values of the system conditional lifetimes in the reliability state subset $\{r, r+1,..., z\}$ and

$$\breve{p}_b,\ 0 \le \breve{p}_b \le 1 \text{ and } \hat{p}_b,\ 0 \le \hat{p}_b \le 1,\ \breve{p}_b \le \hat{p}_b,\ b=1,2,\ldots,\nu, \tag{10}$$

are respectively the lower and upper bounds of the unknown transient probabilities p_b.

Now, we can obtain the optimal solution to the formulated by (8)–(10) the linear programming problem, i.e. we can find the optimal values $\dot{p}_b$ of the transient probabilities p_b, $b = 1,2,\ldots, \nu$, that maximize the objective function given by (8). First, we arrange the system conditional lifetime mean values $\mu_b(r)$, $b = 1,2,\ldots,\nu$, in non-increasing order

$$\mu_{b_1}(r) \ge \mu_{b_2}(r) \ge \ldots \ge \mu_{b_\nu}(r), \quad b_i \in \{1,2,..., \nu\} \text{ for } i = 1,2,..., \nu.$$

Next, we substitute

$$x_i = p_{b_i},\ \breve{x}_i = \breve{p}_{b_i},\ \hat{x}_i = \hat{p}_{b_i} \text{ for } i = 1,2,..., \nu \tag{11}$$

and we maximize with respect to x_i, $i = 1,2,..., \nu$, the linear form (8) that takes the form

$$\mu(r) = \sum_{i=1}^{\nu} x_i \mu_{b_i}(r) \tag{12}$$

for a fixed $r \in \{1,2,..., z\}$ with the following bound constraints

$$\sum_{i=1}^{\nu} x_i = 1, \quad \breve{x}_i \le x_i \le \hat{x}_i, \quad i = 1,2,..., \nu, \tag{13}$$

where $\mu_{b_i}(r)$, $\mu_{b_i}(r) \ge 0$, $i = 1,2,..., \nu$, are fixed mean values of the system conditional lifetimes in the reliability state subset $\{r, r+1,..., z\}$ arranged in non-increasing order and

$$\breve{x}_i,\ 0 \le \breve{x}_i \le 1 \text{ and } \hat{x}_i,\ 0 \le \hat{x}_i \le 1,\ \breve{x}_i \le \hat{x}_i,\ i = 1,2,..., \nu, \tag{14}$$

are the lower and upper bounds of the unknown probabilities x_i, $i = 1,2,..., \nu$, respectively.

We define

$$\breve{x} = \sum_{i=1}^{\nu} \breve{x}_i,\ \hat{y} = 1 - \breve{x} \tag{15}$$

and

$$\breve{x}^0 = 0,\ \hat{x}^0 = 0 \text{ and } \breve{x}^I = \sum_{i=1}^{I} \breve{x}_i,\ \hat{x}^I = \sum_{i=1}^{I} \hat{x}_i \text{ for } I = 1,2,..., \nu. \tag{16}$$

Next, we find the largest value $I \in \{0,1,..., \nu\}$ such that

$$\hat{x}^I - \breve{x}^I < \hat{y} \tag{17}$$

and we fix the optimal solution that maximize (12) in the following way:

i) if $I = 0$, the optimal solution is $\dot{x}_1 = \hat{y} + \breve{x}_1$ and $\dot{x}_i = \breve{x}_i$ for $i = 2,3,..., \nu$; (18)

ii) if $0 < I < \nu$, the optimal solution is

$$\dot{x}_i = \hat{x}_i \text{ for } i = 1,2,..., I,\ \dot{x}_{I+1} = \hat{y} - \hat{x}^I + \breve{x}^I + \breve{x}_{I+1} \text{ and } \dot{x}_i = \breve{x}_i \text{ for } i = I+2, I+3,..., \nu; \tag{19}$$

iii) if $I = \nu$, the optimal solution is $\dot{x}_i = \hat{x}_i$ for $i = 2,3,..., \nu$. (20)

Finally, after making the inverse to (11) substitution, we get the optimal limit transient probabilities

$$\dot{p}_{b_i} = \dot{x}_i \text{ for } i = 1,2,..., \nu, \tag{21}$$

that maximize the system mean lifetime $\mu(r)$ in the reliability state subset $\{r, r+1,..., z\}$, defined by the linear form (8) giving its maximum value in the following form

$$\dot{\mu}(r) = \sum_{b=1}^{\nu} \dot{p}_b \mu_b(r) \text{ for a fixed } r \in \{1,2,..., z\}. \tag{22}$$

From the above, replacing r by u, $u = 1,2,..., z$, we obtain the corresponding optimal solutions for the mean values of the system unconditional lifetimes in the reliability state subsets $\{u, u+1,..., z\}$ of the form

$$\dot{\mu}(u) = \sum_{b=1}^{\nu} \dot{p}_b \mu_b(u) \text{ for } u = 1,2,..., z. \quad (23)$$

Further, according to (3), the corresponding optimal unconditional multistate reliability function of the system is

$$\dot{\boldsymbol{R}}(t,\cdot) = [1, \dot{\boldsymbol{R}}(t,1),..., \dot{\boldsymbol{R}}(t,z)], \quad (24)$$

where

$$\dot{\boldsymbol{R}}_{\boldsymbol{n}}(t,\cdot) \cong \sum_{b=1}^{\nu} \dot{p}_b [\boldsymbol{R}(t,u)]^{(b)} \text{ for } t \geq 0, \ u = 1,2,..., z, \quad (25)$$

and by (5) the optimal solutions for the mean values of the system unconditional lifetimes in the particular reliability states are of the form

$$\dot{\bar{\mu}}(u) = \dot{\mu}(u) - \dot{\mu}(u+1), \ u = 0,1,...,z-1, \ \dot{\bar{\mu}}(z) = \dot{\mu}(z). \quad (26)$$

Moreover, considering (6) and (7), the corresponding optimal system risk function and the moment when the risk exceeds a permitted level δ, respectively are given by

$$\dot{\boldsymbol{r}}(t) = 1 - \dot{\boldsymbol{R}}(t,r) \text{ for } t \in (-\infty,\infty) \text{ and } \dot{\tau} = \dot{\boldsymbol{r}}^{-1}(\delta). \quad (27)–(28)$$

3 OPTIMAL SOJOURN TIMES OF COMPLEX TECHNICAL SYSTEM OPERATION PROCESS

Replacing the limit transient probabilities p_b of the system operation process at the operation states by their optimal values $\dot{p}_b$ and the mean values M_b of the unconditional sojourn times at the operation states by their corresponding unknown optimal values $\dot{M}_b$ maximizing the mean value of the system lifetime in the reliability states subset $\{r, r+1,..., z\}$, we get the system of equations

$$\dot{p}_b = \frac{\pi_b \dot{M}_b}{\sum_{l=1}^{\nu} \pi_l \dot{M}_l}, \ b = 1,2,..., \nu. \quad (29)$$

After simple transformations the above system takes the form

$$\begin{cases} (\dot{p}_1 - 1)\pi_1 \dot{M}_1 + \dot{p}_1 \pi_2 \dot{M}_2 + ... + \dot{p}_1 \pi_\nu \dot{M}_\nu = 0 \\ \dot{p}_2 \pi_1 \dot{M}_1 + (\dot{p}_2 - 1)\pi_2 \dot{M}_2 + ... + \dot{p}_2 \pi_\nu \dot{M}_\nu = 0 \\ . \\ \dot{p}_\nu \pi_1 \dot{M}_1 + \dot{p}_\nu \pi_2 \dot{M}_2 + ... + (\dot{p}_\nu - 1)\pi_\nu \dot{M}_\nu = 0, \end{cases} \quad (30)$$

where $\dot{M}_b$ are unknown variables we want to find, $\dot{p}_b$ are optimal transient probabilities and π_b are steady probabilities.

Since the system of equations is homogeneous and it can be proved that the determinant of its main matrix is equal to zero, then it has nonzero solutions and moreover, these solutions are ambiguous. Thus, if we fix some of the optimal values $\dot{M}_b$ of the mean values M_b of the unconditional sojourn times at the operation states, for instance by arbitrary fixing either one or multiple of them, we may find the values of the remaining ones and using this method arrive at the solution of this equation.

Another very useful and much easier applicable in practice tool that can help in planning the operation processes of complex technical systems are the system operation process optimal mean values of the total system operation process sojourn times $\hat{\theta}_b$ at the particular operation states z_b, $b = 1,2,..., \nu$, during the fixed system operation time θ. They can be obtained by replacing the transient probabilities p_b at the operation states z_b with their optimal values $\dot{p}_b$. This results in the following expession

$$\dot{\hat{M}}_b = \dot{E}[\hat{\theta}_b] = \dot{p}_b \theta, \ b = 1,2,..., \nu. \quad (31)$$

The knowledge of the optimal values $\dot{M}_b$ of the mean values of the unconditional sojourn times and the optimal mean values $\dot{\hat{M}}_b$ of the total sojourn times at the particular operation states during the fixed system operation time may be the basis for changing the complex technical systems operation processes in order to ensure that these systems operate both more reliably and more safely. This knowledge may also be useful in these systems operation cost analysis.

4 THE BULK CARGO TRANSPORTATION SYSTEM RELIABILITY AND RISK

The considered bulk cargo terminal placed at the Baltic seaside is designated for storage and reloading of bulk cargo, but its primary activity is loading bulk cargo on board the ships for export. There are two independent transportation systems: the system of reloading rail wagons and the system of loading vessels.

Cargo is brought to the terminal by trains consisting of self-discharging wagons, which are discharged to a hopper and then by means of conveyors transported into one of four storage tanks (silos). Loading of fertilizers from storage tanks on board the ship is done by means of special reloading system which consists of several belt conveyors and one bucket conveyor which allows the transfer of bulk cargo in a vertical direction. Researched system is a system of belt conveyors, referred to as the transport system.

In the conveyor reloading system we distinguish three bulk cargo transportation subsystems, the belt conveyors S_1, S_2 and S_3. The conveyor loading system is composed of six bulk cargo transportation subsystems, the dosage conveyor S_4, the horizontal conveyor S_5, the horizontal conveyor S_6, the sloping conveyors S_7, the dosage conveyor with buffer S_8, the loading system S_9.

The bulk cargo transportation subsystems are built, respectively:

– the subsystem S_1 : 1 rubber belt, 2 drums, set of 121 bow rollers, set of 23 belt supporting rollers,

– the subsystem S_2: 1 rubber belt, 2 drums, set of 44 bow rollers, set of 14 belt supporting rollers,

– the subsystem S_3: 1 rubber belt, 2 drums, set of 185 bow rollers, set of 60 belt supporting rollers,

– the subsystem S_4: 3 identical belt conveyors, each composed of 1 rubber belt, 2 drums, set of 12 bow rollers, set of 3 belt supporting rollers,

– the subsystem S_5: 1 rubber belt, 2 drums, set of 125 bow rollers, set of 45 belt supporting rollers,

– the subsystem S_6: 1 rubber belt, 2 drums, set of 65 bow rollers, set of 20 belt supporting rollers,

– the subsystem S_7: 1 rubber belt, 2 drums, set of 12 bow rollers, set of 3 belt supporting rollers,

– the subsystem S_8: 1 rubber belt, 2 drums, set of 162 bow rollers, set of 53 belt supporting rollers,

– the subsystem S_9: 3 rubber belts, 6 drums, set of 64 bow rollers, set of 20 belt supporting rollers.

Taking into account the operation process of the considered system we distinguish the following as its three operation states:

– an operation state z_1 – loading fertilizers from rail wagons on board the ship is done using S_1, S_2, S_3, S_6, S_7, S_8 and S_9 subsystems.

– an operation state z_2 – discharging rail wagons to storage tanks or hall when subsystems S_1, S_2 and S_3, are used,

– an operation state z_3 – loading fertilizers from storage tanks or hall on board the ship is done by using S_4, S_5, S_6, S_7, S_8 and S_9, subsystems.

The limit values of the bulk cargo transportation systems operation process transient probabilities $p_b(t)$ at the operation states z_b, $b = 1,2,3$, determined in [5], on the basis of data coming from experts are

$$p_1 = 0.2376,\ p_2 = 0.6679,\ p_3 = 0.0945. \qquad (32)$$

Further, assuming that the system is in the reliability state subset $\{u,u+1,\ldots,z\}$ if all its subsystems are in this subset of reliability states, we conclude that the bulk cargo transportation system is a series system [4] of subsystems S_1, S_2, S_3, S_6, S_7, S_8 and S_9.

Under the assumption that changes of the bulk cargo transportation system operation states have an influence on both the subsystem S_i reliability and the entire reliability structure [8], on the basis of expert opinions and statistical data given in [9], [10], the bulk cargo transportation system reliability structures and their components reliability functions at different operation states can be determined.

Additionally, we assume that subsystems S_i, $i = 1,2,3,...,9$, are composed of four-state exponential components, with the reliability functions

$$[R_i(t,\cdot)]^{(b)} = [1, [R_i(t,1)]^{(b)}, [R_i(t,2)]^{(b)}, [R_i(t,3)]^{(b)}],$$

$$t \in <0,\infty),\ b = 1,2,3,\ u = 1,2,3.$$

At the operation state z_1, at loading of fertilizers from rail wagons on board the ship, the system is composed of seven non-homogenous series subsystems S_1, S_2, S_3, S_6, S_7, S_8, and S_9 forming a series structure.

The conditional reliability function of the system while it is at the operation state z_1 is given by

$$[\boldsymbol{R}(t,\cdot)]^{(1)} = [1, [\boldsymbol{R}(t,1)]^{(1)}, [\boldsymbol{R}(t,2)]^{(1)}, [\boldsymbol{R}(t,3)]^{(1)}],$$

where

$$[\boldsymbol{R}(t,u)]^{(1)} = [\boldsymbol{R}_{147}(t,u)]^{(1)}\,[\boldsymbol{R}_{61}(t,u)]^{(1)}\,[\boldsymbol{R}_{248}(t,u)]^{(1)}\,[\boldsymbol{R}_{88}(t,u)]^{(1)}$$

$$[\boldsymbol{R}_{18}(t,u)]^{(1)}[\boldsymbol{R}_{218}(t,u)]^{(1)}[\boldsymbol{R}_{93}(t,u)]^{(1)} \text{for } t \in <0,\infty), u = 1,2,3,$$

i.e.

$$[\boldsymbol{R}(t,1)]^{(1)} = \exp[-74.426t],\ [\boldsymbol{R}(t,2)]^{(1)} = \exp[-93.472t],$$

$$[\boldsymbol{R}(t,3)]^{(1)} = \exp[-150.206t]. \qquad (33)–(35)$$

The expected values of the conditional lifetimes in the reliability state subsets at the operation state z_1, calculated from the above result given by (33)–(35), are:

$$\mu_1(1) \cong 0.013,\ \mu_1(2) \cong 0.011,\ \mu_1(3) \cong 0.007 \text{ years}, \qquad (36)$$

and further, using (5), it follows that the conditional lifetimes in the particular reliability states at the operation state z_1 are:

$$\overline{\mu}_1(1) \cong 0.002,\ \overline{\mu}_1(2) \cong 0.004,\ \overline{\mu}_1(3) \cong 0.007 \text{ years}.$$

At the operation state z_2, i.e. at the state of discharging rail wagons to storage tanks or hall, the system is built of three subsystems S_1, S_2 and S_3 forming a series structure [4].

The conditional reliability function of the bulk cargo transportation system at the operation state z_2 is given by

$$[\boldsymbol{R}(t,\cdot)]^{(2)} = [1, [\boldsymbol{R}(t,1)]^{(2)}, [\boldsymbol{R}(t,2)]^{(2)}, [\boldsymbol{R}(t,3)]^{(2)}],$$

where

$$[\boldsymbol{R}(t,u)]^{(2)} = [\boldsymbol{R}_{147}(t,u)]^{(2)}\ [\boldsymbol{R}_{61}(t,u)]^{(2)}\ [\boldsymbol{R}_{248}(t,u)]^{(2)}$$

$$\text{for } t \in <0,\infty),\ u = 1,2,3,$$

i.e.

$$[\boldsymbol{R}(t,1)]^{(2)} = \exp[-39.563t],\ [\boldsymbol{R}(t,2)]^{(2)} = \exp[-49.663t],$$

$$[\boldsymbol{R}(t,3)]^{(2)} = \exp[-64.280t]. \qquad (37)–(39)$$

The expected values of the conditional lifetimes in the reliability state subsets at the operation state z_2, calculated from the above result given by (37)–(39), are:

$$\mu_2(1) \cong 0.025,\ \mu_2(2)\ 0.020, 0.016 \text{ years}, \qquad (40)$$

and further, using (5), it follows that the conditional lifetimes in the particular reliability states at the operation state z_2 are:

$$\overline{\mu}_2(1) \cong 0.005,\ \overline{\mu}_2(2) \cong 0.004,\ \overline{\mu}_2(3) \cong 0.016 \text{ years}.$$

At the operation state z_3, i.e. at the loading of fertilizers from storage tanks or hall on board, the bulk cargo transportation system is built of six subsystems one series-parallel subsystem S_4 and five series subsystems S_5, S_6, S_7, S_8, S_9 forming a series structure [4].

The conditional reliability function of the system while it is at the operation state z_3 is given by

$$[\boldsymbol{R}(t,\cdot)]^{(3)} = [1,\ [\boldsymbol{R}(t,1)]^{(3)}, [\boldsymbol{R}(t,2)]^{(3)}, [\boldsymbol{R}(t,3)]^{(3)}],$$

where

$$[\boldsymbol{R}(t,u)]^{(3)} = [\boldsymbol{R}_{3,18}(t,u)]^{(3)} \cdot [\boldsymbol{R}_{173}(t,u)]^{(3)} \cdot [\boldsymbol{R}_{88}(t,u)]^{(3)} \cdot [\boldsymbol{R}_{18}(t,u)]^{(3)}$$

$$\cdot [\boldsymbol{R}_{218}(t,u)]^{(3)} \cdot [\boldsymbol{R}_{93}(t,u)]^{(3)} \text{ for } t \in <0,\infty),\ u = 1,2,3,$$

i.e.

$$[\boldsymbol{R}(t,1)]^{(3)} = \exp[-57.758t] - 3\exp[-55.007t] + 3\exp[-52.256t], \quad (41)$$

$$[\boldsymbol{R}(t,2)]^{(3)} = \exp[-70.974t] - 3\exp[-68.018t] + 3\exp[-65.062t] \quad (42)$$

$$[\boldsymbol{R}(t,3)]^{(3)} = \exp[-89.416t] - 3\exp[-86.140t] + 3\exp[-82.864t]. \quad (43)$$

The expected values of the conditional lifetimes in the reliability state subsets at the operation state z_3, calculated from the above result given by (41)–(43), are:

$$\mu_3(1) \cong 0.020,\ \mu_3(2) \cong 0.016,\ \mu_3(3) \cong 0.013 \text{ years}, \quad (44)$$

and further, using (5), it follows that the conditional lifetimes in the particular reliability states at the operational state z_3 are:

$$\overline{\mu}_3(1) \cong 0.004,\ \overline{\mu}_3(2) \cong 0.003,\ \overline{\mu}_3(3) \cong 0.013 \text{ years}.$$

In the case when the system operation time is large enough, the unconditional reliability function of the bulk cargo transportation system is given by the vector

$$\boldsymbol{R}(t,\cdot) = [1,\ \boldsymbol{R}(t,1),\ \boldsymbol{R}(t,2),\ \boldsymbol{R}(t,3)],\ t \geq 0,$$

where, according to (3) and after considering the values of $p_b, b = 1,2,3$, given by (32), its co-ordinates are as follows:

$$\boldsymbol{R}(t,u) = p_1 \cdot [\boldsymbol{R}(t,u)]^{(1)} + p_2 \cdot [\boldsymbol{R}(t,u)]^{(2)} + p_3 \cdot [\boldsymbol{R}(t,u)]^{(3)} \quad (45)$$

for $t \geq 0,\ u = 1,2,3$, where $[\boldsymbol{R}(t,u)]^{(1)}$ and $[\boldsymbol{R}(t,u)]^{(2)}$ and $[\boldsymbol{R}(t,u)]^{(3)}$ are respectively given by (33)–(35) and (37)–(39) and (41)–(43), i.e.

$$\boldsymbol{R}(t,1) = 0.6679\exp[-39.563t] + 0.0945\exp[-74.426t] +$$
$$+ 0.2376[\exp[-57.758t] - 3\exp[-55.007t] + 3\exp[-52.256t]], \quad (46)$$

$$\boldsymbol{R}(t,2) = 0.6679\exp[-93.472t] + 0.0945\exp[-49.663t] +$$
$$+ 0.2376[\exp[-70.974t] - 3\exp[-68.018t] + 3\exp[-65.062t]], \quad (47)$$

$$\boldsymbol{R}(t,3) = 0.6679\exp[-150.206t] + 0.0945\exp[-64.280t] +$$
$$+ 0.0945[\exp[-89.416t] - 3\exp[-86.140t] + 3\exp[-82.864t]]. \quad (48)$$

The mean values of the system unconditional lifetimes in the reliability state subsets, according to (4) are respectively:

$$\mu(1) \cong 0.016,\ \mu(2) \cong 0.013,\ \mu(3) \cong 0.009. \quad (49)$$

The mean values of the system lifetimes in the particular reliability states, (5), are

$$\overline{\mu}(1) = \mu(1) - \mu(2) = 0.003,\ \overline{\mu}(2) = \mu(2) - \mu(3) = 0.004,$$
$$\overline{\mu}(3) = \mu(3) = 0.009.$$

If the critical reliability state is $r = 2$, then the system risk function, according to (6), is given by

$$\boldsymbol{r}(t) = 1 - [0.6679\exp[-93.472t] + 0.0945\exp[-49.663t] +$$
$$+ 0.2376(\exp[-70.974t] - 3\exp[-68.018t] + 3\exp[-65.062t])]$$
$$\text{for } t \geq 0. \quad (50)$$

Hence, the moment when the system risk function (Fig. 1) exceeds a permitted level, for instance $\delta = 0.05$, from (7), is

$$\tau = \boldsymbol{r}^{-1}(\delta) \cong 0.000627 \text{ years}. \quad (51)$$

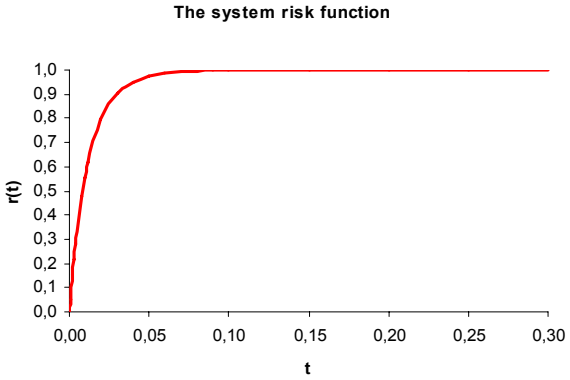


Figure 1 – The graph of the port bulk cargo transportation system risk function

5 OPTIMIZATION OF THE BULK CARGO TRANSPORTATION SYSTEM OPERATION PROCESS

In our case, as the critical state is $r = 2$, then considering the expression for $\mu(2)$, the objective function (8), takes the form

$$\mu(2) = p_1 \cdot 0.011 + p_2 \cdot 0.020 + p_3 \cdot 0.016 = 0.013 \text{ years}. \quad (52)$$

The lower $\breve{p}_b$ and upper $\hat{p}_b$ bounds of the unknown transient probabilities $p_b,\ b = 1,2,3$, coming from experts, respectively are [6]:

$$\breve{p}_1 = 0.150,\ \breve{p}_2 = 0.005,\ \breve{p}_3 = 0.015,$$
$$\hat{p}_1 = 0.850,\ \hat{p}_2 = 0.120,\ \hat{p}_3 = 0.390.$$

Therefore, according to (9)–(10), we assume the following bound constraints

$$\sum_{b=1}^{3} p_b = 1,\ 0.250 \leq p_1 \leq 0.850,$$
$$0.005 \leq p_2 \leq 0.150,\ 0.050 \leq p_3 \leq 0.550.$$

Now, before we find optimal values $\dot{p}_b$ of the transient probabilities $p_b,\ b = 1,2,3$, that maximize the objective function (53), we arrange the system conditional lifetimes mean values $\mu_b(2),\ b = 1,2,3$, in non-increasing order $\mu_2(2) \geq \mu_3(2) \geq \mu_1(2)$.

Next, according to (11), we substitute

$$x_1 = p_2 = 0.0945,\ x_2 = p_3 = 0.2376,\ x_3 = p_1 = 0.6679, \quad (53)$$

$$\breve{x}_1 = 0.005,\ \ \breve{x}_2 = 0.050,\ \ \breve{x}_3 = 0.250,\ \ \hat{x}_1 = 0.150,$$
$$\hat{x}_2 = 0.550,\ \ \hat{x}_3 = 0.850, \qquad (54)$$

where $\breve{x}_i$ and $\hat{x}_i$ are lower and upper bounds of the unknown limit transient probabilities x_i, $i = 1,2,3$, respectively and we maximize with respect to x_i, $i = 1,2,3$, the linear form (52) that according to (13) takes the form

$$\mu(2) = x_1 \cdot 0{,}011 + x_2 \cdot 0{,}020 + x_3 \cdot 0{,}016, \qquad (55)$$

with the following bound constraints

$$\sum_{i=1}^{3} x_i = 1,$$

$$0.005 \le x_1 \le 0.12,\ 0.015 \le x_2 \le 0.390,\ 0.150 \le x_3 \le 0.850. \qquad (56)$$

According to (15)–(17), we calculate and fix the optimal solution that maximizes linear function (55) according to the rule (19). Namely, we get

$$\dot{x}_1 = \hat{x}_1 = 0.120\,,\ \dot{x}^2 = \hat{x}^2 = 0.390\,,$$
$$\dot{x}_3 = 0.830 - 0.490 + 0.150 = 0.490.$$

Finally, according to (21) after making the inverse to (53) substitution, we get the optimal transient probabilities

$$\dot{p}_1 = \dot{x}_3 = 0.490,\ \dot{p}_2 = \dot{x}_1 = 0.120,\ \dot{p}_3 = \dot{x}_2 = 0.390, \qquad (57)$$

that maximize the system mean lifetime in the reliability state subset $\{2,3\}$ expressed by the linear form (53) giving, according to (12) and (57), its optimal value

$$\dot{\mu}(2) = \dot{p}_1 \cdot 0{,}011 + \dot{p}_2 \cdot 0{,}020 + \dot{p}_3 \cdot 0{,}016 = 0.49 \cdot 0.011 +$$
$$+ 0.12 \cdot 0.020 + 0.39 \cdot 0.016 = 0.014. \qquad (58)$$

6 OPTIMAL RELIABILITY CHARACTERISTICS OF THE BULK CARGO TRANSPORTATION SYSTEM

Further, substituting the optimal solution (57) according to (24), we obtain the optimal solution for the mean value of the system unconditional lifetime in the reliability state subset $\{1,2\}$, $\{3\}$ that respectively amounts:

$$\dot{\mu}(1) \cong 0.0172,\ \dot{\mu}(3) \cong 0.0104, \qquad (59)$$

and according to (26), the optimal solutions for the mean values of the system unconditional lifetimes in the particular reliability states are

$$\dot{\breve{\mu}}(1) \cong 0.0032,\ \dot{\breve{\mu}}(2) \cong 0.0036,\ \dot{\breve{\mu}}(3) \cong 0.0104.$$

Moreover, according to (24)–(25), the corresponding optimal unconditional multistate reliability function of the system is of the form

$$\dot{\boldsymbol{R}}(t,\cdot) = [1,\ \dot{\boldsymbol{R}}(t,1),\ \dot{\boldsymbol{R}}(t,2),\ \dot{\boldsymbol{R}}(t,3)] \text{ for } t \ge 0,$$

where according to (3) and after considering the values of $\dot{p}_b$, its co-ordinates are as follows:

$$\dot{\boldsymbol{R}}(t,u) \cong 0.49 \cdot [\boldsymbol{R}(t,u)]^{(1)} + 0.12 \cdot [\boldsymbol{R}(t,u)]^{(2)} + 0.39 \cdot [\boldsymbol{R}(t,u)]^{(3)}$$
$$\text{for } t \ge 0,\ u = 1,2,3, \qquad (60)$$

where $[\boldsymbol{R}(t,u)]^{(1)}$, $[\boldsymbol{R}(t,u)]^{(2)}$, $[\boldsymbol{R}(t,u)]^{(3)}$ are respectively given by (33)–(35) and (37)–(39), (41)–(43).

If the critical reliability state is $r = 2$, then the system risk function, according to (27), is given by

$$\dot{\mathrm{r}}(t) = 1 - \dot{R}(t,2) = 1 - [0.49\exp[-93.472t] + 0.12\exp[-49.663t] +$$
$$+ 0.39(\exp[-70.974t] - 3\exp[-68.018t] + 3\exp[-65.062t]) \qquad (61)$$

where $\dot{R}(t,2)$ is given by (60) for $u = 2$.

Hence, considering (28), the moment when the optimal system risk function (Fig. 2) exceeds a permitted level, for instance δ=0.05, is

$$\tau = \dot{\mathrm{r}}^{-1}(\delta) \cong 0.000676 \text{ years}. \qquad (62)$$

Comparing the bulk cargo transportation system reliability characteristics after its operation process optimization given by (58)–(62) with the corresponding characteristics before this optimization determined by (45)–(51) justifies this action.

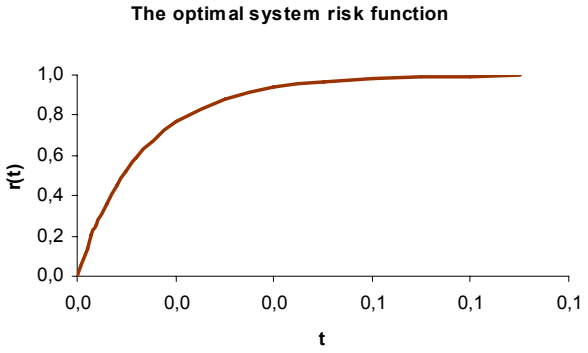


Figure 2 – The graph of the port bulk cargo transportation system optimal risk function

7 OPTIMAL SOJOURN TIMES OF BULK CARGO TRANSPORTATION SYSTEM OPERATION PROCESS AT OPERATION STATES

Having the values of the optimal transient probabilities determined by (57), it is possible to find the optimal conditional and unconditional mean values of the sojourn times of the bulk cargo transportation operation process at the operation states and the optimal mean values of the total unconditional sojourn times of the bulk cargo transportation system operation process at the operation states during the fixed operation time as well.

Substituting the optimal transient probabilities at operation states determined in (57) and the steady probabilities

$$\pi_1 = 0.315,\ \pi_2 = 0.5,\ \pi_3 = 0.185,$$

we get the following system of equations

$$\begin{cases} -0{,}16065\dot{M}_1 + 0.245\dot{M}_2 + .0.09065\dot{M}_3 = 0 \\ 0.0378\dot{M}_1 + (-0.44)\dot{M}_2 + 0.0222\dot{M}_3 = 0. \\ 0.12285\dot{M}_1 + 0.195\dot{M}_2 + (-0.11285)\dot{M}_3 = 0 \end{cases} \qquad (63)$$

with the unknown optimal mean values $\dot{M}_b$ of the system unconditional sojourn times in the operation states. Consequently, we get

$$\dot{M}_1,\ \dot{M}_2 = 0.154286\dot{M}_1,\ \dot{M}_3 = 1.216216\dot{M}_1.$$

Thus, we may fix $\dot{M}_1$ and determine the remaining ones. In our case, after considering expert opinion, we conclude that it is sensible to assume

$$\dot{M}_1 \cong 2.$$

This way the obtained solutions of the system of equations, are

$$\dot{M}_1 \cong 2,\ \dot{M}_2 = 0.308571,\ \dot{M}_3 = 2.432432. \qquad (64)$$

It can be seen that these solutions differ substantially from the values $M_1,\ M_2,\ M_3$.

Other very useful and much easier to apply in practice tools that can help in planning the operation process of the technical system are the system operation process optimal mean values of the total sojourn times at the particular operation states during the fixed system operation time θ. Assuming the system operation time $\theta = 1$ year = 365 days, after aplying (32), we get their values

$$\dot{\hat{M}}_1 = \dot{E}[\hat{\theta}_1] = \dot{p}_1\theta = 0.49 \cdot 365 \cong 179\,\text{days},$$

$$\dot{\hat{M}}_2 = \dot{E}[\hat{\theta}_2] = \dot{p}_2\theta = 0.12 \cdot 365 = 44\,\text{days},$$

$$\dot{\hat{M}}_3 = \dot{E}[\hat{\theta}_3] = \dot{p}_3\theta = 0.39 \cdot 365 = 142\ \text{days}. \qquad (65)$$

In practice, the knowledge of the optimal values of $\dot{M}_b$ and $\dot{\hat{M}}_b$ given respectively by (64)–(65), is impornat and very helpful in planing and improving the operation process, as it allows for more reliable and safer system operation. From the performed analysis of the results of the bulk cargo transportation system operation process optimization it can be suggested to change the operation process characteristics that result in replacing (or the approaching/convergence to) the unconditional mean sojourn times M_b in the particular operation states before the optimization by their optimal values $\dot{M}_b$ after the optimization. The easiest way of reorganizing the system operation process leads to replacing (or the approaching/convergence to) the total sojourn times, $\hat{M}_b = E[\hat{\theta}_b]$, of the bulk cargo transportation system operation process, and in particular operation states during the operation time $\theta = 1$ year, with their optimal values $\dot{\hat{M}}_b = \dot{E}[\hat{\theta}_b]$.

CONCLUSIONS

The joint model of reliability of complex technical systems at variable operation conditions linking a semi-markov modelling of the system operation processes with a multi-state approach to system reliability analysis was constructed. Next, the final results obtained from this joint model and linear programming were used to build the model of complex technical systems reliability optimization. These tools can be used in reliability evaluation and optimization of a very wide class of real technical systems operating at varying conditions that influence their reliability structures and the reliability parameters of their components. The practical application of these tools to reliability and risk evaluation and optimization of a technical system of a bulk cargo transportation system, operating in variable conditions, and the results achieved can be implemented by reliability practitioners from both maritime transport industry and other industrial sectors.

REFERENCES

1. Aven T. Reliability evaluation of multi-state systems with multi-state components / T. Aven // IEEE Transactions on reliability. – 1985. – Vol. 34. – P. 473–479.
2. Grabski F. Semi-Markov Models of Systems Reliability and Operations / F. Grabski. – Warsaw : Systems Research Institute, Polish Academy of Science. – 2015.
3. Klabjan D. Existence of optimal policies for semi-Markov decision processes using duality for infinite linear programming / D. Klabjan, D. Adelman // Siam. J. Control Optim. – 2006. – Vol. 44(6). – P. 2104–2122. DOI: 10.1137/s0363012903437290
4. Kołowrocki K. Reliability of large and complex systems / K. Kołowrocki. – Elsevier, 2014. – 460 p. DOI: 10.1016/b978-0-08-099949-4.00010-6
5. Kołowrocki K. Modelling of operational processes of port bulk cargo system / K. Kołowrocki, B. Kwiatuszewska-Sarnecka, J. Soszyńska // Proc. 2nd Summer Safety and Reliability Seminars – SSARS 2008. – Gdacsk-Sopo. – 2008. – Vol. 2. – P. 217–222.
6. Kołowrocki K. Reliability and risk analysis of large systems with ageing components / K. Kołowrocki, B. Kwiatuszewska-Sarnecka // Reliability Engineering & System Safety – 2008. – Vol. 93. – P. 1821–182. DOI: 10.1016/j.ress.2008.03.008
7. Kołowrocki K. Reliability and availability of complex systems / K. Kołowrocki, J. Soszyńska // Quality and Reliability Engineering International. – 2006. – Vol. 22, Issue 1. – P. 79–99. DOI: 10.1002/qre.749
8. Kołowrocki K. Reliability and Safety of Complex Technical Systems and Processes: Modeling-Identification-Prediction-Optimization / K. Kołowrocki, J. Soszycska-Budny. – London : Springer, 2011. – 405 p. DOI: 10.1007/978-0-85729-694-8
9. Kołowrocki K. Reliability and risk improvement with components quantitative and qualitative redundancy – bulk cargo terminal / K. Koiowrocki, B. Kwiatuszewska-Sarnecka, J. Soszyńska // Proc. 9nd Summer Safety and Reliability Seminars (SSARS 2015). – Gdacsk-Sopot, 2015.
10. Kwiatuszewska-Sarnecka B. Analysis of Reservation Efficiency in Series Systems : thesis ... doctor of philosophy / Kwiatuszewska-Sarnecka Bozena. – Gdynia Maritime University, 2003.
11. Kwiatuszewska-Sarnecka B. Reliability Improvement of Large Multi-state Series-parallel Systems / B. Kwiatuszewska-Sarnecka // International Journal of Automation and Computing. – 2006. – Vol. 2. – P. 157–164. DOI: 10.1007/s11633-006-0157-y
12. Lisnianski A. Multi-state System Reliability. Assessment, Optimisation and Applications / A. Lisnianski, G. Levitin. – Singapore : World Scientific Publishing Co., 2003. – 376 p.
13. Meng F. Component-relevancy and characterization in multi-state systems / F. Meng // IEEE Transactions on reliability. – 1993. – Vol. 42. – P. 478–483. DOI: 10.1109/24.257834

Article was submitted 01.12.2015.
After revision 15.12.2015.

Коловроцкий К.[1], Квапишевска-Сарнецка Б.[2], Сошинська-Будный Й.[3]
[1]Д-р наук, профессор, профессор кафедры математики, Морская Академия в Гдыне, Гдыня, Польша
[2]Д-р философии, доцент, доцент кафедры математики, Морская Академия в Гдыне, Гдыня, Польша
[3]Д-р философии, адъюнкт кафедры математики, Морская Академия в Гдыне, Гдыня, Польша

ОПТИМИЗАЦИЯ НАДЕЖНОСТИ И РИСКОВ СИСТЕМ С НЕСКОЛЬКИМИ УСТОЙЧИВЫМИ СОСТОЯНИЯМИ В ПРИЛОЖЕНИИ К ТРАНСПОРТНОЙ СИСТЕМЕ ПОРТА

Сложность процессов работы технических систем и их влияние на изменение во времени структур систем и параметров надежности их компонентов обусловливают сложности при первой встрече в реальности, а затем в фиксации и анализе этих структур и

параметров надежности. Путем построения объединенной модели надежности сложных технических систем в различных условиях эксплуатации, связывающей полумарковское моделирование процессов работы системы с подходом нескольких состояний в анализе надежности систем, мы находим основные характеристики надежности системы. Затем мы используем линейное программирование для того, чтобы построить модель оптимизации надежности сложных технических систем. Мы исследуем приложение модели в морском транспорте, в частности, в оптимизации надежности и рисков объемной системы грузоперевозок. Инструменты, разработанные нами, могут быть использованы для оценки надежности и оптимизации очень широкого класса реальных технических систем, работающих в различных условиях, которые влияют на их структуру надежности и параметры надежности их компонентов. Следовательно, разработанные нами инструменты могут быть использованы специалистами-практиками в области надежности как в отрасли морского транспорта, так и в других отраслях промышленности.

Ключевые слова: функция надежности, функция риска, рабочий процесс, оптимизация.

Коловроцький К.[1], Квапішевска-Сарнецка Б.[2], Сошинська-Будний Й.[3]
[1]Д-р наук, професор, професор кафедри математики, Морська Академія в Гдині, Гдиня, Польща
[2]Д-р філософії, доцент, доцент кафедри математики, Морська Академія в Гдині, Гдиня, Польща
[3]Д-р філософії, ад'юнкт кафедри математики, Морська Академія в Гдині, Гдиня, Польща

ОПТИМІЗАЦІЯ НАДІЙНОСТІ І РИЗИКІВ СИСТЕМ З КІЛЬКОМА СТІЙКИМИ СТАНАМИ У ЗАСТОСУВАННІ ДО ТРАНСПОРТНОЇ СИСТЕМИ ПОРТУ

Складність процесів роботи технічних систем та їхній вплив на зміну в часі структур систем і параметрів надійності їхніх компонентів обумовлюють складнощі при першій зустрічі у реальності, а потім у фіксації і аналізі цих структур і параметрів надійності. Шляхом побудови об'єднаної моделі надійності складних технічних систем в різних умовах експлуатації, що зв'язує напівмарковське моделювання процесів роботи системи з підходом декількох станів в аналізі надійності систем, ми знаходимо основні характеристики надійності системи. Потім ми використовуємо лінійне програмування для того, щоб побудувати модель оптимізації надійності складних технічних систем. Ми досліджуємо застосування моделі в морському транспорті, зокрема в оптимізації надійності та ризиків об'ємної системи вантажоперевезень. Інструменти, розроблені нами, можуть бути використані для оцінки надійності та оптимізації дуже широкого класу реальних технічних систем, що працюють в різних умовах, які впливають на їх структуру надійності і параметри надійності їхніх компонентів. Отже, розроблені нами інструменти можуть бути використані фахівцями-практиками в галузі надійності як у галузі морського транспорту, так і в інших галузях промисловості.

Ключові слова: функція надійності, функція ризику, робочий процес, оптимізація.

REFERENCES

1. Aven T. Reliability evaluation of multi-state systems with multi-state components, *IEEE Transactions on reliability*, 1985, Vol. 34, pp. 473–479.
2. Grabski F. Semi-Markov Models of Systems Reliability and Operations. Warsaw, Systems Research Institute, Polish Academy of Science, 2015.
3. Klabjan D., Adelman D. Existence of optimal policies for semi-Markov decision processes using duality for infinite linear programming, *Siam*. J. *Control Optim*, 2006, Vol. 44(6), pp. 2104–2122. DOI: 10.1137/s0363012903437290
4. Kołowrocki K. Reliability of large and complex systems. Elsevier, 2014, 460 p. DOI: 10.1016/b978-0-08-099949-4.00010-6
5. Kołowrocki K. Kwiatuszewska-Sarnecka B., Soszyńska J. Modelling of operational processes of port bulk cargo system, *Proc. 2nd Summer Safety and Reliability Seminars. – SSARS 2008.* Gdacsk-Sopot, 2008, Vol. 2, pp. 217–222.
6. Kołowrocki K., Kwiatuszewska-Sarnecka B. Reliability and risk analysis of large systems with ageing components, *Reliability Engineering & System Safety,* 2008, Vol. 93, pp. 1821–182. DOI: 10.1016/j.ress.2008.03.008
7. Kołowrocki K., Soszyńska J. Reliability and availability of complex systems, *Quality and Reliability Engineering International,* 2006, Vol. 22, Issue 1, pp. 79–99. DOI: 10.1002/qre.749
8. Kołowrocki K., Soszycska-Budny J. Reliability and Safety of Complex Technical Systems and Processes: Modeling-Identification-Prediction-Optimization. London, Springer, 2011, 405 p. DOI: 10.1007/978-0-85729-694-8
9. Kołowrocki K., Kwiatuszewska-Sarnecka B., Soszycska J. Reliability and risk improvement with components quantitative and qualitative redundancy – bulk cargo terminal, *Proc. 9nd Summer Safety and Reliability Seminars (SSARS 2015)*. Gdacsk-Sopot, 2015.
10. Kwiatuszewska-Sarnecka B. Analysis of Reservation Efficiency in Series Systems : thesis ... doctor of philosophy. Gdynia Maritime University, 2003.
11. Kwiatuszewska-Sarnecka B. Reliability Improvement of Large Multi-state Series-parallel Systems, *International Journal of Automation and Computing,* 2006, Vol. 2, pp. 157–164. DOI: 10.1007/s11633-006-0157-y
12. Lisnianski A., Levitin G. Multi-state System Reliability. Assessment, Optimisation and Applications. Singapore, World Scientific Publishing Co., 2003, 376 p.
13. Meng F. Component-relevancy and characterization in multi-state systems, *IEEE Transactions on reliability*, 1993, Vol. 42, pp. 478–483. DOI: 10.1109/24.257834

УДК 004.89:656.11

Лахно В. А.

Д-р техн. наук, доцент, заведующий кафедрой организации комплексной защиты информации, Европейский университет, Киев, Украина

МОДЕЛЬ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ УПРАВЛЕНИЯ ГОРОДСКИМИ АВТОБУСНЫМИ ПЕРЕВОЗКАМИ

Решена задача разработки математического обеспечения для интеллектуальной системы управления городскими автобусными перевозками. Объектом исследования является автоматизация процесса принятия решений интеллектуальными системами управления городскими транспортными потоками. Предмет исследования составляют модели диспетчерского управления транспортными потоками автобусных маршрутов. Цель работы: совершенствование моделей для интеллектуальных транспортных систем управляющих городскими потоками автобусов и маршрутных такси. Разработана модель для интеллектуальной транспортной системы управления с учетом влияния наиболее значимых факторов на график движения автобусов по маршруту. Модель позволяет оперативно оценивать влияние возмущающих действий на движение подвижной единицы, в частности, переполнение пассажирами автобусов на маршруте, их сход с линии, отклонение от расписания и др., на показатели качества обслуживания, а также, оптимизировать расписание движения. В качестве критерия оптимизации предложен показатель минимума времени ожидания пассажирами автобусов и маршрутных такси на остановках.

В ходе экспериментов проверена адекватность разработанной модели, которая оценивалась методом однофакторного дисперсионного анализа и полнофакторного эксперимента в реальных городских условиях. Результаты экспериментов позволяют рекомендовать предложенную модель для практического использования в интеллектуальных транспортных системах управления городскими автобусными маршрутами.

Ключевые слова: интеллектуальная транспортная система, управление автобусными перевозками, модель, алгоритм управления.

НОМЕНКЛАТУРА

ВВ – возмущающее воздействие;
ИТС – интеллектуальная транспортная система;
КП – контрольный пункт;
КТС – контур транспортной связи;
ОП – остановочный пункт;
ПЕ – подвижная единица;
ТС – транспортное средство;
ЭВМ – электронная вычислительная машина;

A_i – вместимость i-й подвижной единицы;

d_i^j – потребность в перевозке с j-го ОП в момент прибытия i-й ПЕ;

g_i^j – количество пассажиров, не обслуженных i-й ПЕ j-го ОП;

K_i – коэффициент наполнения ПЕ;

l^j – расстояние между j-м и $j+1$-м остановочными пунктами;

L – длина маршрута;

m – количество подвижных единиц на маршруте;

n – количество остановочных пунктов на маршруте;

N – количество возможных вариантов распределения перевозок;

$N_{КП}$ – номер конечного КП;

P_i – суммарный пассажиропоток, приходящийся на i-ю ПЕ за рейс;

P_i^j – количество пассажиров, пришедших на j-й ОП за время между прибытием i-й и $i-1$-й ПЕ;

$\widetilde{P}_i^j$ – количество вошедших на ОП пассажиров;

q_i^j – количество вышедших на ОП пассажиров;

Q_i^j – наполнение i-й ПЕ после обслуживания j-го ОП;

r_i^j – резерв мест в i-й ПЕ на j-м ОП;

S_i^j – средние затраты времени ожидания;

t_{cj} – среднее время ожидания пассажиров на j-м временном отрезке при сходе на нем ПЕ;

$t_{нj}$ – среднее время ожидания пассажиров на j-м временном отрезке при переключении ПЕ на другой маршрут;

t_i^j – время прибытия подвижной единицы на j-й остановочный пункт;

$t_i^{нчр(k)}$ – время начала k-го рейса;

$t_{inn}^{КП}$ – время прибытия ПЕ на контрольный пункт;

t_{nj} – среднее время ожидания пассажиров на j-м временном отрезке при работе «переключенной» на нем ПЕ;

t_{oj} – среднее время ожидания пассажиров на j-м временном отрезке при движении ПЕ на нем с оперативным интервалом τ_o;

t_{pj} – среднее время ожидания пассажиров при вводе резервной ПЕ на j-м временном отрезке;

t_{uj} – среднее время ожидания пассажиров на j-м временном отрезке при увеличении интервала движения ПЕ на нем на величину τ_u;

t_{yj} – среднее время ожидания пассажира на j-м временном отрезке при увеличении на нем пассажиропотока;

t_{2j} – среднее время ожидания пассажиров на j-м временном отрезке без переключения ПЕ на другой маршрут;

T – среднее время рейса;

T_i – время рейса i-й ПЕ;

$T_i^{дв}$ – время, затрачиваемое на проезд по перегонам между ОП;

DOI 10.15588/1607-3274-2016-2-15

$T_i^{ост}$ – суммарное время стоянок на ОП;

W_{in} – закономерности распределения перевозок по часам суток, заданные в виде весовых оценок;

Z – суммарные затраты времени для всех пассажиров на ожидание транспорта;

γ – допустимая величина уменьшения времени t_0 отстоя на конечном КП;

γ_{kj} – допустимая величина уменьшения планового времени τ_{kj} проезда на участке $k-j$ маршрута;

γ_{kN} – допустимая величина уменьшения планового времени проезда на участке $k-N$ маршрута (времени τ_{kN});

δ_k – опоздание ПЕ на ОП;

ΔJ – общее снижение времени ожидания пассажиров;

$\Delta\tau$ – интервал между ПЕ;

ε – допустимая величина увеличения времени t_0 отстоя на конечном КП;

ε_{kj} – допустимая величина увеличения времени τ_{kj} проезда на участке $k-j$;

ε_{kN} – допустимое увеличение планового времени проезда на участке $k-N$;

η_i – случайная величина, характеризующая интенсивность посадки и высадки для i-й ПЕ;

λ_j – интенсивность пассажиропотока на j-м временном отрезке;

ν_i^j – случайная величина, характеризующая количество пассажиров, выходящих из i-й ПЕ на j-м ОП;

$v(U)$ – случайная величина, характеризующаяся законом распределения скорости движения ПЕ при данном признаке управления в ИТС;

v_{cp} – средняя скорость движения ПЕ;

ρ – суммарный результат пассажиропотока маршрута;

$\rho^j(t)$ – случайная величина, характеризующая плотность пассажиропотока на j-м ОП;

τ – интервал времени между приходами n ПЕ;

τ_{ni}^j – время, затрачиваемое на посадку и высадку пассажиров на j-м ОП i-й ПЕ;

$\tau_i'^j$ – время движения между $i-1$-й и i-й ПЕ после прохождения j-го ОП;

$\tau_{i-1,i}^j$ – интервал между прибытиями $i-1$-й и i-й ПЕ;

$\tilde{\tau}_i^{j+1}$ – время, затраченное на движение по перегону маршрута между j-м и $j+1$-м ОП;

τ_j – плановый момент прибытия на j-й КП;

$\tau_j(t_c)$ – часть j-го временного отрезка, перекрываемая временем схода ПЕ;

$\tau_j(t_н)$ – часть j-го временного отрезка, перекрываемая временем переключения ПЕ (временем τ_n);

$\tau_j(t_n)$ – часть j-го временного отрезка, на котором работает «переключенная» ПЕ;

$\tau_j(t_p)$ – часть j-го временного отрезка, заключенная между моментами схода ПЕ с линии и ввода резервной ПЕ;

$\tau_j(\tau_o)$ – часть j-го временного отрезка, на которой интервал движения ПЕ осуществляется с оперативным интервалом τ_o;

$\tau_j(\tau_u)$ – часть j-го временного отрезка, на которой интервал движения ПЕ увеличен на τ_u;

$\tau_з(U)$ – случайная величина, определяющая время задержки как функцию от признака управления в ИТС.

ВВЕДЕНИЕ

Транспортный комплекс больших городов образуют множество пассажирских и грузовых средств, управляющий ими персонал и дорожно-информационно-коммуникационная инфраструктура. Современные интеллектуальные транспортные системы (ИТС) базируются на использовании наукоемких инфокоммуникационных технологий, востребованных необходимостью повышения эффективности дорожного движения [1, 2].

Объектом исследования является автоматизация процесса принятия решений интеллектуальными системами управления городскими транспортными потоками в современных мегаполисах. Следует отметить, что автоматизированные системы управления движением, базирующиеся на детерминированных математических моделях, часто оказываются неадекватными реальным процессам перевозки пассажиров [1–6].

Предмет исследования составляют модели диспетчерского управления транспортными потоками автобусных маршрутов.

Цель работы – совершенствование математических моделей для интеллектуальных транспортных систем управляющих городскими потоками автобусов и маршрутных такси.

1 ПОСТАНОВКА ЗАДАЧИ

В моделировании транспортных потоков существуют две противоположные тенденции. С одной стороны, стремятся, как можно точнее описать исследуемый объект, чтобы добиться полной адекватности объекта и моделью. А с другой, – упростить модель, привести ее к виду, удобному для нахождения решения одним из известных методов [3, 4, 6]. Однако, в большинстве задач управления городскими транспортными потоками приходится иметь дело со случайными величинами, что значительно усложняет математическую обработку моделей [7–9].

В этих случаях используется стохастическое моделирование. Его сущность состоит в том, что процессу функционирования объекта ставится в соответствие процесс вычисления функционала $\bar{y} = F(x)$, задаваемого математической моделью объекта.

Элементарным звеном функционирования системы является технологический процесс работы отдельной подвижной единицы (ПЕ), например, городского автобуса или маршрутного такси. Он состоит из выхода ПЕ на маршрут, работы на маршруте и заезда на стоянку. Работа на маршруте разбивается на выполнение рейсов и отстои [10].

Основная задача ИТС управления движением ПЕ – это удовлетворение потребности пассажиров в поездках. Она решается за счет выполнения расписания, а также подключения в случае необходимости резервных транспортных средств. Критерием качества удовлетворения потребности в поездках является минимизация времени ожидания пассажирами автобуса или маршрутного такси.

2 ЛИТЕРАТУРНЫЙ ОБЗОР

В [6, 7] говорится, что включение ИТС в систему управления городскими перевозками требует от разработчика: 1) наполнить ее программами, связанными с технологией управления; 2) согласовать ее динамические характеристики с динамикой работы реальных объектов; 3) согласовать сигналы (например, от сенсоров движения, системы GPS и т.п.), поступающие от объекта к ИТС и выдаваемые ИТС на объект.

Если же алгоритм управления транспортным потоком не носит жесткого характера, если в процессе принятия решения о выдаче сигналов на объект управления решаются различные оптимизационные задачи, если, наконец, на первом этапе экспериментальной проверки ИТС происходит накопление новой информации, ранее неизвестной заказчику или разработчику, то применение ЭВМ становится оправданным [4, 8, 9].

С учетом вышесказанного, целью исследования является разработка и совершенствование моделей для ИТС управляющих городскими транспортными потоками.

3 МАТЕРИАЛЫ И МЕТОДЫ

Все нарушения движения транспорта – отклонения от расписания, переполнения, являющиеся следствием неравбномерной скорости движения отдельных ПЕ и флуктуаций пассажиропотока, – вызывают увеличение времени ожидания. Поэтому и эффективность управляющих воздействий имеет смысл оценивать по тому же критерию.

Управляющие воздействия данного уровня можно разделить на две категории. Первые из них вызваны нарушениями расписания отдельными ПЕ и имеют индивидуальный характер: увеличение скорости (нагон в пути), сокращение отстоя и т.п. Во вторую категорию входят воздействия, вызванные как индивидуальными, так и групповыми нарушениями, сходами, ухудшениями условий движения, выполняемые группой ПЕ или резервными ПЕ: раздвижка интервала, вход резерва, переход на оперативный интервал и т.д. Функцию управления на этом уровне может выполнять ИТС. Для этого необходим объем информации о состоянии транспортной городской системы. При имеющемся контуре транспортной связи (КТС) информация в ИТС поступает с каждой ПЕ в контрольных точках маршрута [10].

Модель функционирования маршрута для разрабатываемой ИТС описывается следующей системой уравнений. Система моделирует временные характеристики движения подвижной единицы (ПЕ) по маршруту и процесс пассажирообразования:

$$
\left.
\begin{aligned}
& t_i^1 = t_i^{нчр(k)}; \\
& Q_i^0 = 0; \\
& q_i^1 = 0; \\
& \nu_i^j = \frac{i}{n}\eta_i \rho^j(t); \\
& q_i^j = \nu_i^j Q_i^{j-1}; \\
& r_i^j = A_i - (Q_i^{j-1} - q_i^j); \\
& P_i^j = \int\limits_{t_{i-1}^j}^{t_i^j} \rho^j(t)dt; \\
& d_i^j = P_i^j + g_{i-1}^j; \\
& Q_i^j = \min\left\{A_i, A_i + d_i^j - r_i^j\right\}; \\
& g_i^j = \max\left\{0, d_i^j - r_i^j\right\}; \\
& \widetilde{P}_i^j = d_i^j - g_i^j; \\
& q_i^n = Q_i^{n-1}; \\
& P_i^n = 0; \\
& \tau_{нi}^j = \eta_i(\widetilde{P}_i^j + q_i^j); \\
& \tau_{зi}^j = \tau_з(U); \\
& \tau_i^j = \begin{cases} \tau_{нi}^j, \text{ если } \tau_{зi}^j \le \tau_{нi}^j, \\ \tau_{зi}^j, \text{ если } \tau_{зi}^j > \tau_{нi}^j; \end{cases} \\
& v_i^j = v(U); \\
& \widetilde{\tau}_i^{j+1} = l^j / v_i^j; \\
& t_i^{j+1} = t_i^j + \widetilde{\tau}_i^{j+1} + \tau_i^j; \\
& \tau'^j_i = t_i^j - t_{i-1}^j.
\end{aligned}
\right\} \quad i = \overline{1,m};\ j = \overline{1,n}, \quad (1)
$$

Исходными данными для моделирования являются величины m; n; l^j; $t_i^{нчр(k)}$; A_i; $v(U)$; $\tau_з(U)$; η_i; $\rho^j(t)$. Остальные величины вычисляются рекурсивно в процессе функционирования модели. Случайные величины η_i и $\rho^j(t)$ определяются в результате проведения обследования пассажиропотоков.

Характеристики плотности пассажиропотока формируются на основании суточных объемов перевозок на маршруте A^j и известных общих закономерностей распределения перевозок по часам суток, заданных в виде весовых оценок W_{in}, таких, что $\sum_i W_{in} = 1$, $i = \overline{1,T}$, $n = \overline{1,N}$. Для вычисления величины W_{in} можно использовать алгоритм, описанный в [4, 7].

Моделирование системы сбора информации заключается в присвоении некоторым ОП признаков контрольного пункта (КП). При переходе модели в состояние, в котором i-я ПЕ находится на КП, вычисляются

величины: $t_{inn}^{KП}$; $K_i = 25\%$; (здесь $[\cdot]$ – «целая часть»), являющийся аналогом информации, получаемой ИТС при данном КТС, который вместе с номерами ПЕ и КП поступают на вход модели ДУ движением ПЕ.

Цель ИТС – обеспечить выполнение расписания движения с минимальными отклонениями от запланированного с помощью использования соответствующих диспетчерских воздействий. Поэтому алгоритмы ИТС ориентированы на выбор диспетчерского воздействия (ДВ), компенсирующих возмущающие воздействия (ВВ). Выбор оптимальных алгоритмов управления движением является важнейшей задачей построения ИТС.

Входной информацией для модели ИТС служат плановое расписание движения, а также информация, поступающая из модели маршрута: момент прохода подвижной единицы КП, коэффициенты наполнения ПЕ, сообщения о сходах ПЕ с линии. В [9, 10] установлено взаимное соответствие между видами нарушений движения и управляющими воздействиями, которые могут быть направлены на восстановление движения.

Как видно из данных [2, 4, 6, 8, 9], одни и те же нарушения могут устраняться различными управляющими воздействиями. Важнейшей задачей моделирования является количественный анализ управления движением. Модель позволяет оценивать качество управления по выбранному критерию.

Основным критерием качества удовлетворения потребности в поездках принимаем время ожидания пассажирами транспорта. Структура модели позволяет непосредственно оценивать эту величину. Заметим, что в реальных условиях возможна только приблизительная оценка времени ожидания пассажирами транспорта.

Будем считать, что поток пассажиров на остановке является пуассоновским с параметром ρ. Тогда за время τ на остановку в среднем приходят $\rho\tau$ пассажиров. Время ожидания пассажиром автобуса (ПЕ) в среднем равно половине интервала ожидания. Тогда суммарные затраты времени для всех пассажиров на ожидание транспорта, собравшихся за время τ, $Z = \rho\tau^2/2$.

Время τ сбора пассажиров принимается равным интервалу времени между прибытиями на остановку двух соседних ПЕ. Если пассажир не был обслужен подошедшей ПЕ, то время ожидания им прихода следующей равно уже не половине, а интервалу между ПЕ. Поэтому суммарные затраты времени на j-м ОП за интервал $\tau_{i-1,i}^j$ между прибытиями $i-1$-й и i-й ПЕ

$$Z_i^j = P_i^j \frac{\tau_{i-1,i}^j}{2} + g_{i-1}^j \tau_{i-1,i}^j. \qquad (2)$$

Средние затраты времени ожидания S_i^j вычисляются по формуле

$$S_i^j = Z_i^j / P_i^j. \qquad (3)$$

На основании выражения (2) можно определить затраты для каждой ПЕ за рейс

$$Z_i = \sum_{j=1}^{n} Z_i^j, \qquad (4)$$

$$S_i = \sum_{j=1}^{n} S_i^j \qquad (5)$$

и для каждого ОП за интервал времени между приходами n ПЕ:

$$\tau = t_{K+n}^j - t_K^j; \qquad (6)$$

$$Z^j = \sum_{i=K}^{K+n} Z_i^j, \qquad (7)$$

$$S^j = \sum_{i=K}^{K+n} S_i^j. \qquad (8)$$

Величины Z_i^j могут вычисляться для всех i, j в процессе функционирования модели по формуле (2).

Рассмотрим влияние возмущающих воздействий на величину параметров Z и S.

Отклонение ПЕ от расписания. Допустим, что на достаточно коротком временном интервале плотность пассажиропотока постоянна, $\rho^j(t) = \rho^j$. Тогда

$$P_i^j = \int_{t_{i-1}^j}^{t_i^j} \rho^j dt = \rho^j (t_i^j - t_{i-1}^j) = \rho^j \tau_{i-1,i}^j;$$

$$P_i = \sum_{j=1}^{n} P_i^j = \sum_{j=1}^{n} \rho^j \tau_{i-1,i}^j. \qquad (9)$$

Пусть на маршруте m ПЕ, а среднее время рейса T. Для ПЕ задано расписание с равными интервалами движения. При этом $\Delta\tau = T/m$ – интервал между ПЕ. Будем считать, что расписание выполняется идеально, $\tau_{i-1,i}^j = \Delta\tau$ для всех $i = \overline{1,m}$. Тогда

$$P_i = \sum_{j=1}^{n} \rho^j \Delta\tau = \Delta\tau \sum_{j=1}^{n} \rho^j = \rho\Delta\tau, \qquad (10)$$

где $\rho = \sum_{j=1}^{n} \rho^j$.

Из выражения (10) видно, что все P_i равны между собой и при идеальном выполнении расписания пассажиропоток разделен между ПЕ. Будем считать, что ПЕ достаточно для обслуживания маршрута и при равномерном распределении пассажиропотока переполнений не возникало. Тогда

$$Z_i^j = P_i^j \frac{\tau_{i-1,i}^j}{2} = \rho^j \frac{(\Delta\tau)^2}{2}; \qquad (11)$$

$$Z_i = \sum_{j=1}^{n} \rho^j \frac{(\Delta\tau)^2}{2} = \rho \frac{(\Delta\tau)^2}{2}; \qquad (12)$$

$$Z^j = \sum_{i=1}^{m} \rho^j \frac{(\Delta\tau)^2}{2} = m\rho^j \frac{(\Delta\tau)^2}{2}; \quad (13)$$

$$Z = m\rho \frac{(\Delta\tau)^2}{2}, \quad (14)$$

где $Z = \sum_{j=1}^{n} Z^j$.

Пусть i-я ПЕ отклоняется от расписания на величину δ_i. Тогда

$$\tau_{i-1,i} = \Delta\tau - \delta_{i-1} + \delta_i;\ P_i = \rho\Delta\tau + \rho(\delta_i - \delta_{i-1}); \quad (15)$$

$$Z_i = \rho\frac{(\Delta\tau + \delta_i - \delta_{i-1})^2}{2} = \rho\frac{(\Delta\tau)^2}{2} + \rho\Delta\tau(\delta_i - \delta_{i-1}) + \rho\frac{(\delta_i - \delta_{i-1})^2}{2}; \quad (16)$$

$$Z^j = \sum_{i=1}^{m} \rho^j \frac{(\Delta\tau + \delta_i - \delta_{i-1})^2}{2};$$

$$Z = \sum_{j=1}^{n} Z^j = m\rho\frac{(\Delta\tau)^2}{2} + \rho\sum_{i=1}^{m}\frac{(\delta_i - \delta_{i-1})^2}{2}. \quad (17)$$

Если мы обозначим P_i и Z в формулах (10) и (14) через P_i' и Z', а в формулах (15) и (17) через P_i'' и Z'', то получим

$$\Delta P_i = \rho\Delta\tau + \rho(\delta_i - \delta_{i-1}) - \rho\Delta\tau = P_i'' - P_i' = \rho(\delta_i - \delta_{i-1}), \quad (18)$$

при нарушении регулярности движения появляется неравномерность в распределении пассажиропотока между ПЕ даже при постоянной плотности, что может привести к переполнениям

$$\Delta Z = Z'' - Z' = \rho\sum_{i=1}^{m}\frac{(\delta_i - \delta_{i-1})^2}{2} \geq 0. \quad (19)$$

Таким образом, время ожидания пассажиров при нарушении расписания возрастет. Z будет минимальным, если $\delta_i = \delta_{i-1}$ для всех i из выражения (19) или $\delta_i = 0$.

Средние затраты времени ожидания ПЕ пассажирами

$$\Delta S = \frac{\Delta Z}{P} = \frac{\rho\sum_{i=1}^{m}\frac{(\delta_i - \delta_{i-1})^2}{2}}{m\rho\Delta\tau} = \sum_{i=1}^{m}\frac{(\delta_i - \delta_{i-1})^2}{2m\Delta\tau} =$$

$$= \sum_{i=1}^{m}\frac{(\delta_i - \delta_{i-1})^2}{2T} \geq 0. \quad (20)$$

На основании анализа выражений (19) и (20) можно сделать вывод, что для повышения качества обслуживания пассажиров на маршрутах движения необходим равномерный интервал между ПЕ.

Предположим, что ПЕ прибыла на k-й КП в момент времени t_k с опозданием на время δ_k. Если величина времени опоздания удовлетворяет условию $\delta_k = \gamma_{kj}\tau_{kj}$, то можно формировать управляющее воздействие «нагон на участке $k - j$ маршрута» ($j = k+1, k+2, ..., N_{KII} - 1$). При этом необходимо осуществлять корректировку плановых моментов проследования КП:

$$\tau_j^* = \begin{cases} \tau_j + \delta_k - \gamma_{kj}\tau_{kj}, & \text{если } (\delta_k - \gamma_{kj}\tau_{kj}) > 0; \\ \tau_j & \text{в противном случае.} \end{cases}$$

Если величина времени опоздания такова, что за счет увеличения скорости движения ПЕ обеспечить прибытие ее на конечный КП в плановый момент времени не представляется возможным, можно осуществить сокращение времени отстоя ПЕ на конечном КП. При этом необходимым условием формирования данного управляющего воздействия является $\delta_k \leq \gamma_{kN}\varphi_{kN} + \gamma t_0$.

Если величина опоздания такова, что за счет увеличения скорости движения и уменьшения времени отстоя ТЕ на конечном КП обеспечить отправление в очередной рейсов плановый момент времени не представляется возможным, допустимо формировать управляющие воздействия типа: «увеличение времени отстоя на КП» (так, чтобы ПЕ была отправлена в очередной рейс в плановый момент); «укороченный рейс» (такой, чтобы ПЕ закончила его к плановому моменту начала очередного рейса); «удлиненный рейс» (такой, чтобы ПЕ, пропустив один обычный рейс, вернулась из удлиненного к плановому моменту начала следующего рейса).

Предположим, что ПЕ прибыла на k-й КП в момент времени t_k с опережением планового момента прибытия τ_k на величину δ_k. Если величина $\delta_k \leq \varepsilon_{kj}\tau_{kj}$, то следует формировать управляющее воздействие «увеличение времени проезда». При этом необходимо осуществлять корректировку плановых моментов проследования контрольных пунктов:

$$\tau_j^* = \begin{cases} \tau_j - (\delta_k - \gamma_{kj}\tau_{kj}), & \text{если } (\delta_k - \gamma_{kj}\tau_{kj}) > 0; \\ \tau_j & \text{в противном случае.} \end{cases}$$

Если величина времени опережения такова, что за счет уменьшения скорости движения ПЕ обеспечить прибытие ее на конечный КП в плановый момент времени невозможно, то требуется увеличить время отстоя ПЕ на конечном КП. При этом необходимым условием формирования данного управляющего воздействия является

$$\delta_k \leq \varepsilon_{kN}\varphi_{kN} + \varepsilon t_0.$$

Сход ПЕ с линии. Пусть все предложения предыдущего пункта остаются в силе. Из формулы (10) следует, что для всех i

$$P_i = \rho\Delta\tau = \frac{\rho\tau}{m}. \quad (21)$$

Предположим, что i-я ПЕ сошла с линии. Тогда

$$P_{i+1}' = P_i + P_{i+1} = 2P_i, \quad (22)$$

пассажиропоток, приходящейся на следующую за сошедшей ПЕ, увеличивается в два раза.

Потери времени на ожидание двух ПЕ

$$Z' = Z_i + Z_{i+1} = \rho\frac{(\Delta\tau)^2}{2} + \rho\frac{(\Delta\tau)}{2} = \rho\Delta\tau^2. \quad (23)$$

Потери времени после схода ПЕ

$$Z'' = \rho \frac{2(\Delta\tau)^2}{2} = 2\rho\Delta\tau^2; \quad (24)$$

$$\Delta Z = Z'' - Z' = \rho\Delta\tau^2 > 0. \quad (25)$$

Как видим, и в этом случае ВВ влияют на увеличение критерия Z. Учитывая условия, приведенные в выражении (22), можно сказать, что возникает возможность переполнения на маршруте. В данном случае ИТС формиру следующие управляющие воздействия: «уменьшение времени рейса», «раздвижка интервала», «переход на равномерный интервал», «переключение с маршрута на маршрут» и «ввод резервной ПЕ». Вначале оценим эффективность применения ИТС управляющих воздействий «ввод резервной ПЕ», «раздвижка интервала» и «переход на равномерный интервал». Предположим, что на маршруте выделено пять временных отрезков относительно постоянной интенсивности пассажиропотока (два «пиковых» и три «межпиковых»): $t_0 - t_1, t_1 - t_2, ..., t_4 - t_5$. В момент времени t_c зафиксирован сход ПЕ с линии. Работоспособное состояние ПЕ будет восстановлено через время $\tau_в$, в момент времени $t_в$. Через время τ_p в момент времени t_p может быть осуществлен ввод резервной ПЕ.

Суммарное время ожидания пассажиров на маршруте за время $\tau_в$ при сходе ПЕ с линии составит

$$J_c = \sum_{j=1}^{5} \lambda_j \tau_j((t_c)t_{cj}).$$

Суммарное время ожидания пассажиров на маршруте за время $\tau_в$ при вводе резерва составит $J_p = \sum_{j=1}^{5} \lambda_j \tau_j(t_p)t_{pj} + \sum_{j=1}^{5} \lambda_j \tau_j(t_p)t_{cj}$. Тогда снижение ΔJ_p суммарного времени ожидания пассажиров за счет ввода резерва при сходе ПЕ с линии определиться в виде $\Delta J_p = J_c - J_p$.

Учитывая, что $\tau_j(t_c) = \tau_j(\tau_p + \tau_j(t_p))$, получим следующее выражение для ΔJ_p:

$$\Delta J_p = \sum_{j=1}^{5} \lambda_j \tau_j(t_p)t_{cj} - \sum_{j=1}^{5} \lambda_j \tau_j(t_p)t_{pj} = \sum_{j=1}^{5} \lambda_j \tau_j(t_p)(t_{cj} - t_{pj}).$$

Проведя аналогичный анализ, можно показать, что снижение ΔJ_u суммарного времени ожидания пассажиров за счет раздвижки интервала при сходе ПЕ с линии определяется так $\Delta J_u = \sum_{j=1}^{5} \lambda_j \tau_j(\tau_u)(t_{cj} - t_{uj})$. Аналогично можно показать, что снижение ΔJ_o суммарного времени ожидания пассажиров за счет перехода на равномерный или оперативный интервал определяется в виде

$$\Delta J_o = \sum_{j=1}^{5} \lambda_j \tau_j(\tau_o)(t_{cj} - t_{oj}).$$

Перейдем теперь к оценке эффективности применения управляющего воздействия «переключение ПЕ с маршрута на маршрут». Предположим, что анализируется возможность переключения ПЕ с одного маршрута, интенсивность пассажиропотока на котором λ_1, на другой, где интенсивность пассажиропотока λ_2 в случае схода ПЕ с первого маршрута в момент времени t_c на время $\tau_в$. Начало переключения ПЕ со второго маршрута на первый осуществляется в момент времени $t_н$. При этом переключаемая ПЕ прибывает на первый маршрут в момент t_n.

Переключение ПЕ с одного маршрута на другой можно отождествлять с вводом резервной ПЕ на первый маршрут. Поэтому снижение ΔJ_n суммарного времени ожидания пассажиров на первом маршруте за счет переключения ПЕ с маршрута на маршрут определиться по аналогии с ΔJ_p: $\Delta J_n = \sum_{j=1}^{5} \lambda_j \tau_j(t_n)(t_{cj} - t_{nj})$. Переключение ПЕ со второго маршрута равносильно сходу с него ПЕ, что вызывает увеличение времени ΔJ_1 ожидания пассажиров: $\Delta J_1 = \sum_{j=1}^{5} \lambda_{2j} \tau_j(t_н)t_{нj} - \sum_{j=1}^{5} \lambda_{2j} \tau_j t_{2j}$.

Перевод ПЕ с маршрута на маршрут будет оправдан только в том случае, если снижение времени ожидания пассажиров на новом маршруте будет больше увеличения времени ожидания пассажиров на прежнем маршруте (в рассматриваемом случае при $\Delta J_n > \Delta J_1$). Общее снижение ΔJ времени ожидания пассажиров $\Delta J = \Delta J_n - \Delta J_1$.

При формировании управляющего воздействия «переключение ПЕ с маршрута на маршрут» в случае схода ПЕ с линии необходимо определить такую пару маршрутов, для которой обеспечивается максимум общего снижения времени ожидания пассажиров – $\max\Delta J$. Выбор наиболее эффективного управляющего воздействия при сходе ПЕ с линии может быть осуществлен при определении максимального снижения ΔJ_c суммарного времени ожидания пассажиров: $\Delta J_c = \max(\Delta J_p, \Delta J_u, \Delta J_o, \max \Delta J)$.

Таким образом, при каждом сходе ПЕ с линии целесообразно формировать такое управляющее воздействие, которое обеспечивает максимальное снижение времени ожидания пассажиров (по сравнению со случаем отсутствия управляющих воздействий).

Ухудшение условий движения. Рассмотрим соотношение между интервалом движения и другими параметрами движения: $T_i = T_i^{дв} + T_i^{ост}$,

где

$$T_i^{дв} = \sum_{j=1}^{n} \tau_i^{j,j+1} = \sum_{j=1}^{n} l^{j,j+1} / v_i^{j,j+1} = \frac{1}{v_{ср}} \sum_{j=1}^{n} l^{j,j+1} = L / v_{ср}, \quad (26)$$

$$T_i^{ост} = \sum_{j=1}^{n} \tau_i^j = \sum_{j=1}^{n} \tau_{нi}^j + \sum_{j=1}^{n} \tau_{зi}^j; \quad (27)$$

$$\sum_{j=1}^{n} \tau_{нi}^j = \sum_{j=1}^{n} \eta_i(\tilde{P}_i^j + q_i^j) = \eta_i(\sum_{j=1}^{n} \tilde{P}_i^j + \sum_{j=1}^{n} q_i^j) = 2\eta_i \sum_{j=1}^{n} \tilde{P}_i^j = 2\eta_i\rho\Delta\tau.$$

Следовательно, $T_i = \frac{L}{v_{cp}} + 2\eta_i \rho \Delta\tau + T_{зi} \approx T$; $\Delta\tau = \frac{T}{m}$;

$m\Delta\tau = L / v_{cp} + 2\eta_i \rho \Delta\tau + T_{зi}$.

Отсюда

$$\Delta\tau = \frac{L / v_{cp} + T_{зi}}{m - 2\eta_i \rho}. \quad (28)$$

Из отношения

$$Z = m\rho \frac{(\Delta\tau)^2}{2} \quad (29)$$

видно, что с увеличением $\Delta\tau$ растет и Z. При ухудшении условий движения v_{cp} понижается, $v'_{cp} < v_{cp}$.

Тогда

$$\delta = \Delta\tau' - \Delta\tau = \frac{L}{m - 2\rho\eta_i} \frac{\Delta v}{(v_{cp} - \Delta v)}; \quad \Delta v = v_{cp} - v'_{cp};$$

$$\Delta Z = m\rho \frac{(\Delta\tau + \delta)^2}{2} - m\rho \frac{(\Delta\tau)^2}{2} = m\rho\Delta\tau\delta + m\rho \frac{\delta^2}{2} > 0. \quad (30)$$

При ухудшении условий движения на отдельных маршрутах или маршрутной сети в целом, в качестве основного управляющего воздействия в ИТС рекомендуется управление «переход на равномерный интервал». С учетом возможной скорости движения ПЕ рассчитывается интервал, соблюдая который ПЕ отправляются с конечных КП.

Переполнения на маршруте фиксируется, если не все пассажиры могут быть обслужены прибывшей на ОП подвижной единицей, когда $g_i^j > 0$. Возникнув на одной остановке, переполнение может появиться и на других за счет уменьшения резерва мест в ПЕ. При переполнениях

$$Z_i^j = \rho^j \frac{(\Delta\tau)^2}{2} + g_i^j \Delta\tau; \quad (31)$$

$$\Delta Z_i^j = g_i^j \Delta\tau > 0; \quad (32)$$

$$\Delta Z = \sum_{g_i > 1} \Delta Z_i^j > 0. \quad (33)$$

Сумма берется по всем i, j для которых $g_i^j > 0$.

Внеплановое увеличение пассажиропотоков ведет к переполнению ПЕ, к снижению комфортабельности поездок и увеличению временных затрат пассажиров. Основными управляющими воздействиями при этом являются «переключение ПЕ с маршрута на маршрут» и «ввод резервной ПЕ». Целесообразно формировать такое управляющее воздействие, которое обеспечит максимальное снижение времени ожидания пассажиров (по сравнению со случаем отсутствия управляющих воздействий).

Предположим, на части j-го временного отрезка $\tau_j(t_y)$ в момент t_y произошло внеплановое увеличение λ_j на величину $\Delta\lambda_j$. Тогда увеличение суммарного времени ожидания пассажиров на данном маршруте $\Delta J_y = \sum_{j=1}^{5} \Delta\lambda_j \tau_j(t_n) t_{yj}$.

Если на данный маршрут переключить ПЕ, то ΔJ_y уменьшится на ΔJ_n, на величину снижения суммарного времени ожидания пассажиров за счет переключения ПЕ с маршрута на маршрут. Если ввести резервную ПЕ, то величина ΔJ_y уменьшится на ΔJ_p, на величину снижения суммарного времени ожидания пассажиров за счет ввода резервной подвижной единицы.

Выбор типа управляющего воздействия при внеплановом увеличении пассажиропотока на маршруте может быть осуществлен при определении максимального уменьшения величины ΔJ_y: $J_y = \max\{(\Delta J_y - \Delta J_n), (\Delta J_y - \Delta J_p)\}$.

Таким образом, установлено, что все возмущающие воздействия ухудшают показатели качества обслуживания пассажиров. На описанной модели – это влияние можно подвергнуть количественной оценке. Моделируя ВВ и вычисляя величины Z и S до нарушения движения и после принятия диспетчерского воздействия, можно определить скорость реакции системы на ВВ и ДВ; эффективность ДВ, а в конечном результате – непосредственный экономический эффект от внедрения ИТС. Все указанные вычисления, а также статистический анализ моделирования производит блок анализа.

4 ЭКСПЕРИМЕНТЫ

Предложенная модель реализована в MatLab. В ходе имитационного моделирования исследовались следующие параметры: η_i; $\widetilde{P}_i^j + q_i^j$; $\widetilde{\tau}_i^{j+1}$; $\Delta\tau'$. Для проведения имитационного эксперимента была составлена рабочая таблица планирования в соответствии рекомендациями [8, 10].

5 РЕЗУЛЬТАТЫ

Результаты имитационного моделирования пассажиропотока показаны в таблице 1.

Таблица 1 – Результаты имитационных экспериментов

Номер опыта	η_i	$\widetilde{P}_i^j + q_i^j$	$\widetilde{\tau}_i^{j+1}$	$\Delta\tau'$
1	0,2	1	1	0,34
2	0,8	14	1	0,17
3	0,2	1	11	2
4	0,8	14	11	1
5	0,5	7,5	6	–0,68

Экспериментальное значение критерия Фишера $K_{фе} = 8{,}0 < K_{фт} = 16$ при $P = 0{,}95$, позволило сделать вывод об адекватности разработанной модели. Этот факт дает возможность обоснованно утверждать, что стохастическая имитационная модель является базой для внедрения управленческих действий для создаваемых ИТС управления пассажирским автотранспортном.

6 ОБСУЖДЕНИЕ

Имитационное моделирование с использованием пакета MatLab позволило сделать следующие выводы:

– предложенная математическая модель является работоспособной и позволяет адекватно описывать процесс движения подвижной единицы по маршруту;

– использование пакета MatLab в реальной практике эксплуатиции ИТС не представляется возможным, поскольку его применение предполагает определенную квалификацию пользователя.

ВЫВОДЫ

Научная новизна результатов, полученных в статье, состоит в том, что впервые разработана модель для ИТС управления маршрутами городских пассажирских автобусов с учетом влияния наиболее значимых стохастических факторов. Получена система уравнений, которая моделирует функционирование маршрута.

Практическая значимость полученных результатов заключается в том, что разработано программное обеспечение для имитационного моделирования в среде MatLab, реализующее предложенную модель, на основе которой решена практическая задача выбора управляющих воздействий для ИТС городских автобусных маршрутов.

Перспективы дальнейших исследований состоят в том, чтобы програмно реализовать предложенную модель на языках высокого уровня для проектируемых ИТС крупных городов Украины, что позволит оперативно оценивать влияние возмущающих воздействий (отклонение подвижной единицы от расписания, сход с линии, ухудшение условий движения, переполнение на маршруте) на показатели качества обслуживания пассажиров и составить оптимальное расписание движения транспорта.

БЛАГОДАРНОСТИ

Работа выполнена в рамках госбюджетных научно-исследовательских тем «Информационно-аналитические технологии управления в интеллектуальных транспортных системах многокритериальными и многопродуктовыми потоками в условиях неоднородной неопределенности параметров процессов» (номер г/р 0113U000695) Днепропетровского национального университета железгодорожного транспорта им. академика В. Лазаряна и «Разработка программного обеспечения для модулей АСУ предприятий с использованием объектно-ориентированных языков программирования и технологии клиент-сервер» (номер г/р 0107U006840) Луганского национального аграрного университета.

СПИСОК ЛИТЕРАТУРЫ

1. Transportation & Logistics 2030. [Electronic resource]. – London: PWC, 2014. – Access mode: http://www.pwc.com/tl2030
2. Mikulski J. Modern Transport Telematics / J. Mikulski // 11th International Coference on Transport Systems Tlematics, TST 2011. Katowice-Ustron, 19–22 October 2011.: Abstracts. Katowice-Ustron, Poland. – 2011. – P. 8–11.
3. Continuing Evolution of Travel Time Data Information Collection and Processing / [Tarnoff, Philip John, Bullock, Darcy M, Young, Stanley E, et al.] // Transportation Research Board Annual Meeting. TRB. – 2009. – P. 23–45.
4. Tyagi V. Vehicular Traffic Density State Estimation Based on Cumulative Road Acoustics / V. Tyagi, S. Kalyanaraman, R. Krishnapuram // IEEE Transactions on Intelligent Transportation Systems. – 2012. – Vol. 6. – P. 456–468.
5. Horowitz R. Control design of an automated highway system / R. Horowitz, P. Varaiya // In Proceedings of the IEEE. – 2000. – Vol. 88. – P. 913–925.
6. Dynamic Traffic Light Sequence, Science Publications / [Khalid A. S. Al-Khateeb, Jaiz A. Y. Johari, Wajdi F. Al-Khateeb] // Journal of Computer Science. – 2008. – No. 4 (7). – P. 517–524.
7. Автоматизированные системы обработки информации и управления на автомобильном транспорте : учебник / [А. Б. Николаев, С. В. Алексахин, И. А. Кузнецов, В. Ю. Строганов]; под ред. А. Б. Николаева. – М. : ACADEMIA, 2003. – 223 с.
8. Методика имитационного моделирования работы городского транспорта / [В. Н. Галушко, В. Д. Левчук, И. В. Максимей и др.] // Электрон. моделирование. – 2006. – № 2. – С. 79–95.
9. Дубова С. В. Особенности развития пассажирского транспорта в Киеве / С. В. Дубова // Містобудування та терит. планування. – 2003. – Вип. 15. – С. 68–72.
10. Лахно В. А. Повышение эффективности систем управления автомобильным пассажирским транспортом методами стохастического моделирования : монография / В. А. Лахно, А. И. Пилипенко. – Луганск : Элтон-2, 2007. – 177 с.
11. Transport Logistics. Shared solution to common challenges [Electronic resource]. – Paris: Organisation for economic co-operation and development, 2002. – Access mode: http://www.internationaltransportforum.org/pub/pdf/02LogisticsE.pdf

Статья поступила в редакцию 28.01.2016.
После дрработки 10.02.2016.

Лахно В. А.
Д-р техн. наук, доцент, завідувач кафедри організації комплексного захисту інформації, Європейський університет, Київ, Україна

МОДЕЛЬ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ УПРАВЛІННЯ МІСЬКИМИ АВТОБУСНИМИ ПЕРЕВЕЗЕННЯМИ

Вирішене завдання розробки математичного забезпечення для інтелектуальної системи управління міськими автобусними перевезеннями. Об'єктом дослідження є автоматизація процесу прийняття рішень інтелектуальними системами управління міськими транспортними потоками. Предмет дослідження становлять моделі диспетчерського управління транспортними потоками автобусних маршрутів. Метою роботи є вдосконалення моделей для інтелектуальних транспортних систем керуючих міськими потоками автобусів і маршрутних таксі. Розроблено модель для інтелектуальної транспортної системи управління з урахуванням впливу найбільш значущих чинників на графік руху автобусів за маршрутом. Модель дозволяє оперативно оцінювати вплив збурюючих дій на рух рухомої одиниці, зокрема таких, як переповнення автобусів на маршруті пасажирами, їх схід з лінії, відхилення від розкладу та ін., на показники якості обслуговування, а також, оптимізувати розклад руху. В якості критерію оптимізації запропоновано використовувати показник мінімуму часу очікування пасажирами автобусів і маршрутних таксі на зупинках. Під час експериментів перевірена адекватність розробленої моделі, яка оцінювалася методом однофакторного дисперсійного аналізу і повнофакторного експерименту в реальних міських умовах. Результати експериментів дозволяють рекомендувати запропоновану модель для практичного використання в інтелектуальних транспортних системах управління міськими автобусними маршрутами.

Ключові слова: інтелектуальна транспортна система, управління автобусними перевезеннями, модель, алгоритм управління.

Lakhno V. A.
Dr. Sc., associate professor, Head of Complex Information Security Organization Department, European University, Kyiv, Ukraine

MODEL OF INTELLIGENT MANAGEMENT SYSTEM OF CITY BUS TRANSPORTATIONS

This paper is devoted to improving the mathematical support for the intelligential traffic management system of city buses. It provides an overview of what is the real state of the-art with respect to traffic flow theory. A new mathematical model of the buses motion has been generating in consideration of stochastic factors. The model allows the calculating immediate changes in the city buses schedules connected with speed parameters. The model and program realization make allowance for increasing the efficiency of passenger service when projecting

city passenger transports. With regard to the traffic organization, the automated control system as the element of the intelligential transport systems plays the increasingly important role as a key component of the transport system, which is able to form the right choice for customers across a network, to support safe travel. The software implementing proposed method is developed. The experiments to study the properties of the proposed model are conducted. The experimental results allow to recommend the proposed model for use in practice.
Keywords: modeling, intelligential transport systems, information systems, dispatching control, transport flow, passenger flow.

REFERENCES

1. Transportation & Logistics 2030. [Electronic resource]. London, PWC, 2014. Access mode: http://www.pwc.com/tl2030
2. Mikulski J. Modern Transport Telematics, *11 th International Coference on Transport Systems Tlematics, TST 2011. Katowice-Ustron, 19–22 October 2011*. Abstracts. Katowice-Ustron, Poland.2011. – P. 8–11.
3. Tarnoff, Philip John, Bullock, Darcy M, Young, Stanley E, et al.Continuing Evolution of Travel Time Data Information Collection and Processing, *Transportation Research Board Annual Meeting*. TRB, 2009, pp. 23–45.
4. Tyagi V., Kalyanaraman S., Krishnapuram R. Vehicular Traffic Density State Estimation Based on Cumulative Road Acoustics, *IEEE Transactions on Intelligent Transportation Systems*, 2012, Vol. 6, pp. 456–468.
5. Horowitz R., Varaiya P. Control design of an automated highway system, *In Proceedings of the IEEE*, 2000, Vol. 88, pp. 913–925.
6. Khalid A. S., Al-Khateeb, Jaiz A.Y. Johari, Wajdi F. Al-Khateeb Dynamic Traffic Light Sequence, Science Publications, *Journal of Computer Science*, 2008, No. 4 (7), pp. 517–524.
7. Nikolaev A. B., Aleksahin S. V., Kuznetsov I. A., Stroganov V. Yu. Ed. A. B. Nikolaev Automated systems for information processing and management of road transport : textbook. Moscow, Academy, 2003, 223 p.
8. Galushko V. N., Levchuk V. D., Maksimey I. V., Mogila V. S., Chechet P. L. Methods of simulation of urban transport, *Elektron. Modelirovanie*, 2006, No. 2, pp. 79–95.
9. Dubova S. V. Features of development of passenger transport in Kiev, *Mistobuduvannya ta terit. planuvannya,* 2003, Vyp. 15, pp. 68–72.
10. Lahno V. A., Pilipenko A. I. Improving the efficiency of control systems of road passenger transport methods of stochastic modeling, monografija. Lugansk, Elton-2, 2007, 177 p.
11. Transport Logistics. Shared solution to common challenges [Electronic resource]. Paris, Organisation for economic co-operation and development, 2002. Access mode: http://www.internationaltransportforum.org/pub/pdf/02LogisticsE.pdf

Наукове видання

Радіоелектроніка,

інформатика,

управління

№ 2/2016

Науковий журнал

Головний редактор – д-р фіз.-мат. наук В. В. Погосов

Заст. головного редактора – д-р техн. наук С. О. Субботін

Комп'ютерне моделювання та верстання — С. В. Зуб
Редактор англійських текстів — С. О. Субботін

Оригінал-макет підготовлено у редакційно-видавничому відділі ЗНТУ

Свідоцтво про державну реєстрацію
КВ № 6904 від 29.01.2003.

Підписано до друку 21.06.2016. Формат 60×84/8.
Папір офс. Різогр. друк. Ум. друк. арк. 14,88.
Тираж 300 прим. Зам. № 558.

69063, м. Запоріжжя, ЗНТУ, друкарня, вул. Жуковського, 64

Свідоцтво суб'єкта видавничої справи
ДК № 2394 від 27.12.2005.