UDC 001.891:65.011.56

# THE ALGORITHM TREE METHOD IN SOLVING THE TASK OF CLASSIFYING HYDROGRAPHIC DATA

**Povkhan I. F.** – Dr. Sc., Associate Professor, Associate Professor at the System Software Department, State Higher Education Institution Uzhhorod National University, Uzhhorod, Ukraine.

**Mitsa O. V.** – Dr. Sc., Associate Professor, Head of the Information Control Systems and Technologies Department, State Higher Education Institution Uzhhorod National University, Uzhhorod, Ukraine.

**Mulesa O. Y.** – Dr. Sc., Associate Professor, Associate professor at the Cybernetics and Applied Mathematics, State Higher Education Institution Uzhhorod National University, Uzhhorod, Ukraine.

**Polishchuk V. V.** – PhD, Associate Professor, Associate Professor at the System Software Department, State Higher Education Institution Uzhhorod National University, Uzhhorod, Ukraine.

## ABSTRACT

**Context.** The work is dedicated to the identification of a simple and effective mechanism by which it is possible to build algorithmic classification trees (algorithmic tree models) on the basis of fixed initial information in the form of a discrete data training sample. The constructed algorithmic classification tree will unmistakably classify (recognize) the entire training sample on which the model is built, have a minimum structure (structural complexity) and consist of components – autonomous classification and recognition algorithms as the vertices of the structure (attributes of the tree).

**Objective.** The aim of this work is to create a simple, effective and universal method of constructing classification (recognition) models based on the concept of algorithmic trees for arrays of real hydrographic data, where the obtained schemes of classification systems (classification tree structure) are characterized by a tree structure (construction) and autonomous classification algorithms (sets of generalized features) as their structural elements (construction blocks).

**Method.** The general scheme of synthesizing classification trees in a form of algorithmic trees on the basis of a procedure of approximation of an array of discrete data by a set of elementary classifiers, which for the set initial training sample builds a tree-like structure, i.e. a model of the algorithmic tree, is suggested. Moreover, the constructed scheme consists of a set of autonomous classification and recognition algorithms evaluated at each step/stage of constructing the classification tree for this initial sample. A method for constructing an algorithmic classification tree has been developed, the main idea of which is to approximate step-by-step the initial sample of an arbitrary volume and structure by a set of elementary classification algorithms. The method of algorithmic tree in the formation of the current algorithmic tree vertex, node, generalized feature provides selection of the most effective, high-quality elementary classifiers from the initial set and completion of only those paths in the tree structure where the largest number of errors (failures) occurs. The structural complexity of the algorithmic tree design is estimated based on the number of transitions, vertices and tiers of the model structure, which allows one to improve the quality of its subsequent analysis, provide an effective decomposition mechanism, and build algorithmic tree structures under fixed constraint sets. The method of the algorithmic tree synthesis allows one to build different types of tree-like recognition models with different initial sets of elementary classifiers with predetermined accuracy for a wide class of problems of the artificial intelligence theory.

**Results.** The developed method of building algorithmic tree models allows one to work with training samples of a large amount of different types of information (discrete data) and provides high speed and economy of hardware resources in the process of generating the final classification scheme, as well as to build classification trees with predetermined accuracy.

**Conclusions.** An approach to the synthesis of new recognition algorithms (schemes) based on a library (set) of already known algorithms (methods) and schemes has been developed. That is, an effective scheme for recognizing discrete objects based on step-by-step evaluation and selection of classification algorithms (generalized features) at each step of the scheme synthesis is presented. Based on the suggested concept of algorithmic classification trees, a model of the structure of the algorithm tree was built, which provided classification of flood situations for the Uzh river basin.

**KEYWORDS:** classification tree, algorithmic classification tree, discrete object, feature, recognition function, recognition algorithm, branching criterion.

## ABBREVIATIONS

TS is a training selection;
ST is a test selection;
RS is a recognition system;
IR is an image recognition;
GA is a generalized attribute;
RF is a recognition function;
SS is a software system;
LCT is a logical classification tree;
ACT is an algorithmic classification tree;
DFT is a generalized feature tree;
CMT is a classification models tree;
GAT is a generalized attribute tree;

BAS is a branched attribute selection.

## NOMENCLATURE

$G$ is a set of signals (discrete objects);

$R$ is a partitioning into classes (patterns) $H_i$ specified in the initial data set $G$;

$n$ is a total number of the problem attributes (the attribute space dimensionality;

$w_i$ is a discrete object of the initial TS;

$m$ is a total number of training pairs (known classification objects) of the initial TS;

$f_R$ is a recognition function;

$f_j$ is a generalized feature constructed at $j-$ step of generating the algorithmic classification tree model;

$a_i$ is a fixed independent classification and recognition algorithms in the scheme of the algorithmic classification tree;

$x_i$ is a discrete objects (signals) of the initial TS;

$l$ is a value of a class of discrete object $x$ belonging;

$H_i$ is a set of classes set by the initial TS;

$M$ is a total number of independent classification algorithms $a_i$ in the set;

$k$ is a total number of classes (patterns) of the set of signals $G$;

$W(f)$ is the generalized attribute $f$ information quality value;

$G_{a_1,...,a_i}$ is a subset of the initial $G$ which is approximated using the set of classification algorithms $a_i$;

$S_{a_1,...,a_n}$ is a number of occurrences in the TS of training pairs which meet the condition $x_i \in G_{a_1,...,a_n}$;

$S^l_{a_1,...,a_n}$ is a number of all training pairs of the initial TS which are correctly classified (when the basic membership requirement is fulfilled $x_s \in G_{a_1,...,a_i}, f_R(x_s) = l$;

$\delta_{a_1,...,a_i}$ is a frequency of occurrence of the members of the sequence of discrete objects from the initial TS in $G_{a_1,...,a_i}$ data set;

$\rho_{a_1,...,a_i}$ is a maximum frequency out of frequencies $\psi^j_{a_1,...,a_i}$ is a regarding all the classes of the current problem;

$F_S$ is a value which characterizes the quality of approximation of the initial TS using the set of classification algorithms $a_i$;

$K_{Stop}$ is a branching stopping criterion at the ACT model construction;

$F$ is a function of recognizing the structure of the classification tree;

$En_{tr}$ is a classification errors, rejections (of the first type) in the initial training data set;

$Et_{tr}$ is a classification errors, rejections (of the first type) in the training data set;

$T_{Uz}$ is a characterizes the total time (hardware time) which is necessary for building the current generalized feature $f_j$;

$E_{Uz}$ is a information capacity (structural complexity) of the constructed GF $f_j$ within the current step of generating the algorithmic classification tree model;

$S_{Uz}$ is a total number of discrete objects $x_i$ of the initial TS which are generalized (described) by the specified generalized feature $f_j$;

$P_{pt}(TS)$ is a power (volume) of the initial TS (or its fixed part for the current step of the scheme of constructing the algorithmic classification tree);

$Er_{All}$ is a total number of errors of the algorithmic classification tree model in the data of the initial test set as well as of the training set;

$M_{All}$ is a total power (volume) of the data of the training set as well as of the test set;

$Fr_{All}$ is a characterizes the number of vertices of the obtained algorithmic classification tree model with the resulting value $f_R$ (RF, i. e. the leaves of the classification tree);

$V_{All}$ is a represents the total number of all types of vertices in the structure of the algorithmic classification tree model;

$O_{Uz}$ is a total number of generalized features used in the classification tree model;

$P_{All}$ is a total number of transitions between the vertices in the structure of the constructed classification tree model;

$N_{All}$ is a total number of different classification algorithms that are used in the classification tree model;

$I_{Main}$ is a indicator of generalizing data of the initial TS using the classification tree;

$Q_{Main}$ is a integral indicator of the quality of the algorithmic classification tree model.

## INTRODUCTION

Tasks related to image recognition, are quite diverse and arise in the contemporary world in all spheres of economics and social content of human activity, this leads to the need to construct and study mathematical models of the corresponding systems. As of now, there is no universal approach to their solving. There have been proposed several fairly general theories and approaches that allow solving many types (classes) of tasks, but their application differs greatly by the high level of sensitivity to the specifics of the task itself or subject area of application [1]. Many theoretical results have been obtained for special cases and subtasks, and it should be noted that the weak point of successful real recognition systems is the necessity to perform a huge amount of computation and orientation on powerful hardware tools. However, a large number of application tasks in various fields of natural science, such as geology, geophysics, geochemistry, medicine, sociology, archeology, biology, etc., where classification problems are solved using software and hardware systems, determines the intensity and relevance of this study [2–5]. So, to date, four thousand recognition methods and schemes are known. All of them are characterized by a variety of approaches to the classification

problem, but have sets of strict restrictions. The limitations of classification schemes are largely related to the efficiency of the recognition scheme, the speed of classification, and the rational use of information system resources. Well-known classification schemes and methods are also characterized by high sensitivity to classes of applied problems and their specifics. It is clear that such limitations of classification methods and schemes also have a negative impact on real information systems (recognition systems) [6]. The representation of training samples and large amounts of discrete information in the form of logical and algorithmic tree structures has its own significant advantages. This allows you to provide an economical description and efficient data analysis, a simple mechanism of operation [7]. So, the coverage of the training set with a set of elementary features in the case of logical classification trees, or the coverage of the training set with a fixed set of autonomous recognition and classification algorithms in the case of algorithmic classification trees, generates a fixed tree-like data structure (a tree model) that enables compression and transformation of the initial data of the training set. This method allows significant software optimization, significant savings in the resources of the classification system [8] in problems with complex data structures. The scope of application of the concept of classification trees, various types of LCTs / ACTs structures is currently extremely voluminous. The set of problems and problems solved by this device can be represented by three basic classes: problems of describing and presenting complex structural data, a class of classification and forecasting problems, and a class of regression analysis problems [9]. Note that modern algorithms and methods based on step – by-step selection of traits of decision tree synthesis belong to the class of greedy algorithms and are known from the literature – division and capture. When applying the classical scheme, the construction of the classification tree will be carried out in stages in the direction from top to bottom [10].

Decision trees, namely the structures of algorithmic classification trees are **the object of the study.**

The central issues of classification tree methods remain the problem of selecting the basic branching criterion in the schema structure, the branching stop criterion, and the logic tree branch rejection criterion. There is a fundamental question of the theory of classification trees – the question of the possible construction of all variants of logical tree structures corresponding to the initial TS and the selection of the minimum depth and structural complexity of the classification tree [11].

The methods and algorithms for constructing algorithmic classification trees (decision trees) are **the subject of research**.

A problematic issue of existing methods for processing training samples in recognition problems when constructing classification rules is that they do not effectively allow us to adjust their complexity and accuracy in the process of model synthesis [12]. This disadvantage is largely absent in the methods and schemes for constructing recognition systems based on the concept of algo-

rithmic, logical classification trees. Note that a special feature of the algorithmic tree concept is the possibility of integrated use of many well – known recognition algorithms to solve each specific problem of constructing a recognition scheme. This allows you to organize the modular principle of constructing classifiers. The concept of algorithm trees is based on a basic scheme – optimal approximation of training sample data by a set of generalized features, classifiers that are included in a certain structure built in the learning process [13].

The **objective of the work** will be to build an effective, universal method for synthesizing recognition models based on the concept of algorithm trees for arrays of real hydrographic data in the problem of predicting flood phenomena. In this case, the constructed recognition (classification) systems are a graph – schema structure (classification tree) and contain sets of different types of recognition algorithms and sets of classifiers (generalized features).

## 1 PROBLEM STATEMENT

Suppose within some set $G$ of discrete objects $x$ there is a partition $R$ into a finite number $k$ of subsets (classes, images) $H_i, (i = 1,...,k)$, $G = \bigcup_{i=1}^{k} H_i$. The respective sets $H_1,...,H_k$ are called images and the elements of set $G$ are called images or representatives of images $H_1,...,H_k$. Objects (images) $x$ are defined by sets of values of some features $x_j, (j = 1,...,n)$. If $x \in H_i$, then we consider that this object belongs to image $H_i$. In the general case images $H_1,...,H_k$ can be specified by probabilistic partitions $p(H_1 / x),..., p(H_i / x)$, where $p(H_i / x)$ is the probability ( or in the continuous case probability density) of belonging $x, (x \in H_i)$ to image $H_i$. Initial training information in the form of HB is set by the problem condition as a fixed sequence of some training pairs of the following form:

$$(x_1, f_R(x_1)),...,(x_m, f_R(x_m)). \qquad (1)$$

To check the effectiveness of the constructed classification models, a test sample is also given – in the form of a set of objects of known class affiliation, and it can be represented as a fixed part of the initial TS data. In this interpretation of the initial problem, TS is a collection of some sets of discrete objects of known class affiliation.

Moreover, each discrete TS data object is a set of values of fixed elementary features, object attributes, and values of recognition functions (RF) on this data set. Note that the value of the recognition function (class marker) refers a fixed discrete TS data object to the corresponding class [14].

Within the framework of this study, the basic task will be to construct a model of the algorithm tree (ADC struc-

ture) with a set of fixed design parameters $p$. Moreover, the structure $L$ (ACT models) of which would be optimal $F(L(p, x_i), f_R(x_i)) \rightarrow opt$ according to the conditions of the problem (a set of initial TS data, the structure of the initial array of discrete data. Note that in this study, we will be interested in constructing ACT structures for real-world training samples of hydrographic data.

## 2 REVIEW OF THE LITERATURE

The research is a continuation of the works devoted to the main problematic issues of the concept of tree-like classification schemes for discrete objects [6–8, 11] in pattern recognition problems. The main attention in these works is paid to the problems of constructing, representing, using, and optimizing the structures of classification trees. It is known from [6] that the resulting classification scheme, which is constructed by an arbitrary method or algorithm based on the criteria of branched feature selection, has a tree-like logical structure. In other words, it can be represented in graph – schematic form. A logical classification tree consists of a set of vertices, branches of the structure, which are grouped by tiers and obtained at a certain stage of building a classification tree model [15]. An actual problem that arises from the work [16] is the question of synthesizing recognition trees, which will actually be represented by a tree of algorithms. Note that unlike existing classification schemes, the main feature of tree recognition systems is that the importance of individual features, a group of features, or algorithms is determined relative to the function that sets the division of objects into classes [17]. Thus, the work [18] deals with the main issues related to generating decision trees in the case of uninformative features. We note a characteristic feature of classification tree models (LCT/ACT structures) to provide effective one-dimensional branching, which allows analyzing attributes, individual features in the object structure, and working with generalized features of various types. In this case, sets of generalized features can be represented as complex predicates, and for structures of algorithm trees in the form of sequences by autonomous classification and recognition algorithms [8]. This representation of classification models (LCT/ACT structures) is actively used for big data mining, where there is a need to build a classification model that predicts the value of one or more target variables based on data from the original TS array [19]. To date, there are a significant number of methods and schemes that implement the concept of decision trees based on basic branching criteria, but two of their representatives (C4.5/C5.0 and CART) have received the greatest use and distribution. So the scheme of the logical classification tree C4.5/C5.0 uses the so – called information-theoretic criterion as a node selection criterion (branching criterion). The CART scheme is based on the calculation of the Gini index (criterion). This criterion is based on calculating the relative distances between the distributions of primary classes [20] based on the metric given by the condition of the problem. Schemes of all methods and algorithms of

branched feature selection, algorithms for ACT structures, can be reduced to optimal coverage (approximation) of the initial data array by a set of fixed classification (recognition) algorithms. For the LCT case, TS data is approximated by sets of estimated elementary features. A basic question arises-the question of choosing an effective branching criterion in the structure of the classification tree scheme. The question boils down to the problem of efficient selection of sets of vertices, features and their combinations of discrete objects for schemes of logical classification trees and algorithms for ADC methods. A set of these basic questions is considered in [21], which highlights the problem of informational evaluation of individual discrete features, their sets and fixed combinations based on fixed performance criteria. This allows you to introduce an effective mechanism for implementing branching in the design of the classification tree model. The structures of classification tree models (LCTs / ACTs models) constructed from the initial TS data have a relatively small structural complexity, but the classification tree structure itself is sparse and unevenly filled in depth with attributes compared to the class of regular logic trees [22]. Important issues remain the convergence of the process of constructing classification tree structures using branched feature selection methods, the issue of maximum complexity of the tree model synthesis procedure itself, and the choice of an effective criterion for stopping the logical tree synthesis process [23]. An important point is also that the concept of classification trees does not contradict the fundamental ability to use not only individual attributes, attributes of objects that combine them, and sets as attributes, vertices of the classification tree structure. Moreover, individual independent recognition algorithms estimated from TS data can be considered as branches. In this case, the output will be a new structure – the ACT model [8]. This paper will be devoted to ACT structures in practice.

## 3 MATERIALS AND METHODS

At the first stage, by analogy with the schemes of selection of elementary features (LCT structures) [7], we propose a basic scheme of algorithmic classification trees. Moreover, for ACT structures, it will consist in approximating the array of the initial TS with a set of autonomous, independent classification algorithms of various types and structures.

At the first stage, the condition of the problem is the initial TS of the general form and structure (1). It is a set of elements with class labels (RF values) of the total volume $m$. A system of different types of algorithms (classifiers) is also defined as a condition of the problem $a_1(x), a_2(x), ..., a_M(x)$. We introduce the following basic sets, which represent the breakdown of the initial HB data by the corresponding independent classification algorithms $a_i$:

$$G_{a_1,...,a_i} = \{x \in G / a_i(x) = 1\}, (i = 1,...,M). \tag{2}$$

Each of the fixed autonomous classification algorithms $a_i(x)$ of the initial set generates a value from the set $\{0,1\}$ at the output. Here $a_i(x)=1$, in the case of a successful approximation of the object $x$, $a_i(x)=0$ in the opposite case.

Note that the set system $G_{a_1,...,a_i}$ is a complete step – by-step division of the set $G$ in the process of increasing the $i$ value – that is, the classification algorithms involved. This division is implemented by a set of independent classification algorithms $a_1, a_2,...,a_M$. Note that depending on the initial set of classification algorithms $a_1, a_2,...,a_M$, some of the sets $G_{a_1,...,a_i}$ may be empty. This situation is possible if one fixed or several algorithms for approximating the current TS data fail.

Enter a value $S_{a_1,...,a_M}$ that determines the number of training pairs $(x_s, f_R(x_s))$, $(1 \le s \le m)$ that meet the condition $x_s \in G_{a_1,...,a_i}$. Additionally, we introduce a value $S^j_{a_1,...,a_M}$, $(j=0,1,...,k-1)$ that determines the number of initial pairs $(x_s, f_R(x_s))$, $(s=1,2,...,m)$ that meet the conditions $x_i \in G_{a_1,...,a_n}$ and $f_R(x_s)=j$.

Then, according to these values and by analogy with the methods of selecting sets of elementary features [23], the following values can be introduced, which should be considered as basic criteria for constructing branching for ACT structures:

$$\delta_{a_1,...,a_i} = \frac{s_{a_1,...,a_i}}{m}; \qquad \psi^j_{a_1,...,a_i} = \frac{s^j_{a_1,...,a_i}}{s_{a_1,...,a_i}};$$

$$\rho_{a_1,...,a_i} = \max_j \psi^j_{a_1,...,a_i}. \qquad (3)$$

If the condition $x_s \notin G_{a_1,...,a_i}$ for all is met $(s=1,...,m)$, then $\delta_{a_1,...,a_i}=0$ and $\psi^j_{a_1,...,a_i}=0$ when $(j=0,1,...,k-1)$.

The value $\delta_{a_1,...,a_i}$ characterizes the frequency of occurrence of sequence sets $x_1, x_2,...,x_m$ in the set $G_{a_1,...,a_i}$. The value $\psi^j_{a_1,...,a_i}$ characterizes the frequency of an object $x$ belonging to the $H_j$ class and $x \in G_{a_1,...,a_i}$.

The condition $x \in G_{a_1,...,a_i}$ is equivalent to the condition that for a sequence $a_1,...,a_i$ there is such an element $a_y$ that $a_y(x)=1$. The value $\delta_{a_1,...,a_i}$ characterizes the information efficiency of recognition for an object $x$ up to one of the classes $H_0, H_1,...,H_{k-1}$ under the condition $x \in G_{a_1,...,a_i}$.

At the next stage, the fundamental question of whether an object $x$ belongs to classes $H_0, H_1,...,H_{k-1}$ again arises – the question of forming a classification rule for the ACT scheme. Here you should assign the classification object $x$ to a class $H_j$ when performing the following basic relation:

$$\rho_{a_1,...,a_i} = \psi^j_{a_1,...,a_i}. \qquad (4)$$

Note that this ratio is a fixed classification rule, and $\{0 \le j \le k-1\}$. The larger the value $\rho_{a_1,...,a_i}$, the higher its efficiency.

According to the problem condition, the only information that represents image splitting $H_0, H_1,...,H_{k-1}$ is the initial TS. In this case, the class $H_j$ is a collection of all training pairs $(x_s, f_R(x_s))$ of TS that satisfy the condition of belonging $f_R(x_i)=j$ [24–27].

By analogy, the average efficiency of recognizing the set of images $H_0, H_1,...,H_{k-1}$ which are specified by the training dataset with the help of recognition algorithms $a_1, a_2,...,a_M$, is estimated by the following value:

$$F_S(a_1,...,a_M) = \sum_{a_1,...,a_i} \delta_{a_1,...,a_i} \cdot \rho_{a_1,...,a_i}. \qquad (5)$$

The value $F_S(a_1,...,a_M)$ is an estimate of the approximation of the initial TS using the initial set of independent classification algorithms $a_1, a_2,...,a_M$ or the approximation ability of a fixed set of algorithms.

The value $F_S(a_1,...,a_M)$ can be obtained in parallel according to the following scheme. Let the recognition function $F$ be a certain display that matches each set $a_1,...,a_i$ with a certain element of the set $\{0,1,...,k-1\}$. Moreover, each element of this set encodes a class label.

In this case $F$, the RF is a function of the form $F(a_1,...,a_i)$, the arguments of which take values from the set $a_1,...,a_i$. Then the RF $F(a_1,...,a_i)$ uniquely refers the object $x,(x \in G)$ to the class $H_j,(0 \le j \le k-1)$ for which the condition is fulfilled:

$$F(a_1,...,a_i) = l, (0 \le l \le k-1). \qquad (6)$$

Let the condition of the problem be the initial TS in the form of a sequence of training pairs $(x_j, f_R(x_j))$ and a fixed set of different types of classification algorithms $(a_1,...,a_m)$. The total number of training pairs in HB is – $M$, the dimension of the feature space is – $n$. It has to be admitted that the work of the constructed models of classification trees is checked using an array of training dataset whose capacity is equal to $T$ (whose class belonging is also known).

Admittedly, here the data of the initial TS specify some partition into classes $(H_1,...,H_k)$, and it is possible that the corresponding algorithms $a_i$ may be unrelated by a single concept of recognition, and they can implement various methods and algorithms of classification (for example, these may be conventional geometric algorithms – the principle of whose work lies in approximating the TS by the corresponding geometric objects, algorithms for calculating estimates, potential functions, etc.).

The result of each of the fixed, selected from the algorithm library, autonomous classifiers $a_i$, at the corresponding stage of ACT generation, is one or more generalized features – $f_j$. Here, GFs are classification rules that approximate a certain part of the initial training sample. For the case of known geometric recognition algorithms, the resulting generalized features will be geometric objects that cover TS in the feature space of the dimension problem – $n$.

In applied problems, there may be cases when the corresponding classification algorithm $a_i$ cannot construct a generalized feature $f_j$. This situation is possible due to the complex arrangement of classes $H_k$ in the feature space of the problem or certain conceptual and implementation limitations of the classification algorithm itself. It is also possible that the generalized features $f_j$ constructed by the classification algorithm $a_i$ do not fully approximate the initial TS, or this situation is provided for by the ACT generation algorithm scheme itself. Here it is possible that there is an initial limitation in the scheme of the tree algorithm for generating no more than one generalized feature $f_j$ at each step of constructing the ADC structure.

Objects of the initial TS that do not fall under the constructed scheme of sets of GFs $f_j$ approximation of the sample belong to classification errors of the first type – $En_{tr}$. Accordingly, for TS data, failures classified of this type are classified as errors of the first type – $Et_{tr}$.

Therefore, we assume that the ACT structure (type II) will have a general construction of the form – (Fig. 1).

Moreover, each tier of such a classification tree defines the stage of building an ACT model. That is, the stage of approximation by a fixed classification algorithm $a_i$ for a certain part of TS data. It is clear that this approach allows you to effectively adjust the complexity and accuracy of the classification tree model that is being built.
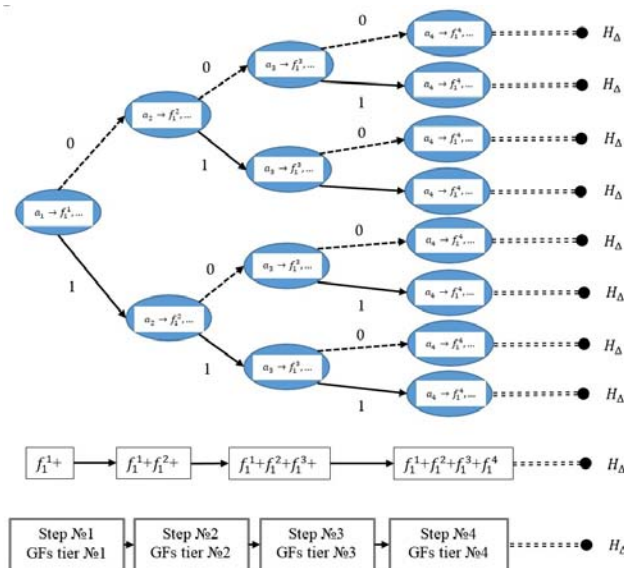


Figure 1 – General scheme of the ACT structure (type II)

For each stage of ACT model generation – (Fig. 1) submit its own classification algorithm $a_i$ and its own corresponding part of TS, a subset of the initial TS. Note that the entire initial TS is provided only at the first step, then with the next stages of building the classification tree, the power of the TS data array will fall due to the set of constructed GFs sets $f_j$. At each step of the sequence, the GFs will approximate a certain part of the original TS data. Depending on the conditions of the problem, the structure of the current ACT construction scheme, and the features of a fixed algorithm $a_i$, it is possible to generate sets $f_j$ and sequences of ACT at each step of the ACT scheme.

We introduce in the following form the basic criterion $K_{Stop}$ for stopping the branching scheme of the method for constructing the algorithm tree model, which consists in checking the volume of the initial sample $P_{pt}(TS)$ or its fixed part:

$$K_{Stop} = \begin{cases} 0, if\ P_{pt}(TS)=0; \\ 1, if\ P_{pt}(TS)>0. \end{cases} \qquad (7)$$

The general scheme for constructing the ACT model continues until the condition is met $K_{Stop}=1$. In a situation where $K_{Stop}=0$ the ACT model synthesis scheme is completed. There is a transition to the stage of testing the model based on TS data and evaluating the quality of the classification tree based on Integral indicators of the model quality.

For the ACT method, the fundamental question becomes the choice of the branching criterion, the effective choice of the current classification algorithm $a_i$ in the structure of the classification tree model under construc-

tion. Then, by analogy with the methods of approximating HB with a set of elementary features – as a basic branching criterion, we can offer a step-by-step assessment of the effectiveness of a set of algorithms $(a_1,...,a_i)$ in the following form:

$$W(a_i) = \frac{\dfrac{1}{P_{pt}(TS)} \cdot \sum\limits_{j=1}^{k}(T_{Uz} + S_{Uz} + \dfrac{E_{Uz}}{S_{Uz}})}{k}. \qquad (8)$$

In formula (8), summation is performed for all classes that are defined by the initial TS data array. Note that there may be restrictions on summation, which are due to the structure, set of initial parameters of the corresponding classification tree construction algorithm.

An important point in the methods of constructing ACT – models (Fig. 1) is that each step of the tree algorithm actually builds its own fixed one or more GFs $f_j$. The total number of GFs will increase with each step of the classification tree algorithm. In this case, the ACT scheme with a set of fixed classification algorithms $(a_1,...,a_m)$ generates a tree – like construction – a tree of generalized features with the corresponding set of GFs $(f_1, f_2,..., f_z)$.

Then, we can propose one of the possible algorithmic schemes for constructing ACT (type II).

At the first stage of building a model, the classification tree is selected based on the basic performance criterion in interactive mode or according to a random scheme. A set of autonomous classification and recognition algorithms $(a_1,...,a_m)$ for initial TS data is recorded. Classification algorithms are selected for their total number in the set (value $m$) depending on the TS structure, conditions, and constraints of the applied problem. This stage is of fundamental importance due to the impact of the final complexity of the constructed ACT model.

At the second stage of building a classification tree model, a complete regular logical tree is constructed. In this ACT structure, a selected classification algorithm from the fixed initial set is located on each of the tiers $(a_1,...,a_m)$. Notably, in this logical tree, each vertex has two transitions to the next tier (two descendants), which are denoted by the value from the binary set $\{0,1\}$. Since we are dealing with a regular logical tree, on each of the tiers of this structure there are labels (variables) of the same type (order), this applies only to the classification algorithms $a_i$, and not the GFs they generate.

Thus, at the second stage of generating the ACT (type II), an array of the training dataset is consistently provided to the classification algorithms $a_i$ (according to the structure of the constructed classification tree – (Fig. 1) in order to obtain a set of corresponding GFs, and their total number in the tree structure and the number for

each classification algorithm (a step in the scheme of the tree, a tier of the logical tree) depends on the initial parameters of the initialization of the algorithm of constructing the ACT (set up interactively or automatically) and the peculiaritites of the application task for which the ACT model is built.

After constructing a set of all GFs $f_j$ for this application task they are located at the appropriate vertices of the obtained classification tree in order to complete the procedure of its construction. The main point of this stage is that the set of constructed GFs should overlap the entire array of the training set to ensure one hundred percent recognition of the initial data. Moreover, there may be some deviations if the ACT model is built with the predetermined accuracy and complexity (this limitation of the task can be implemented by changing the number and capacity of the GFs $f_j$ being built at the second stage). It must be stressed that this condition can also be realized by limiting the steps (the number of tiers of the structure) in the procedure of building the ACT model, additional restrictions on the number of classification algorithms used in the structure of the classification tree.

At the third stage of the scheme of constructing the ACT after constructing the basic structure of a classification tree it is possible to go directly to the mode of testing the received ACT model. It has to be noted that for each test object, which is provided at the input of the classification tree, there are calculated the corresponding values $\varphi(a_j)$ (using a set of previously constructed GFs – for each vertex of the corresponding tier of the tree) that ensure (determine) the corresponding route in the structure of the constructed ACT of the second type. Thus, the GFs of each of the vertices of the ACT – in the case of a possible approximation of the object of unknown classification, ensure that the corresponding counter of the class membership is incremented and leave it unchanged in case of refusal (impossibility) of classification. At the output of the ACT structure, the object of unknown classification refers to the class whose membership counter will be maximal, and in case of their zero equity we are dealing with classification refusal.

**Remarks.** Based on the scheme of constructing the ACT of the second type, which was presented above, we can see that the number of GFs (parametric complexity, capacity) generated by the same selected classification algorithm $a_j$ on some tier of the classification tree for each of the paths of the ACT structure may be different; and following this direction, we come to the fact that the structure of the ACT model does not necessarily belong to the class of regular structures (logical trees) – that is, in each tier of the ACT structure under construction, together with a different quantity and type (general parameters) of GFs the presence of different classification and recognition algorithms $a_j$ is allowed.

## 4 EXPERIMENTS

One of the possible areas of applying the concept of algorithmic classification trees is related to the problems associated with forecasting and classifying flood events (based on arrays of meteorological and hydrological data). It is known that flood events cause tremendous harm to both the economic and ecological sectors of the national economy: they flood and prevent economic circulation of agricultural land, flood settlements, destroy residential buildings and industrial enterprises, dams, highways and railways, and human casualties occur – which is extremely important for the Transcarpathian region. That is why the study of the conditions for the formation of flood situations in the context of classifying meteorological phenomena under changing climatic conditions is a necessary stage for further substantiation of new methods for calculations and forecasts for overcoming critical ecological situations.

It should be noted that the surface water flow regime of the Uzh River basin is characterized by significant territorial unevenness due to climatic, temperature factors and the influence of the Carpathian mountains, which have a decisive impact on the formation of seasonal runoffs. To analyze the conditions and causes of floods on the river Uzh in the Transcarpathian region in the autumn – spring period there were used the data from 2 hydrological (observation) posts with a monitoring period from the beginning of 1992 to 2010 including (at the 30-km observation site). It should be highlighted that according to water regime, the river Uzh can be referred to rivers with flood regime of the Carpathian subtype – where flood situations are observed mainly during the autumn – spring period. Such flood situations are usually called cold season floods, and they are observed in the period from October / November to April / May. It is clear that the modern climate has been changing, that is why the boundaries of warm and cold periods are not constant from year to year, and the basis for determining their beginning and ending were mainly the basic meteorological conditions (air temperature, types of precipitation and snow cover) .

An important peculiarity of the Uzh river basin, in contrast to other river basins in the western region of Ukraine, is the formation of flood phenomena of mixed origin (based on snowmelt and rain runoffs, often with a predominance of rain component), which occur, by most observations, during autumn – spring; notably, this water regime is due to the climatic peculiarities of Transcarpathia, the presence of mountains, the movement of atmospheric fronts and other factors. We emphasize that in the basins of the rivers of the Transcarpathian region, rain (snow-rain) flood phenomena of different intensity and duration repeat with a frequency of up to 4-6 times per calendar year of observations. On the other hand, it should be noted that as a result of intense snowfall (the corresponding process of snow thaws) for the basin of the river Uzh only snow floods are observed, and it should be said that annual maximums are quite common only due to snowmelt, not water arteries of the Transcarpathian re-

gion. Thus, the annual maximums of floods of the cold period often exceed significantly the maximums of the warm period, and flood phenomena of the cold period are longer in time and occur more frequently, and the maximum water discharge of cold period floods is almost 2.5 times higher than the maximum water discharge of warm period floods (certainly in this scheme depending on the year there are some exceptions). The peculiarity of the Uzh river basin is that the maximum intensity of water rise is from 1.5 to 2.5 m per day, in some cases (1992) this figure may exceed 3 m per day.

The initial parameters of this application task of classifying flood phenomena of the river Uzh are presented in (Table 1).

It must be noted that the monitoring of the river basin has specific features (it may even be unique within a certain area of observation) which arise from the fact that the river is one of rather dynamic objects of nature, and requires, to a great extent, an individual approach (due to geography, formation conditions, water regime etc.). The total water flow of the river should be considered not only as a certain hydrological factor, but also as a geomorphological factor (this should be taken into account at the stage of monitoring and selecting the  observation feature space).

So the monitoring of the river bed is connected with a large space (the entire river basin) and significant time consumption, a change in the frequency of observations; along with this the tasks of the observation post include the complete characteristic feature of not only merely hydrological objects, but also the general geological and geomorphological conditions of the river basin. It should be noted that the general model of the flood phenomenon is described on the basis of 18 features (attributes), which have a different nature and are formed on the grounds of long-term observations of the Uzh river basin, and among these features the following defining characteristics (sets of parameters) of the river basin can be distinguished:

1)   The total density of the river cover within the observation sector is the ratio of the length of all surface flows of the riverbed (in kilometers), to the total amount of area of the basin (in square kilometers) in the respective sector of observation. Notably, this parameter is calculated separately for each observation post.

2)   The coefficient of curvature of the riverbed in the observation area is the ratio of the actual length of the river section (observation area) to the length of the line from the source to the mouth.

3)   The parameter of the type of annual terraces is a feature characterizing the geological peculiarities of the structure of the riverbed within the observation site and ranges as follows – {1 ... 6}.

4)   The parameter of the average width of the riverbed in the observation area for the basin the Uzh river basin is characterized by sharp differences in the values depending on the observation post where the measurements are made.

5)   The parameter of the reference water level – usually, the average water level for rivers is taken as the

reference level; the average water level is observed during the year within the period when the river is free of ice cover, and this average water level is determined as the arithmetic mean of daily level values for a certain period from flood recession to the beginning of ice phenomena). There are approximately determined the period and height of the low, constant and intermediate level, which is taken as the reference (initial) one; at the next stage, the excess of the operating level over the reference one is calculated. In case of exceeding, it is used as the value of growth for bringing all current observations to the datum level.

6) The parameter of the average depth of the river-bed in the observation area for the Uzh river is characterized by rather large differences for different observation areas.

7) The parameter of the general regime of the river is a feature that characterizes the peculiarities of water supply and flow of the riverbed in the observation area and ranges in the following way – {1 ... 10}.

8) The parameter of the temperature regime of the river is a feature characterizing the peculiarities of the temperature regime of the river at the observation site and is within the range of {1 ... 50}.

9) The parameter of the average river flow speed is an attribute characterizing the average rate of water flow in the middle of the riverbed at the observation site.

10) The parameter of the maximum surface speed of the river flow is different from the parameter of the average river flow and is measured on a permanent basis for the site of observing the riverbed.

11) The indicator of the average daily flow rate of the river at the observation site calculated for the period of 24 hours (the indicator is calculated once a day) – the calculating can be done not on the basis of the current measurements, but on the basis of the parameter of the average flow speed.

12) The parameter of the riverbed relief type is a feature characterizing the degree of complexity of the riverbed relief and is within the range of {1 ... 20}.

Based on the set of the presented hydrographic characteristics, at Uzhhorod National University, there were constructed the models of classifying flood phenomena as ACT structures for the river Uzh for the 12-year period (1992 – 2010). To build models of classification trees, the software system "Orion III" was used for generating autonomous recognition and classification systems, where the algorithmic library of the system has 15 algorithms (recognition methods and schemes). Thus, the TS consisted of objects (each of them is described by 18 features) of three basic classes, and at the stage of examining the built classification system (the ACT model) should provide the efficient recognition of objects of the unknown classification in relation to these three classes. The general parameters (characteristics) of the constructed structures (ACT models) are presented in (Table 2).

Notably, the array of the initial training data was dominated by training pairs of class $H_3$ (objects of the situational state of the neutral zone, the green marker), in

the second place with a significant gap in number there were training pairs of class $H_2$ (objects of the situational state of the observation zone, the yellow marker) and in the third place there were directly training pairs of the flood phenomena (objects of the red marker) – class $H_1$. It should be stressed that the capacity of class $H_2$ slightly exceeds the capacity of class $H_1$, this can be explained by the dynamics of changes in the flood situation over time, which can return to the normal state (the neutral zone) – phenomena of class $H_3$, and in most cases goes into crisis (the red zone of flood phenomenon) – class $H_1$ ( Fig. 2). The training dataset consisted of 8391 objects (sets of the known classification) for two monitoring posts in the city of Uzhhorod, and the efficiency of the constructed recognition system was evaluated with the help of the test set of 500 objects for each of the observation posts, and the test dataset was a separate part of the initial training dataset (consisted of discrete objects of the known classification).
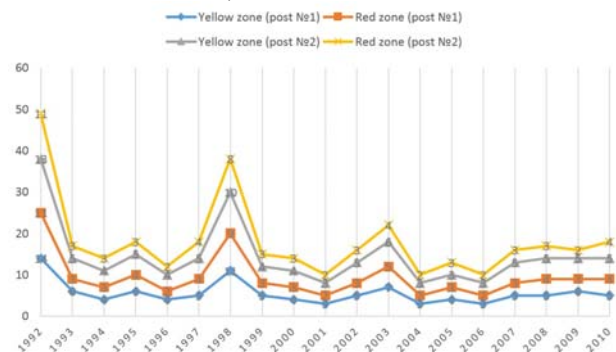


Figure 2 – Flood phenomena of yellow and red zones of the Uzh River during 1992–2010

It must be underlined that this size of the test set is not sufficient for a comprehensive analysis of the quality of the constructed models (structures) of classification trees, but due to the limited nature of the TS itself, even this test set enables the evaluation and analysis of the main parameters of the synthesized ACT structures. The constructed ACT structures (models) provided the necessary level of accuracy and efficiency of the classification specified by the application task, the required speed and the consumption of the working memory of the system, but showed different structural complexity of the constructed classification trees (the parameters of the complexity of the LCT structure – sets of algorithms, sets of generalized features, the number of the vertices of the classification tree structure).

On the basis of the ACT method proposed in this study, classification models of flood phenomena of the Uzh River were built, and admittedly, the ACT structure enables the efficient regulation of the complexity of the classification tree being constructed, or the construction of recognition tree models with the predetermined accuracy as specified by the problem. Clearly, in order to compare and select the specific model of the classification

tree from a fixed set, it is necessary to identify their most important characteristics (feature space dimension, the number of vertices, the number of transitions of the tree structure, etc.) and determine their error in regard to the input array.

An important point is the analysis of quality criteria of the obtained information models which depend on the error of the model, the capacity of the initial array of the training set and test set (the number of training pairs and the feature space dimension specified in the problem), the number of structural parameters of the model and so on. It is evident that the critical parameters of the constructed ACT model that need to be minimized are the model errors respectively within the training dataset, test set and each of the classes (parts, subsets of the initial TS), which are initially specified by the current application task. It has to be emphasized that one of the most important indicators that characterizes the basic properties of the obtained ACT models is the basic indicator of generalizing the initial training dataset by the classification tree (classification model) which is calculated as follows:

$$I_{Main} = \frac{m \cdot O_{Uz}}{Fr_{All} + V_{All} + N_{All} + 2P_{All}}. \quad (9)$$

The basic indicator of generalization of the classification tree model (the algorithm tree structure) reflects its main parameters and characteristics of classification trees. This indicator can also be used as an optimality criterion in the procedure for evaluating an arbitrary tree recognition scheme – the LCTK structure. In a fixed classification problem, $I_{Main}$ it is important to maximize the parameter, this allows you to achieve the most optimal structure of the classification model, and ensures maximum compression of the initial TS data. This refers to the problem of representing an array of source data with a tree of minimal structural complexity [28]. So an important indicator of the quality of the model constructed as a classification tree taking into account the parameters of the structure of the ACT model is the general integrated quality indicator presented in the following way:

$$Q_{Main} = \frac{Fr_{All}}{O_{Uz} \cdot \sum_i p_i} \cdot e^{-\frac{Er_{All}}{M_{All}}}. \quad (10)$$

The set of parameters $p_i$ for the integrated assessment of the quality of the classification tree model represents the most important characteristics of the classification tree (according to the LCT / ACT structures) which is being assessed (e. g. the number of elementary features or generalized features that are used in the classification tree model, the number of transitions between vertices, tiers of classification trees, etc.). A fragment of the main results of comparing constructed structures (structures of algorithm trees based on integral quality assessment), conducted tests of ACT models of

various types, with different sets of classifiers, parameters of structural complexity, restrictions on generating generalized features for the data array of the flood classification problem is reflected in (Table. 3).

## 5 RESULTS

The constructed classification trees (ACT models) provided the necessary quality and speed of schemes of classifying flood phenomena of the Uzh river, at the same time the structure of the tree sheme itself is quite compact – (Table 4). The sets of independent classification algorithms that were selected to generate the groups of GFs also proved their efficiency within this application task. The possible further research may be related to expanding the list of classification algorithms in the ACT scheme as well as additional conditions and restrictions concerning the generation of sets of GFs for each step of the classification tree scheme (ACT structure).
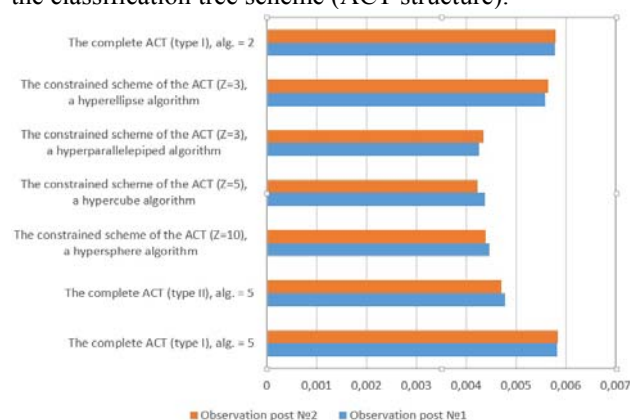


Figure 3 – Integrated indicator of the quality of the constructed ACT models

The presented models of classification trees can be used to assess the general condition of the Uzh river basin (at the observation site) and to identify the situation of the red (flood) zone on the basis of the current measurements of the observation posts. It should be noted that the conducted practical tests of the ACT structures (models) have confirmed the efficiency of the mathematical tool and proposed methods and algorithms for constructing the ACT, the developed software which enables the production of the recommendation on employing this approach (the concept of ACT models) and its software implementation for solving a wide range of classification and recognition tasks in the practical plane.

The only fundamental point to consider when generating ACT models is that usually the consumption of working memory and CPU time of the information system is much higher compared to the structures (models) of the LCT, and it largely depends on the peculiarities of implementing recognition algorithms ( classifiers), the number of algorithms in the ACT scheme, the scheme (the model type) of the ACT structure being generated.

## 6 DISCUSSION

The quality indicators of structures (ACT models) presented in this study can be extended to the structures (trees) of LCTs of this application task. It must be highlighted that this integrated indicator of the quality of the ACT model ranges from zero to one, and the lower it is, the worse the quality of the constructed classification tree will be, and the higher the indicator, the better the obtained model will be. Thus, the proposed integral assessment of the quality of classification tree structures (ACT models) reflects its basic characteristics and can be applied as an optimality criterion in the scheme of qualitative assessment of an arbitrary tree classifier (classifier tree models). The algorithm tree model operates only with ready-made, constructed sets of generalized features, independent classification algorithms. Each of the constructed schemes using the algorithmic tree method will represent a general recognition system, a model of the algorithm tree that can be used for practical work – processing large arrays of experimental data in the form of discrete sets. The constructed classification schemes and models of algorithm trees will represent to a certain extent a new recognition algorithm. Moreover, it is synthesized from well-known recognition algorithms and methods, and the resulting ACT structure, the classification scheme is characterized by high versatility in terms of application and a relatively compact model structure. The algorithm tree requires a large amount of hardware to store generalized features or sets of them and initially evaluate the quality of classification algorithms based on TS data. However, the constructed ACT models in comparison with LCT structures have high performance of classification rules, comparable hardware costs for storing and operating the tree structure itself, and high classification quality.

## CONCLUSIONS

The paper solves the actual problem of constructing ADC models for an array of hydrographic data based on graph – diagram structures of algorithms (schemes for covering data arrays with a set of different types of recognition and classification algorithms).

**The scientific novelty** of the results obtained in this paper lies in the fact that a method for constructing ACT structures for classifying flood phenomena of the Uzh river is proposed based on a scheme for qualitative evaluation of sets of different types of classifiers for synthesizing the structure of the classification tree. In the scheme of the algorithm tree (ACT methods), a fixed subset of data from the initial initial sample is approximated at each stage of constructing the classification tree. The basic scheme for evaluating the effectiveness of algorithms in the ACT method (branching criterion) can be used both to construct the structure of the algorithm tree and to evaluate their approximation ability relative to TS data. This approach makes it possible to achieve low structural efficiency of the synthesized ACT model relative to the initial data set in the future. In this paper, a set of general indicators and parameters is proposed, which makes it possible to effectively represent the general characteristics of the ACT model. It can be used to select the optimal ACT from a set of random classification trees – LCT structures constructed on the basis of methods.

**The practical value** of the obtained results (models of algorithm trees) lies in the fact that the proposed method for constructing ACT models (classification schemes in the form of algorithm trees) was implemented in the algorithm library of the universal software system "ORION III" for solving various types of applied classification problems and predicting large arrays of discrete objects. This toolkit made it possible to carry out an effective classification of flood phenomena of the Uzh River. The conducted practical tests confirmed the operability of mathematical support and constructed ACT models, the developed software, which makes it possible to make a recommendation on the use of this approach and its software implementation for a wide range of applied problems of classification and recognition of discrete objects.

**The perspective research direction** for solving the task of classifying the situational condition of the Uzh River can also be the employment of methods, algorithms and schemes for constructing LCT models with the possibility of comparing them (efficiency, structural complexity and speed of classification) to the constructed ACT structures (models), expanding the library of recognition algorithms (classifiers), optimization of software implementations proposed in this study of constructing the ACT (the evolutionary transition from the C4.5 algorithm to C5.0 can be regarded as an example of such an efficient software optimization), as well as practical testing of the constructed models of classification trees using the training dataset of a larger number of posts of observing the Uzh river basin.

Table 1 – The initial parameters of the task of classifying flood phenomena of the Uzh River basin.

| Number of an observation post № | Dimension of the feature space – N | Capacity of the data array of the initial TS – M | Capacity of the data array of the test set – S | Total number of classes according to partitioning the training dataset – l | The ratio of objects of different classes of the TS ($H_1 / H_2 / H_3$) |
|---|---|---|---|---|---|
| 1 | 18 | 4252 | 500 | 3 | 73/102/4107 |
| 2 | 18 | 4139 | 500 | 3 | 68/97/3974 |

Table 2 – General parameters of the constructed ACT models

| Number of the constructed ACT model | Method (algorithm) of constructing the ACT model | Total number of different classification algorithms applied in the ACT – $N_{All}$ | Total number of GFs (sets of GFs) in the ACT structure $O_{Uz}$ | Total number of all vertices (together with the resulting ones) in the ACT structure | Total time of generating the ACT structure of the current TS |
|---|---|---|---|---|---|
| Observation post №1 | | | | | |
| 1 | The complete ACT method (type I) | 5 restrictions on the sequential use of one-at-a-time algorithm | 42 | 84 | 686 s. |
| 2 | The complete ACT method (type II) | 5 restrictions on the number of GF generations for one step of constructing the ACT | 43 | 98 | 712 s. |
| 3 | The constrained ACT method (Z=10) | 1 a hypersphere algorithm | 35 | 71 | 406 s. |
| 4 | The constrained ACT method (Z=5) | 1 a hypercube algorithm | 58 | 117 | 839 s. |
| 5 | The constrained ACT method (Z=3) | 1 a hyperellipse algorithm | 37 | 75 | 442 s. |
| 6 | The complete ACT method (type I) | 2 a hypersphere algorithm, a hypercube algorithm | 40 | 81 | 706 s. |
| Observation post №2 | | | | | |
| 1 | The complete ACT method (type I) | 5 restrictions on the sequential use of one-at-a-time algorithm | 39 | 79 | 676 s. |
| 2 | The complete ACT method (type II) | 5 restrictions on the number of generations of GFs for one step of constructing the ACT | 40 | 95 | 701 s. |
| 3 | The constrained ACT method (Z=10) | 1 a hypersphere algorithm | 34 | 69 | 380 s. |
| 4 | The constrained ACT method (Z=5) | 1 a hypercube algorithm | 57 | 115 | 828 s. |
| 5 | The constrained ACT method (Z=3) | 1 a hyperparallelepiped algorithm | 48 | 98 | 793 s. |
| 6 | The complete ACT method (type I) | 2 a hypersphere algorithm, a hypercube algorithm | 38 | 77 | 695 s. |

Table 3 – The comparative table of classification tree models (ACT structures) of classifying flood phenomena of the Uzh River basin (of posts №1–№2)

| № | Method (scheme) of synthesizing the structure (model) of the classification tree (LCT / ACT) | Integrated indicator of the quality of the classification tree model $Q_{Main}$ | Total number of errors of the model in the training dataset and test dataset $Er_{All}$ |
|---|---|---|---|
| 1 | The complete ACT method (type I) (the number of algorithms – 5, restrictions on the sequential use of one-at-a-time algorithm) | 0.005821 <br> 0.005845 | 0 <br> 0 |
| 2 | The complete ACT method (type II) (the number of algorithms – 5, restrictions on the number of GF generations for one step of constructing the ACT) | 0.004778 <br> 0.004712 | 0 <br> 0 |
| 3 | The constrained ACT method (Z=10) (the number of algorithms – 1, a hypersphere algorithm) | 0.004464 <br> 0.004389 | 0 <br> 0 |
| 4 | The constrained ACT method (Z=5) (the number of algorithms – 1, a hypercube algorithm) | 0.004387 <br> 0.004228 | 12 <br> 13 |
| 5 | The constrained ACT method (Z=3) (the number of algorithms – 1, a hyperparallelepiped algorithm) | 0.004256 <br> 0.004354 | 6 <br> 6 |
| 6 | The constrained ACT method (Z=3) (the number of algorithms – 1, a hyperellipse algorithm) | 0.005582 <br> 0.005645 | 1 <br> 0 |
| 7 | The complete ACT method (type I) (the number of algorithms – 2, a hypersphere algorithm, a hypercube algorithm) | 0.005790 <br> 0.005801 | 0 <br> 0 |

## REFERENCES

1. Murphy K. Machine Learning: A Probabilistics Perspective, *The MIT Press, Cambridge, Massachusetts*, 2012, 423 p.
2. Gupta Y. Selection of important features and predicting wine quality using machine learning techniques, *Procedia Computer Science*, 2018, Vol. 125, pp. 305–312. DOI:https://doi.org/10.1016/j.procs.2017.12.041
3. Denisko D., Hoffman M. Classification and interaction in random forests, *Proceedings of the National Academy of Sciences of the United States of America*, 2018, Vol. 115, No. 8, pp. 1690–1692. DOI:10.1073/pnas.1800256115.
4. Jordan M. I., Mitchell T. M. Machine learning: trends, perspectives, and prospects, Science 2015, Vol. 349(6245), pp. 255–260. DOI:https://doi.org/10.1126/science.aaa8415
5. Rokach L., Maimon O. Feature set decomposition for decision trees, *Journal of Intelligent Data Analysis,* 2005, Vol. 9, № 2, pp. 131–158. DOI: https://doi.org/10.3233/ida-2005-9202
6. Hyafil L., Rivest R. Constructing optimal binary decision trees is npcomplete, *Information Processing Letters*, 1976, Vol. 5, № 1, pp. 15–17. DOI: https://doi.org/10.1016/0020-0190(76)90095-8
7. Vasilenko Y. A., Vasilenko E. Y., Kuhayivsky A. I., Papp I.O. Construction and optimization of recongnizing systems, *Scientific and technical journal "Information technologies and systems"*, 1999, №1, pp. 122–125.
8. Povkhan I., Lupei M., Kliap M., Laver V. The issue of efficient generation of generalized features in algorithmic classification tree methods, *International Conference on Data Stream Mining and Processing: DSMP 2020 Data Stream Mining & Processing*. Springer, Cham, 2020, pp. 98–113. DOI: https://doi.org/10.1007/978-3-030-61656-4_6
9. Shilen S. Nonparametric classification using matched binary decision trees, *Pattern Recognition Letters*, 1992, No. 13, pp. 83–87. DOI: https://doi.org/10.1016/0167-8655(92)90037-z
10. Perner P. Improving the accuracy of decision tree induction by feature preselection, *Applied Artificial Intelligence*, 2001, Vol. 15, № 8, pp. 747–760. DOI: https://doi.org/10.1080/088395101317018582
11. Povkhan I. A constrained method of constructing the logic classification trees on the basis of elementary attribute selection, *CEUR Workshop Proceedings: Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020)*, Zaporizhzhia, Ukraine, April 15–19, 2020. Zaporizhzhia, 2020, Vol. 2608, pp. 843–857. DOI: https://doi.org/10.15588/1607-3274-2020-2-10
12. Murthy S. K., Kasif S. and Salzberg S. A system for induction of oblique decision trees, *Journal of Artificial Intelligence Research. August* 1994, № 2, pp. 1–33. DOI: https://doi.org/10.1613/jair.63
13. Dovbysh A. S., Moskalenko V. V., Rizhova A. S. Information-Extreme Method for Classification of Observations with Categorical Attributes, *Cibernetica and Systems Analysis*, 2016, Vol. 52, № 2, pp. 45–52. DOI: 10.1007/s10559-016-9818-1
14. Witten I. H., Frank E. Data Mining. Practical Machine Learning Tools and Techniques, Second Edition. San Francisco: Elsevier Inc., 2005, 558 p. DOI: https://doi.org/10.1016/b978-0-12-374856-0.00015-8
15. Geurts P., Irrthum A., Wehenkel L. Supervised learning with decision tree-based methods in computational and systems biology, *Molecular Biosystems*, 2009, Vol. 5, No. 12, pp. 1593–1605. DOI: https://doi.org/10.1039/b907946g
16. Yang J., Li Y. Orthogonal relief algorithm for feature selection, *Lecture Notes in Computer Science*, 2006, pp. 227–234. DOI: https://doi.org/10.1007/11816157_22
17. Rokach L., Maimon O. Data Mining with decision trees: Theory and Applications, *2nd Edition. Singapore: World Scientifc Publishing Co. Pte. Ltd.*, 2015, 305 p.
18. Subbotin S.A. Construction of decision trees for the case of low-information features, *Radio Electronics, Computer Science, Control*, 2019, No. 1, pp. 121–130. DOI: https://doi.org/10.15588/1607-3274-2019-1-12
19. Rodriguez J. J., Kuncheva L. I. and Alonso C. J. Rotation forest: A new classifier ensemble method, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, Vol. 28, No. 10, pp. 1619–1630. DOI: https://doi.org/10.1109/tpami.2006.211

20. What is the C4.5 algorithm and how does it work (2019). Retrieved from https://towardsdatascience.com/what-is-the-c4-5-algorithm-and-how-does-it-work-2b971a9e7db0

21. C5.0 Classification Models (2020). Retrieved from https://cran.r-project.org/web/packages/C50/vignettes/C5.0.html

22. Papagelis A., Kalles D. Breeding S. Decision Trees Using Evolutionary Techniques, *Machine Learning: Proceedings of the Eighteenth International Conference (ICML), June 28–July 1 2001*. Morgen Kaufmann Publishers, 2001, pp. 393–400.

23. Povhan I. F. Logical recognition tree construction on the basis a step-to-step elementary attribute selection, *Radio Electronics, Computer Science, Control*, 2020, № 2, pp. 95–106. DOI: https://doi.org/10.15588/1607-3274-2020-2-10

24. Murthy S., Salzberg S. Decision Tree Induction: How Effective Is the Greedy Heuristic, *Proceedings of the First International Conference on Knowledge Discovery and Data Mining, Montreal*, Kanada, August 20–21 1995, AAAI Press, 1995, pp. 222–227.

25. Subbotin S., Kirsanova E. The regression tree model building based on a clusterregression approximation for data-driven medicine, *CEUR Workshop Proceedings*, 2018, Vol. 2255, pp. 155–169.

26. Harrington P. Machine Learning in Action, Shelter Island, Manning Publications Co, 2012, 354 p.

27. Hssina B., Merbouha A., Ezzikouri H., Erritali M. A comparative study of decision tree ID3 and C4.5, *International Journal of Advanced Computer Science and Applications (IJACSA)*, 2014, pp. 13–19. DOI: https://doi.org/10.14569/specialissue.2014.040203

28. Page D, Ray S. An efficient alternative to lookahead for decision tree induction, *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, August 9–15 2003*. Acapulko, Mexico, Publisher Not Avail, 2003, pp. 601–612.

29. Kaftannikov I. L., Parasich A. V. Decision Tree's Features of Application in Classification Problems, *Bulletin of the South Ural State University. Ser. Computer Technologies,*

30. *Automatic Control, Radio Electronics*, 2015, Vol. 15, № 3, pp. 26–32. DOI: https://doi.org/10.14529/ctcr150304

30. Povhan I. Logical classification trees in recognition problems, *Kwartalnik Naukowo-Techniczny: Informatyka Automatyka Pomiary w gospodarce o ochronie srodowiska*. Krakow, 2020, No. 2, pp. 12–16. DOI: https://doi.org/10.35784/iapgos.927

31. Freund Y., Schapire R. Experiments with a New Boosting Algorithm, *Proceedings Thirteenth of the International Conference on Machine Learning (ICML'96)*, Morgan Kaufmann Publishers Ins, 1996, pp. 148–156.

32. Wang H., Hong M. Online ad effectiveness evaluation with a two-stage method using a Gaussian filter and decision tree approach, *Electronic Commerce Research and Applications*, 2019, Vol. 35, Article 100852. DOI: 10.1016/j.elerap.2019.100852.

33. Giatzitzoglou D. G., Sotiropoulos D. N., Tsihrintzis G. A. AIRS-x: An eXtension to the Original Artificial Immune Recognition Learning Algorithm, 2019 International Conference on Computer Information and Telecommunication Systems (CITS), Beijing, China, 2019, pp. 1–5. DOI: https://doi.org/10.1109/cits.2019.8862043

34. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), Las Vegas, NV, USA, June 27–30*, 2016, pp. 770–778. DOI: https://doi.org/10.1109/cvpr.2016.90

35. Witten I., Eibe F., Hall M. Data Mining: Practical Machine Learning Tools and Techniques, *3rd Edition, Morgan Kaufmann Publishers Ins.*, 2011, 664 p. ISBN 9780123748560.

36. Sirichotedumrong W., Maekawa T., Kinoshita Y., Kiya H. Privacypreserving deep neural networks with pixel-based image encryption considering data augmentation in the encrypted domain, *2019 IEEE International Conference on Image Processing (ICIP), IEEE*, 2019, pp. 674–678. DOI:https://doi.org/10.1109/icip.2019.8804201

УДК 001.891:65.011.56

## МЕТОД ДЕРЕВА АЛГОРИТМІВ В ЗАДАЧІ КЛАСИФІКАЦІЇ ГІДРОГРАФІЧНИХ ДАНИХ

**Повхан І. Ф.** – д-р техн. наук, доцент, доцент кафедри програмного забезпечення систем ДВНЗ Ужгородський національний університет, м. Ужгород, Україна.

**Міца О. В.** – д-р техн. наук, доцент, зав. кафедри інформаційних управляючих систем і технологій. ДВНЗ Ужгородський національний університет, м. Ужгород, Україна.

**Мулеса О. Ю.** – д-р техн. наук, доцент, доцент кафедри кібернетики та прикладної математики ДВНЗ Ужгородський національний університет, м. Ужгород, Україна.

**Поліщук В. В.** – канд. техн. наук, доцент, доцент кафедри програмного забезпечення систем ДВНЗ Ужгородський національний університет, м. Ужгород, Україна.

## АНОТАЦІЯ

**Актуальність.** Робота присвячена виявленню простого та ефективного механізму за допомогою якого можна будувати алгоритмічні дерева класифікації (моделі дерева алгоритмів) за фіксованою початковою інформацією у вигляді навчальної вибірки дискретних даних. Побудоване алгоритмічне дерево класифікації буде безпомилково класифікувати (розпізнавати) всю навчальної вибірки за якою побудована модель, мати мінімальну структуру (структурну складність) та складатися з компонентів – автономних алгоритмів класифікації та розпізнавання в якості вершин конструкції (атрибутів дерева).

**Мета.** Метою даної роботи є створення простого, ефективного та універсального метода побудови моделей класифікації (розпізнавання) на основі концепції дерев алгоритмів для масивів реальних даних гідрографічного характеру, де отримані схеми систем класифікації (структури дерев класифікації) характеризуються деревоподібної структурою (конструкцією) та наявністю автономних алгоритмів класифікації (наборів узагальнених ознак) в якості своїх структурних елементів (блоків конструкції).

**Метод.** Пропонується загальна схема синтезу дерев класифікації у вигляді дерев алгоритмів на основі процедури апроксимації масиву дискретних даних набором елементарних класифікаторів, яка для заданої початкової навчальної вибірки будує деревоподібну структуру – модель дерева алгоритмів. Причому побудована схема складається з набору автономних алгоритмів класифікації та розпізнавання оцінених на кожному кроці, етапі побудови дерева класифікації за даною початковою вибіркою. Розроблений метод побудови алгоритмічного дерева класифікації основна ідея якого полягає в по кроковій апроксимації начальної вибірки довільного об'єму та структури набором елементарних алгоритмів класифікації. Метод дерева алгоритмів при формуванні поточної вершини дерева алгоритмів, вузла, узагальненої ознаки,  забезпечує виділення найбільш ефективних, якісних елементарних класифікаторів з початкового набору та добудову лише тих шляхів в структурі дерева де відбувається найбільша кількість помилок (відмов) класифікації. Структурна складність конструкції дерева алгоритмів оцінюється на основі кількості переходів, вершин та ярусів структури моделі, що дозволяє підвищити якість його наступного аналізу, забезпечити ефективний механізм декомпозиції, та будувати структури дерев алгоритмів в умовах фіксованих наборів обмежень. Метод синтезу дерев алгоритмів дозволяє будувати різнотипні деревоподібні моделі розпізнавання з різними початковими наборами елементарних класифікаторів з наперед заданою точністю для широкого класу задач теорії штучного інтелекту.

**Результати.** Розроблений метод побудови моделей дерев алгоритмів дозволяє працювати з навчальними вибірками великого об'єму різнотипної інформації (дискретними даними) та забезпечує високу швидкість та економність апаратних ресурсів в процесі генерації кінцевої схеми класифікації, будувати дерева класифікації з наперед заданою точністю.

**Висновки.** Розроблений підхід синтезу нових алгоритмів (схем) розпізнавання на основі бібліотеки (набору) вже відомих алгоритмів (методів) та схем. Тобто представлена ефективна схема розпізнавання дискретних об'єктів на основі покрокової оцінки та вибору алгоритмів класифікації (узагальнених ознак) на кожному кроці синтезу схеми. На базі запропонованої концепції алгоритмічних дерев класифікації побудована модель структури дерева алгоритмів, яка забезпечила класифікацію паводкових ситуацій для басейну річку Уж.

**КЛЮЧОВІ СЛОВА:** дерево класифікації, алгоритмічне дерево класифікації, дискретний об'єкт, ознака, функція розпізнавання, алгоритм розпізнавання, критерій розгалуження.

УДК 001.891:65.011.56

## МЕТОД ДЕРЕВА АЛГОРИТМОВ В ЗАДАЧЕ КЛАССИФИКАЦИИ ГИДРОГРАФИЧЕСКИХ ДАННЫХ

**Повхан И. Ф.** – д-р техн. наук,, доцент, доцент кафедры программного обеспечения систем ГВУЗ Ужгородский национальный университет, г. Ужгород, Украина.

**Мица А. В.** – д-р техн. наук, доцент, зав. кафедры информационных управляющих систем и технологий ГВУЗ Ужгородский национальный университет, г. Ужгород, Украина.

**Мулеса О. Ю.** – д-р техн. наук, доцент, доцент кафедры кибернетики и прикладной математики ГВУЗ Ужгородский национальный университет, г. Ужгород, Украина.

**Полищук В. В.** – канд. техн. наук, доцент, доцент кафедры программного обеспечения систем ГВУЗ Ужгородский национальный университет, г. Ужгород, Украина.

## АННОТАЦИЯ

**Актуальность.** Работа посвящена выявлению простого и эффективного механизма с помощью которого можно строить алгоритмические деревья классификации (модели дерева алгоритмов) по фиксированной начальной информации в виде учебной выборки дискретных данных. Построено алгоритмическое дерево классификации будет безошибочно классифицировать (распознавать) всю обучающую выборку по которой построена модель, иметь минимальную структуру (структурную сложность) и состоять из компонентов автономных алгоритмов классификации и распознавания в качестве вершин конструкции (атрибутов дерева).

**Цель.** Целью данной работы является создание простого, эффективного и универсального метода построения моделей классификации (распознавания) на основе концепции деревьев алгоритмов для реальных массивов данных гидрографического характера, где полученные схемы систем классификации (структуры деревьев классификации) характеризуются древовидной структурой (конструкцией) и наличием автономных алгоритмов классификации (наборов обобщенных признаков) в качестве своих структурных элементов (блоков конструкции).

**Метод.** Предлагается общая схема синтеза деревьев классификации в виде деревьев алгоритмов на основе процедуры аппроксимации массива дискретных данных набором элементарных классификаторов, которая для заданной исходной обучающей выборки строит древовидную структуру – модель дерева алгоритмов. Причем построенная схема состоит из набора автономных алгоритмов классификации и распознавания, оцененных на каждом шаге, этапе построения дерева классификации по данной исходной выборке. Разработан метод построения алгоритмического дерева классификации основная идея которого заключается в по шаговой аппроксимации начальной выборки произвольного объема и структуры набором элементарных алгоритмов классификации. Метод дерева алгоритмов при формировании текущей вершины дерева алгоритмов, узла, обобщенной признаки, обеспечивает выделение наиболее эффективных, качественных элементарных классификаторов из начального набора и достройку только тех путей в структуре дерева где происходит наибольшее количество ошибок (отказов) классификации. Структурная сложность конструкции дерева алгоритмов оценивается на основе количества переходов, вершин и ярусов структуры модели, что позволяет повысить качество его последующего анализа, обеспечить эффективный механизм декомпозиции, и строить структуры деревьев алгоритмов в условиях фиксированных наборов ограничений. Метод синтеза деревьев алгоритмов позволяет строить разнотипные древовидные модели распознавания с различными начальными наборами элементов классификаторов с заранее заданной точностью для широкого класса задач теории искусственного интеллекта.

**Результаты.** Разработанный метод построения моделей деревьев алгоритмов позволяет работать с учебными выборками большого объема разнотипной информации (дискретными данными) и обеспечивает высокую скорость и экономность аппаратных ресурсов в процессе генерации конечной схемы классификации, строить деревья классификации с заранее заданной точностью.

**Выводы.** Разработан подход синтеза новых алгоритмов (схем) распознавания на основе библиотеки (набора) уже известных алгоритмов (методов) и схем. То есть представлена эффективная схема распознавания дискретных объектов на основе пошаговой оценки и отбора алгоритмов классификации (обобщенных признаков) на каждом шаге синтеза схемы. На базе предложенной концепции алгоритмических деревьев классификации построена модель структуры дерева алгоритмов, которая обеспечила классификацию паводковых ситуаций для бассейна реки Уж.

**КЛЮЧЕВЫЕ СЛОВА:** дерево классификации, алгоритмическое дерево классификации, дискретный объект, признак, функция распознавания, алгоритм распознавания, критерий ветвления.

### ЛІТЕРАТУРА / ЛИТЕРАТУРА

1. Murphy K. Machine Learning: A Probabilistics Perspective / K. Murphy // The MIT Press, Cambridge, Massachusetts, 2012. – 423 p.
2. Gupta Y. Selection of important features and predicting wine quality using machine learning techniques / Y. Gupta // Procedia Computer Science. – 2018. – Vol. 125. – P. 305–312. DOI:https://doi.org/10.1016/j.procs.2017.12.041
3. Denisko D. Classification and interaction in random forests / D. Denisko, M. Hoffman // Proceedings of the National Academy of Sciences of the United States of America. – 2018. – Vol. 115, № 8. – P. 1690–1692. DOI:10.1073/pnas.1800256115.
4. Jordan M. I. Machine learning: trends, perspectives, and prospects / M. I. Jordan, T. M. Mitchell // Science 2015. – Vol. 349 (6245). – P. 255–260. DOI:https://doi.org/10.1126/science.aaa8415
5. Rokach L. Feature set decomposition for decision trees / L. Rokach, O. Maimon // Journal of Intelligent Data Analysis. – 2005. – Vol. 9, № 2. – P. 131–158. DOI: https://doi.org/10.3233/ida-2005-9202
6. Hyafil L. Constructing optimal binary decision trees is npcomplete / L. Hyafil, R. Rivest // Information Processing Letters. – 1976. – Vol. 5, № 1. – P. 15–17. DOI: https://doi.org/10.1016/0020-0190(76)90095-8
7. Construction and optimization of recongnizing systems / [Y. A. Vasilenko, E. Y. Vasilenko, A. I. Kuhayivsky, I. O. Papp] // Scientific and technical journal "Information technologies and systems". – 1999. – №1. – P. 122–125.
8. The issue of efficient generation of generalized features in algorithmic classification tree methods / [I. Povkhan, M. Lupei, M. Kliap, V. Laver] // International Conference on Data Stream Mining and Processing: DSMP 2020 Data Stream Mining & Processing.  – Springer, Cham. – 2020. – P. 98–113. DOI: https://doi.org/10.1007/978-3-030-61656-4_6
9. Shilen S. Nonparametric classification using matched binary decision trees / S. Shilen // Pattern Recognition Letters. – 1992. – № 13. – P. 83–87. DOI: https://doi.org/10.1016/0167-8655(92)90037-z
10. Perner P. Improving the accuracy of decision tree induction by feature preselection / P. Perner // Applied Artificial Intelligence. – 2001. – Vol. 15, № 8. – P. 747–760. DOI: https://doi.org/10.1080/088395101317018582
11. Povkhan I. A constrained method of constructing the logic classification trees on the basis of elementary attribute selection / I. Povkhan // CEUR Workshop Proceedings: Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020), Zaporizhzhia, Ukraine, April 15–19, 2020. – Zaporizhzhia, 2020. – Vol. 2608. – P. 843–857. DOI: https://doi.org/10.15588/1607-3274-2020-2-10

12. Murthy S. K. A system for induction of oblique decision trees / S. K. Murthy, S. Kasif, S. Salzberg // Journal of Artificial Intelligence Research. August 1994. – № 2. – P. 1–33. DOI: https://doi.org/10.1613/jair.63
13. Dovbysh A. S. Information-Extreme Method for Classification of Observations with Categorical Attributes / A. S. Dovbysh, V. V. Moskalenko, A. S. Rizhova // Cibernetica and Systems Analysis. – 2016. – Vol. 52, № 2. – P. 45–52. DOI: 10.1007/s10559-016-9818-1
14. Witten I. H. Data Mining. Practical Machine Learning Tools and Techniques / I. H. Witten, E. Frank // Second Edition. San Francisco : Elsevier Inc. – 2005. – 558 p. DOI: https://doi.org/10.1016/b978-0-12-374856-0.00015-8
15. Geurts P. Supervised learning with decision tree-based methods in computational and systems biology / P. Geurts, A. Irrthum, L. Wehenkel // Molecular Biosystems. – 2009. – Vol. 5, № 12. – P. 1593–1605. DOI: https://doi.org/10.1039/b907946g
16. Yang J. Orthogonal relief algorithm for feature selection / J. Yang, Y. Li // Lecture Notes in Computer Science. – 2006. – P. 227–234. DOI: https://doi.org/10.1007/11816157_22
17. Rokach L. Data Mining with decision trees: Theory and Applications / L. Rokach, O. Maimon // 2nd Edition. Singapore: World Scientifc Publishing Co. Pte. Ltd. – 2015. – 305 p.
18. Subbotin S. A. Construction of decision trees for the case of low-information features / S. A. Subbotin // Radio Electronics, Computer Science, Control. – 2019. – № 1. – P. 121–130. DOI: https://doi.org/10.15588/1607-3274-2019-1-12
19. Rodriguez J. J. Rotation forest: A new classifier ensemble method / J. J. Rodriguez, L. I. Kuncheva, C. J. Alonso // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2006. – Vol. 28, № 10. – P. 1619–1630. DOI: https://doi.org/10.1109/tpami.2006.211
20. What is the C4.5 algorithm and how does it work (2019). Retrieved from https://towardsdatascience.com/what-is-the-c4-5-algorithm-and-how-does-it-work-2b971a9e7db0
21. C5.0 Classification Models (2020). Retrieved from https://cran.r-project.org/web/packages/C50/vignettes/C5.0.html
22. Papagelis A. Decision Trees Using Evolutionary Techniques / A. Papagelis, D. Kalles, S. Breeding // Machine Learning: Proceedings of the Eighteenth International Conference (ICML), June 28–July 1 2001. – Morgen Kaufmann Publishers. – 2001. – P. 393–400.
23. Povhan I. F. Logical recognition tree construction on the basis a step-to-step elementary attribute selection / I. F. Povhan // Radio Electronics, Computer Science, Control. – 2020. – № 2. – P. 95–106. DOI: https://doi.org/10.15588/1607-3274-2020-2-10
24. Murthy S. Decision Tree Induction: How Effective Is the Greedy Heuristic / S. Murthy, S. Salzberg // Proceedings of

the First International Conference on Knowledge Discovery and Data Mining, Montreal, Kanada. – August 20–21 1995, AAAI Press. – 1995. – P. 222–227.

25. Subbotin S. The regression tree model building based on a clusterregression approximation for data-driven medicine. / S. Subbotin, E. Kirsanova // CEUR Workshop Proceedings. – 2018. – Vol. 2255. – P. 155–169.

26. Harrington P. Machine Learning in Action. / P. Harrington // Shelter Island: Manning Publications Co. – 2012. – 354 p.

27. A comparative study of decision tree ID3 and C4.5 / [B. Hssina, A. Merbouha, H. Ezzikouri, M. Erritali] // International Journal of Advanced Computer Science and Applications (IJACSA). – 2014. – P. 13–19. DOI: https://doi.org/10.14569/specialissue.2014.040203

28. Page D. An efficient alternative to lookahead for decision tree induction. / D. Page, S. Ray // Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence. – August 9–15 2003. – Acapulko, Mexico, Publisher Not Avail. – 2003. – P. 601–612.

29. Kaftannikov I. L. Decision Tree's Features of Application in Classification Problems. / I. L. Kaftannikov, A. V. Parasich // Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics. – 2015. – Vol. 15, № 3. – P. 26–32. DOI: https://doi.org/10.14529/ctcr150304

30. Povhan I. Logical classification trees in recognition problems. / I. Povhan // Kwartalnik Naukowo-Techniczny: Informatyka Automatyka Pomiary w gospodarce o ochronie srodowiska. – Krakow, 2020. – № 2. – P. 12–16. DOI: https://doi.org/10.35784/iapgos.927

31. Freund Y. Experiments with a New Boosting Algorithm. / Y. Freund, R. Schapire // Proceedings Thirteenth of the International Conference on Machine Learning (ICML'96), Morgan Kaufmann Publishers Ins. – 1996. – P. 148–156.

32. Wang H. Online ad effectiveness evaluation with a two-stage method using a Gaussian filter and decision tree approach. / H. Wang, M. Hong // Electronic Commerce Research and Applications. – 2019. – Vol. 35. – Article 100852. DOI: 10.1016/j.elerap.2019.100852.

33. Giatzitzoglou D. G. AIRS-x: An eXtension to the Original Artificial Immune Recognition Learning Algorithm / D. G. Giatzitzoglou, D. N. Sotiropoulos, G. A. Tsihrintzis // 2019 International Conference on Computer Information and Telecommunication Systems (CITS), Beijing, China. – 2019. – P. 1–5. DOI: https://doi.org/10.1109/cits.2019.8862043

34. Deep residual learning for image recognition / [K. He, X. Zhang, S. Ren, J. Sun] // Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), Las Vegas, NV, USA, June 27–30. – 2016. – P. 770–778. DOI: https://doi.org/10.1109/cvpr.2016.90

35. Witten I. Data Mining: Practical Machine Learning Tools and Techniques. / I. Witten, F. Eibe, M. Hall // 3rd Edition, Morgan Kaufmann Publishers Ins. – 2011. – 664 p. ISBN 9780123748560.

36. Privacypreserving deep neural networks with pixel-based image encryption considering data augmentation in the encrypted domain. / [W. Sirichotedumrong, T. Maekawa, Y. Kinoshita, H. Kiya] // 2019 IEEE International Conference on Image Processing (ICIP), IEEE. – 2019. – P. 674–678. DOI: https://doi.org/10.1109/icip.2019.8804201