A Constrained Method of Constructing the Logic Classification Trees on the Basis of Elementary Attribute Selection

Igor Povkhan[1[0000-0002-1681-3466]],

[1] Uzhgorod National University, 89 B, Zankovetska str., Uzhgorod, 88000, Ukraine
Igor.povkhan@uzhnu.edu.ua

Abstract. A problem of constructing the logic classification tree model on the basis of a constrained elementary attribute selection method for the geologic data array has been considered. The method of the real data array approximation by a set of elementary attributes with a fixed criterion of the branching procedure stopping at the stage of constructing the classification tree has been suggested. This method allows the desired model accuracy to be provided, its structural complexity to be reduced and the necessary efficiency indices to be achieved. Based on the suggested elementary attribute selection method modification, the software has been developed enabling a set of different-type applied problems to be used.

Keywords: logic classification tree, image recognition, classification, attribute, branching criterion.

1 Introduction

The classification tree methods, the regressive trees used actively both for the artificial intellect theory problems, being the means of decision adoption support and big data array analysis, and for the related practical field of economy, management etc [1] still remain an important segment of the decision tree concept application. The principal available methods of training selection processing at the recognition function constructing do not allow the required accuracy level of the recognition system to be achieved and their complexity in the data system construction process to be regulated [2, 3]. This shortcoming is not peculiar for the recognition system construction methods based on the logic classification tree methods.

Classification trees are one of the basic methods of automatic data analysis. The first comprehensive studies and concepts of the idea of using the decision/ classification trees date back to the works of S. Hawland and E. Hunt [4]. Note that the logic classification tree (LCT) flexibility, i.e. the ability to take into account and investigate sequentially the effect of the influence of certain variables and structure attributes, is their important specific feature. Respectively, there are a series of reasons providing the LCT structures with higher flexibility as compared to the traditional data analysis methods and tools. For instance, the ability to perform one-dimensional branching to analyze the influence (importance, quality) of certain variables allows one to work with different-type variables in a form of predicates [5].

The problem of selecting the branching stopping rule (the LCT construction stopping criterion) is a basic issue in the classification tree construction. Analyzing the available logic tree construction methods and algorithms, one may distinguish two basic approaches in this direction, i.e. the use of the statistical methods of estimation of the necessity of further manifold partitioning/branching (the so called early stopping or pre-pruning) and application of the logic tree depth (number of layers) limitation scheme.

Thus, let us note that the LCT construction method scheme described, e.g., in [6, 7], has one principal disadvantage related to the fact that the number of elementary attributes in the tree increases considerably with the number of vertices in the logic tree structure (the LCT layer increase). Certainly, such complication of the resulting LCT affects negatively the apparatus possibilities of the recognition system (i.e. memory, processor time). The LCT method modification (the constrained method) on the basis of a stage-by-stage selection of elementary attributes shall be suggested to overcome the above negative issues.

In practice, the LCT construction algorithms and methods quite often provide at the output the structurally complicated logic trees (in the meaning of the number of vertices, number of branches, belonging to the class of non-regular trees) that are not uniformly filled with the data and have different number of branches. Such complicated tree-like structures are quite difficult to be used in the external analysis due to a large number of nodes (vertices) and step-by step partitions of the initial training selection (TS) that contain minimal number of objects (in the worst case, possibly, even single objects). Obviously, it is better to have a certain LCT with a minimal number of vertices (nodes) that correspond to the absolute majority of the initial TS objects. Just at this stage a principal problem of the LCT theory appears, i.e. the problem of possible construction of all the logic tree variants that correspond to the initial TS and the minimal by depth (number of layers) logic tree selection [8–12].

2 Formal problem statement

Let the TS be defined in the following form:

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

Note that here $x_i \in G, f_R(x_i) \in \{0,1,\dots,k-1\}, (i=1,2,\dots,m), m$ is a number of objects with TS; $f_R(x_i)$ is a certain finitely significant function that defines the partition R of the manifold G into the classes (images) $H_0, H_1,\dots,H_{k-1}$. Relation $f_R(x_i)=l$, $(l=0,1,,\dots,k-1)$ means that $x_i \in H_l$, where $x_i=\{x_{i_1}, x_{i_2},\dots,x_{i_n}\}$, x_{i_1} are the values of the $j-$ th attribute for the object x_i, $(j=1,2,\dots,n)$, and n is a number of attributes in the TS.

Thus, the TS is the array (more strictly, the sequence) of certain sets, and each set is the array of values of certain attributes and functions of the above set. One may say that the array of the attribute values is a certain pattern, whereas the value of the function relates the above pattern to the relevant image [13]. A problem is set to construct LCT $-L$ on the basis of the initial TS (1) array and determine the values of its structural parameters p (i.e. $F(L(p,x_i), f_R(x_i)) \to opt$).

3 Literature review

Present study continues a cycle of works dedicated to the problem of the tree-like recognition schemes and the discrete object logic classification methods and algorithms [3, 7–13]. They deal with the issues of constructing, analyzing and optimizing the logic classification trees. For instance, it is known [14] that the resulting classification rule (scheme) constructed by the arbitrary method or by the branched attribute selection algorithm has a tree-like logic structure. The logic tree consists of the vertices (attributes) grouped in layers that are obtained at a certain step (stage) of the recognition tree construction [15–17]. The problem of synthesizing the recognition trees that will be represented by the algorithm tree (graph) is an important one, according to Ref. [18–21]. Unlike the available methods, the main peculiarity of the tree-like recognition systems is that the importance of certain attributes (attribute groups or algorithms) is determined with respect to the function that defines the object partition into the classes [22]. For instance, in Ref. [23], the case of the decision tree construction for a set of low-informative attributes is considered.

The ability of the LCT to execute one-dimensional branching to analyze the influence (importance, quality) of certain variables gives a possibility to work with the different-type variables in a form of predicates (while in case of the algorithmic classification trees (ACT) the relevant autonomous classification/recognition algorithms shall be applied) [24, 25]. Such a logic tree concept is being used actively in the intellectual data analysis, where the target goal is to synthesize a model that shall predict the value of the target variable on the basis of a set of initial data at the system input [26]. Since the principal idea of the branched attribute selection methods and algorithms could be defined as the optimal approximation of a certain initial TS by a set of elementary attributes (the object attributes), then their central problem, i.e. the issue of selecting an efficient branching criterion (vertices, attributes and discrete object attributes) comes on line [27–29]. The work [17] is devoted just to these principal problems by raising the questions of qualitative estimation of certain discrete attributes, sets and their fixed combinations. This allows an efficient mechanism of branching realization to be implemented.

Note that the absolute majority of known decision tree construction algorithms belong to the ‘greedy algorithm’ class. In other words, if at a certain stage some vertex (attribute, node) was selected and used for the initial selection partitioning into the initial selection subsets, the algorithm fails to return at the next stage to selecting the other vertex (node) with higher-quality partition indices. This means, in fact, that at the stage of the logic tree construction it is impossible to determine whether the selected branching vertex will perform some optimal partition [30, 31] at the final stage. For instance, at the decision tree synthesis, the issues of choosing the criterion of the attribute (the LCT vertex), after which the initial TS will begin, the training (the LCT structure construction) stopping criterion and the logic tree (the LCT sub-tree) branch withdrawal criterion will remain the central ones [32].

The logic classification tree structures obtained on the basis of the branched attribute selection methods are characterized, on the one hand, by the compactedness, and, on the other hand, by the non-uniformity of the layer filling (rarity) as compared to the regular trees (the algorithm with a single estimation of the attribute importance [10]. Note that the issues of convergence of the LCT construction process according to the elementary attribute selection methods and those of selecting the criterion of the logic tree synthesis process stopping (e.g. limitation by the logic tree depth or complexity, the accuracy or the structure error number) still remain important [33]. The solution of these issues from the viewpoint of a constrained (modified) LCT construction method shall be presented below.

Thus, notwithstanding a certain range of problems that arise when constructing and using the decision trees (the LCT concept in the general sense), one has to fix their following advantages:

i) the tree-like models are distinguished by a principally fast stage of training the recognition system;
ii) there exists a possibility to synthesize a manifold of decision rules within the area, where it is difficult even for a qualified expert to form a set of recommendations;
iii) synthesis of rules (classification rules, decision rules) in the natural language;
iv) the resulting tree-like model obtained is intuitively understandable;
v) the prediction constructed at the final stage of model operation is characterized by a high accuracy even as compared to the statistical and neural network models;
vi) a possibility to construct non-parametric model.

Hence, taking into account the aforementioned, one may conclude that the decision tree models together with the neural network approach [22, 34] are an important and relevant tool for the data structure extensive analysis and presentation.

4 A concept of training selection data approximation by a set of ranked elementary attributes

Consider some defined general-form (1) training selection (HS) and a certain system (set) of elementary attributes for the initial selection array $\varphi_1(x), \varphi_2(x), ..., \varphi_n(x)$. Let us introduce the following manifolds:

$$G_{r_1,...,r_n} = \{x \in G / a_1(x) = r_1, ..., a_n(x) = r_n\} . \quad (2)$$

Note that the system of manifolds $G_{r_1,\dots,r_n}$ is a complete partition of the manifold G realized by the elementary attributes $\varphi_1, \varphi_2, \dots, \varphi_n$. One has to keep in mind that some of the manifolds $G_{r_1,\dots,r_n}$ may be empty.

Let us set $S_{r_1,\dots,r_n}$ as the number of entrances into the training selection of the pairs $(x_i, f_R(x_i)), (1 \le i \le m)$ that satisfy condition $x_i \in G_{r_1,\dots,r_n}$.

Similarly, we shall set $S^j_{r_1,\dots,r_n}$, $(j = 0,1,\dots,k-1)$ as the number of entrances into the training selection of the pairs $(x_i, f_R(x_i)), (i = 1,\dots, m \le m)$ that satisfy conditions $x_i \in G_{r_1,\dots,r_n}$ and $f_R(x_i) = j$.

Let us introduce the following quantities:

$$\delta_{r_1,\dots,r_n} = \tfrac{S_{r_1,\dots,r_n}}{m} \,;\; \psi^j_{r_1,\dots,r_n} = \tfrac{S^j_{r_1,\dots,r_n}}{S_{r_1,\dots,r_n}} \,;\; \rho_{r_1,\dots,r_n} = \max_j \psi^j_{r_1,\dots,r_n} \,. \tag{3}$$

Note that if $x_i \notin G_{r_1,\dots,r_n}$ for any $(i = 1,\dots,m)$, then $\delta_{r_1,\dots,r_n} = 0$ and $\psi^j_{r_1,\dots,r_n} = 0$, $(j = 0,\dots,k-1)$.

The $\delta_{r_1,\dots,r_n}$ quantity characterizes the frequency of entrances of the members of sequence $x_1, x_2, \dots, x_m$ into the manifold $G_{r_1,\dots,r_n}$. The $\psi^j_{r_1,\dots,r_n}$ quantity characterizes the frequency of the object x belonging to the image (class) H_J, provided $x \in G_{r_1,\dots,r_n}$. It should be noted here that condition $x \in G_{r_1,\dots,r_n}$ is equivalent to the elementary attribute one: $\varphi_1(x) = r_1, \varphi_2(x) = r_2, \dots, \varphi_n(x) = r_n$.

The $\delta_{r_1,\dots,r_n}$ quantity characterizes the information efficiency of recognizing the belonging of x to one of the classes $H_0, H_1, \dots, H_{k-1}$, subject to the basic condition: $x \in G_{r_1,\dots,r_n}$, certainly.

Assume that at the next stage $x \in G_{r_1,\dots,r_n}$. Then the question arises, to what of the classes $H_0, H_1, \dots, H_{k-1}$ x could be related. Obviously, x must be related to the class H_l, for which the following relation holds true:

$$\rho_{r_1,\dots,r_n} = \psi^l_{r_1,\dots,r_n}, \{0 \le l \le k-1\}. \tag{4}$$

Note that this relation is a certain final classification rule. It is obvious that the larger is $\rho_{r_1,\dots,r_n}$, the higher is its general efficiency.

As mentioned above, the initial TS is meant a single information that represents the images $H_0, H_1, \dots, H_{k-1}$. Therefore, the class H_l means a manifold of all the TS pairs $(x_i, f_R(x_i))$ that satisfy the basic relation $f_R(x_i) = l$.

The mean efficiency of recognizing the images $H_0, H_1, ..., H_{k-1}$ defined by the TS data with the help of the elementary attributes $\varphi_1, \varphi_2, ..., \varphi_n$ shall be estimated by the following quantity:

$$o_S(\varphi_1, \varphi_2, ..., \varphi_n) = \sum_{r_1, ..., r_n} \delta_{r_1, ..., r_n} * \rho_{r_1, ..., r_n} . \tag{5}$$

The quantity $o_S(\varphi_1, \varphi_2, ..., \varphi_n)$ could be considered the estimate of the TS approximation by a set of elementary attributes $\varphi_1, \varphi_2, ..., \varphi_n$. From relation (5), the following expressions for $\delta_{r_1, ..., r_n}$, $\psi^j_{r_1, ..., r_n}$ and $\rho_{r_1, ..., r_n}$ result:

$$\delta_{r_1, ..., r_n} \geq 0, \sum_{0 \leq r_1, ..., r_n \leq 1} \delta_{r_1, ..., r_n} = 1 ;$$

$$\psi^j_{r_1, ..., r_n} \geq 0, \sum_{j=0}^{k-1} \psi^j_{r_1, ..., r_n} = 1 ;$$

$$\frac{1}{k} \leq \rho_{r_1, ..., r_n} \leq 1, (r_1, ..., r_n \in \{0,1\}) . \tag{6}$$

From relation (6), the following properties of the quantity $o_S(\varphi_1, \varphi_2, ..., \varphi_n)$ result directly:

i) $\frac{1}{k} \leq o_S(\varphi_1, \varphi_2, ..., \varphi_n) \leq 1$;

ii) $o_S(\varphi_1, \varphi_2, ..., \varphi_n) = 1 \Leftrightarrow \rho_{r_1, ..., r_n} = 1$; (7)

Note that the above properties are valid for any $r_1, ..., r_n$ that satisfy relation $\exists i (1 \leq i \leq m, x_i \in G_{r_1, ..., r_n})$. It follows from the second part of relation (7) that the set of elementary attributes $\varphi_1, ..., \varphi_n$ shall then and only then realize a complete recognition of images $H_0, H_1, ..., H_{k-1}$ (defined by the initial TS). This means that the set of elementary attributes will be a test given $o_S(\varphi_1, \varphi_2, ..., \varphi_n) = 1$, while formula (5) enables one to find such algorithm sets.

Consider one more important peculiarity of the functional estimation of the set of elementary attributes with respect to the initial TS. Let us have a certain number $b, (\frac{1}{k} \leq b < 1)$ and let M_b be a total number of entrances of the objects x_i into the sequence $x_1, ..., x_m$, for which the following relations hold true:

$$\begin{cases} x_i \in G_{r_1, ..., r_n} \\ \rho_{r_1, ..., r_n} > b \end{cases} . \tag{8}$$

The above relations represent the recognition efficiency for the objects x_i by means of a certain set of elementary attributes $\varphi_1,...,\varphi_n$ larger than b.

The number $\gamma_b = \dfrac{m_b}{m}$ is a fraction of entrances of the objects x_i into the sequence $x_1,...,x_m$, for which the recognition efficiency by means of the elementary attribute set $\varphi_1,...,\varphi_n$ is larger than b.

The number γ_b is expressed as:

$$\gamma_b = \sum_{r_1,...,r_n > b} \delta_{r_1,...,r_n} . \tag{9}$$

Note that expression $\sum\limits_{r_1,...,r_n > b}$ means summation over all $r_1,...,r_n$ that satisfy relation $\rho_{r_1,...,r_n} > b$. Let us assume that $o_S(\varphi_1,\varphi_2,...,\varphi_n) = c$ and $\dfrac{1}{k} \le b < c$. Then from expression (5) and relation for $o_S(\varphi_1,\varphi_2,...,\varphi_n) = c$, we have:

$$c = \sum_{\rho_{r_1,...,r_n} > b} \delta_{r_1,...,r_n} * \rho_{r_1,...,r_n} + \sum_{\rho_{r_1,...,r_n} \le b} \delta_{r_1,...,r_n} * \rho_{r_1,...,r_n} . \tag{10}$$

From relations (6) and (9), we obtain as follows:

$$\sum_{\rho_{r_1,...,r_n} \le b} \delta_{r_1,...,r_n} = 1 - \gamma_b . \tag{11}$$

From relations (6), (9), (10) and (11), we have:

$$c \le \gamma_b + (1-\gamma_b) * b . \tag{12}$$

It results directly from relation (12) that:

$$\gamma_b \ge \frac{c-b}{1-b} . \tag{13}$$

Let us substitute $c = 1-\varepsilon^2$ and $b = 1-\varepsilon$ into relation (13), where ε is an infinitely small number. Given such conditions, we have:

$$\gamma_b \ge 1-\varepsilon . \tag{14}$$

It results from relation (14) that, if $c \to 1$, then there exists the quantity b (dependent on c) that:

$$\frac{1}{k} \le b < c, \lim_{c \to 1} b = 1 .$$

Note that the quantity b in this case does not depend on the initial TS (and, thus, on its capacity and number m, as well).

5 A constrained method of classification tree construction on the basis of elementary attribute selection

Note that the main idea of the stage-to-stage elementary attribute selection is to maximize the value of the attribute $W_m(f)$ quality [2]. The latter means that such generalized attribute f must be found in the logic tree algorithms for the training selection (1), for which the value of $W_m(f)$ is as large as possible.

It should be noted that the elementary attribute importance (informativeness) means here the value that could be calculated, as a variant, by the following functionals [13].

It should be noted that each LCT vertex (Fig. 1) comprises either certain attribute (mark) φ_i^j or number m_i^j that belongs to the manifold $\{0,1,...,k-1\}$. Since the vertex containing m_i^j is known as the final LCT vertex, it is also called usually the logic tree leaf.

Two guiding lines (arrows) denoted 0 and 1 go out from the vertex with the attribute φ_i^j. The guiding line 0 corresponds to the value $\varphi_i^j = 0$, whereas the guiding line 1 – to $\varphi_i^j = 1$. The logic tree is conditionally divided into the layers (levels), and the j - th LCT layer contains the corresponding attributes $\varphi_1^j, \varphi_2^j, ...,$.

Note that all the elementary attributes at all the LCT layers, beginning from the first one and ending by the n - th one, represent, in fact, the attributes obtained after the n steps (stages) of the LCT construction. The elementary attributes located at the n - th layer are those derived, respectively, at the n - th step (stage) of the logic tree construction process.

Note that such classification tree (Fig. 1) realizes (represents), in fact, some generalized attribute (x) defined at the manifold G that takes a value from the manifold $\{0,1,...,k-1\}$. After constructing the attribute $f_i(x)$, a stage of a check-test begins. Thus, in the check-test regime, a total number S of all those pairs $(x_i, f_R(x_i))$ from selection (1), for which relation $f_R(x_i) = f_i(x)$ holds true, is calculated.

At the next stage, we have to check condition $\frac{S}{m} \geq \delta$. Here δ is a parameter that characterizes the estimate of training efficiency with respect to the current problem (TS). If this condition is satisfied, the classification tree construction process terminates, and the generalized attribute $f_i(x)$ represented by the structure tree (Fig. 1) ensures selection (1) approximation.

Let us emphasize that such scheme of constructing the logic classification tree has one principal disadvantage related to the fact that the number of elementary attributes φ_i^j in the tree increases considerably (here i is the elementary attribute number in the set, j is that of the attribute location layer). Certainly, such complication of the resulting classification tree affects negatively the apparatus capabilities of the recognition system (i.e. memory, processor time).

To overcome the above negative issues, one may suggest the following modification of the TS data approximation method using the elementary attribute set.

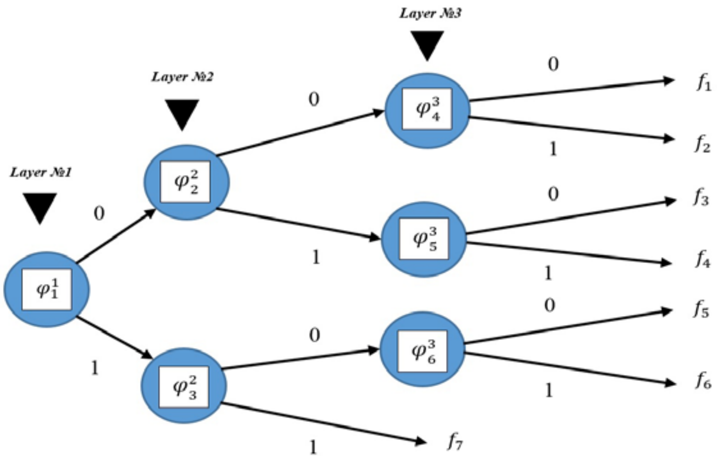

Fig. 1. Finished logic classification tree structure.

Let us fix some positive number Z. Consider a logic tree with structure shown in Fig. 1 that reflects a certain predicate (generalized attribute) $f_i(x)$.

So, at the test stage we calculated a certain number S that occurs in relation $\frac{S}{m} \geq \delta$. Now, besides the number S, for each unfinished path $r_1 r_2 r_3$ of the logic tree (Fig. 1), we calculate also the number $S_{r_1 r_2 r_3}$, where $S_{r_1 r_2 r_3}$ is a number of all pairs $(x_i, f_R(x_i))$ from TS that, in fact, belong to the path $r_1 r_2 r_3$ and satisfy relation $f_R(x_i) = l(r_1 r_2 r_3)$. Thus, $S_{r_1 r_2 r_3}$ is a number of all those errors made by a certain predicate $f_i(x)$ (generalized attribute) represented by the LCT (Fig. 1) on the path $r_1 r_2 r_3$.

At the next step, we choose the number S of such paths $(r_1 r_2 r_3)_1, \ldots, (r_1 r_2 r_3)_Z$, for which the number $S_{r_1 r_2 r_3}$ will be maximal. For example, let $Z = 3$, and the following relation holds true: $S_{000} \geq S_{100} \geq S_{101} \geq S_{001} \geq S_{010} \geq S_{011}$.

Then the paths $000, 100, 101$ will be chosen only. The further construction/selection of vertices (elementary attributes) $\varphi_{r_1 r_2 r_3}$ will be realized for the above paths only.

We shall call this classification tree scheme a constrained method of the LCT structure construction, because, according to this scheme, the paths with the maximal number of errors will continue only.

It should be noted here that, in case of using the above processes at the end of the paths $r_1r_2r_3$ that do not belong to the selected Z paths, the values $l(r_1r_2r_3)$ shall be preserved.

Note that the process of the constrained method of the LCT structure construction could be applied in case when the initial TS (1) is not fixed, i.e. when at each process step a separate selection is made.

6 Model selection

Note that the LCT construction scheme suggested above allows one to regulate the classification tree model complexity or to develop a model with prescribed accuracy. Here the task of choosing the classification tree model from a set of constructed LCTs for a certain problem is determined by an array of parameters having an essential importance with respect to a current applied problem (the TS data set).

Obviously, in order to compare the LCT models and select a particular one, we have to distinguish its principal parameters (i.e. the image space dimensionality, the number of vertices etc) and determine their errors with respect to the array of the input data.

It is principally important to consider, at this stage of study, the quality criteria of obtained information models dependent on the model errors, initial TS data array capacity (the number of training pairs and the problem attribute space dimensionality) and the number of the model parameters etc.

It is apparent that the model errors at the TS and ST data arrays for each of the classes defined by the initial condition of the current applied problem are critically important parameters of developed LCT model, and they must be necessarily minimized.

Note that reduction of the LCT structure complexity (i.e. the number of attributes in the LCT structure, the total number of vertices in the LCT model and that of transitions in the LCT structure), as well as the parameters of the information system total memory consumption and processor time remain here a principal moment. Thus, a total integral quality index in the following form:

$$Q_{Main} = \frac{Fr_{All}}{V_{All} \cdot \sum_i p_i} \cdot e^{-\frac{Er_{All}}{M_{All}}}. \qquad (15)$$

is an important quality index of constructed model in a form of the classification tree with the parameters of the LCT model structure being taken into account.

Note that here the parameter Er_{All} is a total number of the LCT model errors at the data arrays of the initial test and training selections, respectively, while M_{All} is a total capacity of these two data arrays. The Fr_{All} parameter characterizes a number of vertices in the obtained LCT model with the resulting values f_R (RF, i.e. the classification tree leaves), whereas the V_{All} parameter is a total number of all the types of vertices in the LCT model structure. A set of parameters p_i is the most important characteristic of the classification tree to be estimated (the number of elementary attributes used in the classification tree model, the number of transitions between the classification tree vertices, layers etc).

It should be noted that this integral LCT model quality index will take values from zero to unit. The less is this index, the worse is the quality of the constructed tree, and, in contrary, the larger is this index, the better the model obtained is.

7 Experiments and results

The suggested constrained method of constructing the LCT models (i.e. the modified method of elementary attribute selection) has been approved for the problem of classifying an array of geological data (an autonomous recognition system was constructed for it on the basis of a tree-like classification model), namely, a problem of oil-bearing formation separation. The initial parameters of the above applied problem are listed in the following table.

Table 1. Initial parameters of the problem.

Attribute space dimensionality – N	Initial TS data array capacity – M	ST data array capacity – S	Total class number over the TS data partition – l	Different TS class object ratio – (H_1 / H_2)
22	1,250	240	2	756/494

Information about partitioning into two classes is presented in the TS. At the stage of examination, the constructed classification system must provide effective recognition of objects with unknown classification with respect to these two classes. Note that at the initial stage the training and the test selections were automatically checked for the correctness (searching and deleting the similar objects of different belonging, i.e. the first and second-type errors).

It should be noted that the training pairs of the H_1 class (the oil-bearing formations) in the $\approx (1.5/1)$ proportion dominated in the training information array, while the TS itself consisted of 1,250 objects. The constructed recognition system efficiency was estimated at the test sampling of 240 objects, whereas the ST was a segregated TS part (consisting of the discrete objects of known classification). The training and the test selection array data were obtained on the basis of the geological survey on the territory of Transcarpathian province during the 2001–2011 period.

A fragment of the principal results of the above experiments (comparative tests of the LCT model construction methods at the applied problem data array) is presented in Table 2. The constructed LCT models provided the desired level of accuracy preset by the problem condition, system response rate and operating memory consumption, but demonstrated different structural complexity of classification tree and generalized attribute set construction (in the case of the algorithmic classification tree model [7]).

Note that estimation of the classification tree model quality suggested here reflects the basic parameters (characteristics) of classification trees and could be used as the optimality criterion in the procedure of estimating an arbitrary tree-like recognition system, for instance, in case of the random LCT construction and selection methods taken from Ref. [17].

The constrained method of elementary attribute selection (the modified LCT construction method) suggested in this work was compared to the complete LCT method, the algorithm with a single (initial) estimation of the discrete attribute importance and the algorithmic classification tree method, and has shown, in general, a reasonable result.

Table2. Comparative table of classification tree models /methods.

No.	Logic tree structure synthesis method	Integral model quality index Q_{main}	Total number of errors of the model at the TS and ST Er_{All}
1	Complete LCT method on the basis of elementary attribute selection	0,004789	2
2	LCT model with a single attribute importance estimation	0,002263	3
3	Constrained LCT construction method	0,003181	2
4	Algorithmic tree method (type I)	0,005234	0
5	Algorithmic tree method (type II)	0,002941	0

And, finally, please note that the principal idea of classification tree method on the basis of the autonomous algorithms in the own structure is to provide the stage-to-stage approximation of the initial TS data array by the selected algorithm set (that is reflected in the tree construction itself). On the one hand, the ACT structure obtained is characterized by a high flexibility with respect to the applied problems and a relatively compact structure of the model itself, while, on the other hand, it requires the sufficient apparatus costs (i.e. memory and processor time) to save the generalized attributes (classification rule parameters) and the initial quality estimation of the fixed classification algorithms in accordance with the TS data. Therefore, as compared to the ACT concept, the LCT method (i.e. the constrained elementary attribute selection method) possesses high classification scheme response rate, relatively low apparatus costs for the tree structure (the data structure) saving and operating and high quality (regulated complexity) of the discrete object classification.

8 Conclusions

Thus, one may conclude that the problem of the LCT construction automation on the basis of a constrained method of the TS approximation by a set of elementary attributes has been solved and the scheme of the classification tree construction with the prescribed accuracy has been suggested. Present paper analyzes the constrained method of the LCT model construction that allows the prescribed-accuracy classification trees to be constructed by regulating the complexity of the scheme under generation. Such approach enables the optimal structure of the LCT model under construction to be achieved and provides the necessary and sufficient accuracy of the classification tree obtained.

The scientific novelty of the obtained results is that for the first time a simple constrained method/scheme of the LCT construction has been suggested for the first time on the basis of selecting the elementary attributes with the permanent estimation of their importance at each step of classification tree construction with the ability of a further priority-related construction of the fixed LCT construction blocks. Note that this method at each branching (LCT generating) step takes into account the influence of a certain object attribute value on the resulting RF value f_R in the classification tree structure.

Branching criteria in the LCT structure presented in this paper could be effectively used not only to estimate the informativeness efficiency of certain elementary attributes, but to calculate the importance of the attribute sets and combinations as well. This allows the question of specific features of the LCT model construction for less informative attributes to be raised [23]. We have also suggested a general integral index of the LCT model quality that allows the general classification tree characteristics to be represented effectively. Moreover, it can be used to select the most optimal LCT in case of the methods of random LCT construction [17].

The practical merit of the results obtained lies in the fact that suggested constrained LCT model construction method provides possibility to construct cost-saving and efficient classification models with prescribed accuracy. This method was realized in the ORION III system algorithm library to solve different applied classification problems. The above practical applications have confirmed the performance ability of the LCT models constructed and software developed. These studies may be addressed in future towards the further development of the LCT methods, optimization of the software realizations of the suggested constrained method of the LCT construction, as well as towards its practical approval for a number of real problems in the field of classification and recognition.

References

1. Quinlan, J.R.: Induction of Decision Trees. Machine Learning, no. 1, pp. 81-106 (1986).
2. Vasilenko, Y.A., Vasilenko, E.Y., Povkhan, I.F.: Conceptual basis of image recognition systems based on the branched feature selection method. European Journal of Enterprise Technologies, no. 7(1), pp. 13-15 (2004).
3. Povhan, I.: General scheme for constructing the most complex logical tree of classification

in pattern recognition discrete objects. Collection of scientific papers. Electronics and information technologies, vol. 11, pp. 112-117. Lviv (2019).
4. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer, Berlin (2008).
5. Srikant, R., Agrawal, R.: Mining generalized association rules. Future Generation Computer Systems, vol. 13, no. 2, pp. 161-180 (1997).
6. Vasilenko, Y.A., Vasilenko, E.Y., Povkhan, I.F.: Branched feature selection method in mathematical modeling of multi-level image recognition systems. Artificial Intelligence, no. 7, pp. 246-249 (2003).
7. Povkhan, I.F.: Features of synthesis of generalized features in the construction of recognition systems using the logical tree method. In: Materials of the international scientific and practical conference Information technologies and computer modeling ITKM-2019. Ivano-Frankivsk, pp. 169-174 (2019).
8. Povhan, I.: Designing of recognition system of discrete objects, IEEE First International Conference on Data Stream Mining & Processing (DSMP), Lviv, 2016, Ukraine, pp. 226-231 (2016).
9. Shai, B.: Decision Trees. Understanding Machine Learning. Cambridge University Press, (2014).
10. Vasilenko, Y.A., Vashuk, F.G., Povkhan, I.F.: The problem of estimating the complexity of logical trees recognition and a general method for optimizing them. European Journal of Enterprise Technologies, 6/4(54), pp. 24-28 (2011).
11. Vtogoff, P.E.: Incremental Induction of Decision Trees. Machine Learning, no. 4, pp. 161-186 (2009).
12. Elmanova, N.A.: The construction of decision trees. Computer Press, no. 12, pp. 38-45 (2003).
13. Povkhan, I.F.: The problem of functional evaluation of a training sample in discrete object recognition problems. Scientific notes of the Tauride national University. Series: technical Sciences, vol. 29(68), no. 6, pp. 217-222 (2018).
14. Breiman, L.L., Friedman, J.H., Olshen, R.A., Stone, C.J.: Classification and regression trees. Boca Raton, Chapman and Hall/CRC (1984).
15. Laver, V.O., Povkhan, I.F.: The algorithms for constructing a logical tree of classification in pattern recognition problems. Scientific notes of the Tauride national University, Series: technical Sciences, vol. 30(69), no. 4, pp. 100-106 (2019).
16. Povkhan, I.F.: The concept of function and algebraic recognition scheme in classification problems of discrete objects. Scientific notes of the Tauride national University, Series: technical Sciences, vol. 30(69), no. 2, pp. 171-177 (2019).
17. Povkhan, I.F.: Features random logic of the classification trees in the pattern recognition problems. Scientific notes of the Tauride national University, Series: technical Sciences, vol. 30(69), no. 5, pp. 152-161 (2019).
18. Deng, H., Runger, G., Tuv, E.: Bias of importance measures for multi-valued attributes and solutions. In: Proceedings of the 21st International Conference on Artificial Neural Networks (ICANN), pp. 293-300 (2011).
19. Kamiński, B., Jakubczyk, M., Szufel, P.: A framework for sensitivity analysis of decision trees. Central European Journal of Operations Research, 26 (1), pp. 135-159 (2017)
20. Karimi, K., Hamilton, H.J.: Generation and Interpretation of Temporal Decision Rules. International Journal of Computer Information Systems and Industrial Management Applications, vol. 3, pp. 314-323 (2011).
21. Kotsiantis, S.B.: Supervised Machine Learning: A Review of Classification Techniques. Informatica, no. 31, pp. 249-268 (2007).
22. Bodyanskiy, Y., Vynokurova, O., Setlak, G. and Pliss, I.: Hybrid neuro-neo-fuzzy system and its adaptive learning algorithmю. In: Xth Scientific and Technical Conference Computer Sciences and Information Technologies (CSIT), 2015, Lviv, pp. 111-114 (2015)

23. Subbotin, S.A.: Construction of decision trees for the case of low-information features. Radio Electronics, Computer Science, Control, no. 1, pp. 121-130 (2019).
24. Vasilenko, Y.A., Vasilenko, E.Y., Povkhan, I.F.: Defining the concept of a feature in pattern recognition theory. Artificial Intelligence, no. 4, pp. 512-517 (2002).
25. Subbotin, S., Oliinyk, A.: The dimensionality reduction methods based on computational intelligence in problems of object classification and diagnosis. Recent Advances in Systems, Control and Information Technology [eds.: R. Szewczyk, M. Kaliczyńska]. Springer, Cham, pp. 11-19 (Advances in Intelligent Systems and Computing, vol. 543). (2017).
26. Koskimaki, H., Juutilainen, I., Laurinen, P., Roning, J.: Two-level clustering approach to training data instance selection: a case study for the steel industry. In: Neural Networks: International Joint Conference (IJCNN-2008), Hong Kong, 1-8 June 2008 - proceedings. Los Alamitos, IEEE, pp. 3044-3049 (2008). doi: 10.1109/ijcnn.2008.4634228
27. Subbotin, S.A.: Methods and characteristics of localitypreserving transformations in the problems of computational intelligence. Radio Electronics, Computer Science, Control, no. 1, pp. 120-128 (2014).
28. Subbotin, S.: The neuro-fuzzy network synthesis and simplification on precedents in problems of diagnosis and pattern recognition. Optical Memory and Neural Networks (Information Optics), vol. 22, no. 2, pp. 97-103 (2013). doi: 10.3103/s1060992x13020082
29. Subbotin, S.A.: Methods of sampling based on exhaustive and evolutionary search. Automatic Control and Computer Sciences, vol. 47, no. 3, pp. 113-121 (2013). doi: 10.3103/s0146411613030073
30. Miyakawa, M.: Criteria for selecting a variable in the construction of efficient decision trees. IEEE Transactions on Computers, vol. 38, no. 1, pp. 130-141 (1989).
31. Painsky, A., Rosset, S.: Cross-validated variable selection in tree-based methods improves predictive performance. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 11, pp . 2142-2153 (2017). doi:10.1109/tpami.2016.2636831
32. De Mántaras, R. L.: A distance-based attribute selection measure for decision tree induction. Machine learning, vol. 6 (1), pp. 81-92 (1991).
33. Deng, H., Runger, G., Tuv, E.: Bias of importance measures for multi-valued attributes and solutions. In: 21st International Conference on Artificial Neural Networks (ICANN), Espoo, 14-17 June 2011 - proceedings. Springer-Verlag, Berlin, vol. 2, pp. 293-300 (2011)
34. Lupei, M., Mitsa, A., Repariuk, V., Sharkan, V.: Identification of authorship of Ukrainian-language texts of journalistic style using neural networks. Eastern-European Journal of Enterprise Technologies, 1 (2 (103)), pp. 30-36 (2020). doi: https://doi.org/10.15587/1729-4061.2020.195041