The problem of convergence of classifiers construction procedure in the schemes of logical and algorithmic classification trees

Igor Povkhan[a], Oksana Mulesa[a], Olena Melnyk[a], Yuriy Bilak[a], Volodymyr Polishchuk[a]

[a] *Uzhhorod National University, Zankoveckoy str., 89B, Uzhhorod, 88000, Ukraine*

Abstract

The paper considers the problem of convergence in the procedure of classifier schemes synthesis by methods of logical and algorithmic classification trees. It suggests the upper evaluation of complexity for the scheme of algorithms tree in the problem of approximation of real data array by a set of generalized features with a fixed criterion of termination of the branching procedure at the stage of the classification tree construction. This approach provides the required accuracy of the model, evaluates its complexity, reduces the number of branches, and achieves the required performance features. The paper represents the evaluation of convergence for the procedure of recognition schemes construction is logical and algorithmic classification trees on conditions of weak and robust separation of primary initial sampling classes.

Keywords

logical tree, algorithmic tree, classifier, pattern recognition, feature, initial sample.

1. Introduction

Problems united by pattern recognition are various and currently widely spread in both the economic and social content of the human activity, leading to the necessity of building and investigating the mathematical models of the relevant systems [1-5]. A universal approach to their solution is still missing, while plenty of general theories and approaches help solve different problem types (classes). However, their practical application varies by high sensitivity to the specific parameters of the problem itself or subject area of application [6]. Various theoretical results derive from exceptional cases and subproblems, pointing out that the bottleneck of successful real recognition systems requires performing a considerable amount of computation and focusing on powerful hardware tools. The classification trees (solution trees) have fixed a significant part of the above shortcomings. It enables to work effectively with problems of arbitrary scale data (where the information is set in natural form) [7-10].

Today there are various relevant approaches to constructing classification systems in the form of logical trees and algorithmic classifications (LCT/ACT) [11-13]. Moreover, the interest in recognition methods using LCT is caused by several valuable properties. From one side, the complexity of the class of recognition functions (RF) in the form of LCT models, under certain conditions, does not exceed the complexity of the class of linear recognition functions (the simplest of the known). On the other side, RF in the form of classification trees allows distinguishing in the process of classification both causal links (and unambiguously take them into account in the future) and factors of chance or uncertainty. Additionally, it considers both functional and stochastic links between properties and

behavior of the entire system and the process of missing interactions classification between environmental objects, different animal species, and people (except the objects, information about which is transmitted genetically (hereditary) or other). It takes place according to the so-called logical decision tree [14].

In most forecasting and classification problems, which use unstructured data (such as sets of discrete patterns or text arrays), an artificial neural network (of selected type) outperforms all other types of algorithms or decision tree frameworks. Otherwise (in the case of structured high-volume discrete data arrays), the methods and algorithms of the decision tree concept are vastly preferred [15-17]. Practically, LCT algorithms and construction methods usually give structurally complex logical trees at the output (in terms of the number of vertices, number of branches belonging to the class of irregular trees), which are unevenly filled with data and have different numbers of branches. Such complex tree-like structures are difficult to perceive for external analysis due to a massive number of nodes (vertices) and a massive number of stepwise partitions of the primary initial sample (IS) containing a minimum number of objects (maybe even single objects in the worst case). In practice, let us review the fundamental question regarding methods of classification trees (classification models) – the question of convergence of procedure for constructing a classification tree (methods of classification trees), structures LCT/ACT.

2. Formal problem statement

Let it be the IS in the following form:

$$(x_1, f_R(x_1)),...,(x_m, f_R(x_m)) \tag{1}$$

Consider $x_i \in G, (i = 1,...,m)$, where $m-$ is the quantity of objects with the IS, $f_R(x_i)$ – is a finite value function that designates the partition of R for the set G into classes (patterns) $H_0,...,H_{k-1}$. Ratio $f_R(x_i) = l$ $(l = 1,...,k-1)$ means $x_i \in H_l$, $x_i = \{x_{i_1},...,x_{i_n}\}$, $x_{i_j} -$ value $j-$ kind features for object x_i, $(j = 1,...,n)$, $n-$ quantity of features in IS.

As a result, IS is a population (or the sequence) of the set, where each set is a population of indicated values and functional values [18]. Additionally, the set of indicated values is a particular image, where the value of the function relates it to the particular pattern. The problem is to build a structure LCT/ACT $-L$ based on the array of primary IS base of the type (1) and determine the value of its structural parameters p (that is $F(L(p,x_i), f_R(x_i)) \to opt$).

3. Literature review

All basic approaches in the theory of recognition have their advantages and disadvantages and form a single toolkit for solving practical problems of artificial intelligence theory — especially the integrally elaborated classical algebraic approach developed by Y.I. Zhuravlyov [19]. This direction of recognition theory development is connected with the classification algorithms model construction and the optimized quality recognition algorithm selection. The research focuses on the actual concept of decision trees (LCT structures). In particular, The classification scheme, which is given by an arbitrary approach and method and algorithm of the classification tree, has a tree-like logical structure [1,3,20]. Moreover, the structure of the logical tree consists of vertices (features), grouped by circles and built (selected) at a certain step (stage) of classification tree model construction [21]. The main peculiarity of tree-like recognition systems is that the importance of individual features (groups of features or sets of features) is determined relative to the function that defines the division of objects into classes [22].

In the research [21], a generation scheme for the classification tree structure is suggested based on the step-by-step selection of elementary attributes, the disadvantage of which is the high dependence of the model complexity on the effectiveness of the final minimization, tree cutting procedure. In

researches [23-25], the modular scheme of classifiers construction has been offered in the form of classification tree structures, which circumvents the limitations of traditional decision tree methods. The research [24] proposes an efficient scheme for generating generalized features based on constructing sets of geometric objects. However, the disadvantage is the limitation of the initial training structure of the sample and the lack of practical application versatility. In addition, challenges of LDK models structural complexity estimating the minimization stage evaluated in research [20].

Thus, the research [22] shows that the resulting classification rule has a tree-like logical structure built by an arbitrary method or the indication branched selectional algorithm. In this case, the question of qualitative branching criterion selection comes first. The logical tree consists of vertices grouped on tiers and received at a particular step of recognition tree construction [25]. Thereby, it challenges the effective structure minimization for the constructed model of the classification tree. A significant problem that arises from article [23] is the synthesis of recognition trees, which the actual tree of algorithms will represent. An essential area of LDK structures research remains the issues of the generation of decision trees for the case of uninformative features [14] and the actual issue of classification trees theory. It questions the possibility of constructing all logical tree variants, which correspond to primary IS, and selecting minimum depth or structural complexity (number of tiers) classification tree [26-30].

4. Convergence of synthesis of logical and algorithmic classification tree structure

Consider that at every step during building a logical tree (some LCT model), only one selected elementary feature is picked from the set of the fixed features $(\varphi_1,...,\varphi_n)$. Then on $n-$ th step of the tree classification construction procedure, the LCT scheme represents by itself a predicate p_n (generalized feature, which is built from the set of elementary features) [23, 30], that is the most effective approximation of the initial IS of general view (1) (applicable for the ACT structure case).

In particular, p_n represents some tree-like scheme (classification tree), which consists of n vertices. The structure of the predicate p_n includes only n elementary features (attributes of the IS discrete object) from the initial set.

The sequence of predicates $p_1,...,p_j$ (generalized features) coincides with the primary IS of the (1) kind. In case it starts from some Q, the following condition fulfills:

$$f_{Q+m} = f_R(x_i),(i=1,...,m),(m \geq 0).$$

Let us denote with φ_n some elementary features selected (fixed) on $n-$ th step in the scheme of the LCT model construction. The feature φ_n corresponds to some fixed path $r_1, r_2,...,$ which ends by the given attribute (the vertex of classification tree – LCT model). For example, on Fig. 1 there is the LCT in which the vertex φ_2 (feature) corresponds to the path $\{0\}$, and the vertex φ_5 corresponds to the path $\{0,1\}$.

The path that corresponds to the elementary sign φ_n as indicated, let us denote as T_n, and with D_n let us denote the set of those pairs $(x_i, f_R(x_i))$ of the primary IS of the general view (1), where objects w_i belong to the path T_n. For example, for the LCT structure (Fig. 1), let $\varphi_n = \varphi_4$, then the path T_n looks as $\{1,0\}$.

In such a case some object w_i belongs to the path $\{1,0\}$, if the following conditions are fulfilled: $\varphi_1(w_i)=1$ and $\varphi_1(w_i)=0$.

We consider further, that the elementary feature φ_n weakly separates the set D_n, if in D_n there are such pairs $(x_i, f_R(x_i))$ and $(x_j, f_R(x_j))$, that $\varphi_n(x_i) = 0$ and $\varphi_n(x_i) = 1$ (that is $\varphi_n(x_i) \neq \varphi_n(x_j)$).
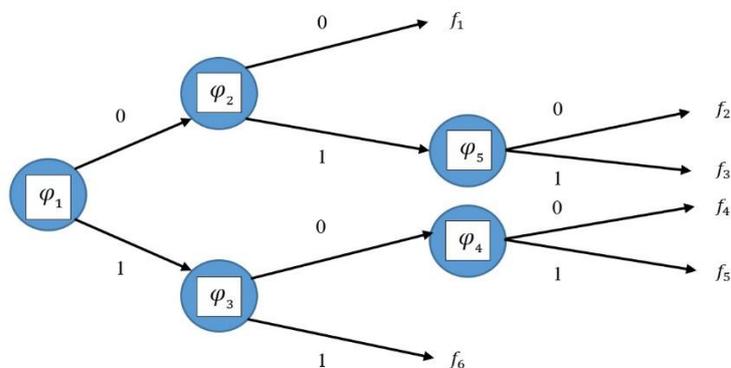
.

Figure 1: LCT structure with elementary features as vertices.

In the final power of the scheme of classification tree method (models LCT/ACT), we call the number of all finite vertices (certain sheets) of the scheme. For example, for the LCT on Fig. 1 the power is 6.

Obviously, the final power of the classification tree method scheme is also equal to the number of all final paths in it. It is clear that with induction of n, it can be easily proved that the final power of each of the above schemes p_n (predicates) is equal to $n+1$. Indeed, the fact that the final power p_1 which includes only one feature or algorithm (cases with LCT/ACT), equals 2, is obvious.

Let the final power of the scheme p_n equals to $n+1$. Let us calculate the final power p_{n+1}. It is clear that this scheme is based on the scheme p_n, when in some final vertex, a new vertex is successively added (feature, algorithm) with the number $n+1$. Obviously, when adding this feature (algorithm) to the scheme p_n one end vertex disappears and two new end vertices are added. Therefore, we can conclude that the number of all end vertices of the scheme p_n equals to $n+2$.

Let us assume that on each $n-$ th step of the classification tree construction procedure (of the LCT model) the set D_n is weakly separated by some feature φ_n. Next, we consider the scheme p_n. According to this scheme, it has $n+1$ of the final paths. Due to the fact that D_n at each step is weakly separated, every such path contains at least one pair of the primary IS of general view (1). It is also obvious that different end paths in p_n do not have common pairs from the sample (1). Therefore, we can conclude that the scheme (predicate) p_n divides IS (based on the basic branching criterion introduced by the current method of classification tree) in $n+1$ non-empty parts (subsets) that do not intersect. Since in the primary IS in total there are m training pairs, the scheme p_{m-1} (or a predicate with a smaller number) completely separates the primary IS, that is p_{m-1} fully recognizes the sample.

Thereby, if on each $n-$ th step the selected elementary feature φ_n weakly separates the set D_n, then in this case the LCT construction process coincides with the primary IS and ends in no more than in $m-1$ steps, where $m-$ the number of all training pairs of primary IS.

Let us notice that the condition of weak separation of classes for primary IS is relatively weak – therefore, it provides low convergence of the construction procedure for the classification tree. Thus, it is essential to consider the convergence of the process under more vital conditions. Therefore, we will assume that we are dealing with a case where IS contains information about two classes (patterns)

H_0 and H_1, and IS itself has a deterministic nature. Let $n_j -$ is a number of training pairs $(x_i, f_R(x_i))$ within primary IS, which satisfy the ratio $f_R(x_i) = j, (j = 0,1)$, and for simplification and certainty, we put it as $n_0 \geq n_1$.

Having fixed $f_R(x) \equiv 0$, we obtain some generalized feature (scheme) f_0, which approximates (in whole or in part) the primary IS. Obviously, in this case (that is in a situation where the choice of any elementary feature has not yet been made φ_n), generalized feature (scheme) f_0 is the best approximation of primary IS. Further the value n_1 we call the unconditional number of errors in the primary IS.

Let in the first step of building a classification tree some elementary features has been selected (arbitrarily) φ_1 – and this feature will break the initial sample into two parts (subsets) H_0 and H_1, where $H_j -$ is the set of all pairs $(x_i, f_R(x_i))$ of primary IS, for which the ratio is satisfied $f_1(x_i) = j, (j = 0,1)$.

Let $n_m^j -$ the set of all pairs $(x_i, f_R(x_i))$ from sample $H_j, (j = 0,1)$, for which the relation is fulfilled $f_R(x_i) = m, (m = 0,1)$. Feature φ_1 can be considered a generalized feature f_1 (scheme), which is built on the first step of the LCT construction process.

Let us enter the value $\rho = \max(n_0^0, n_1^0) + \max(n_0^1, n_1^1)$, which is the number of correct answers (classifications) that are realized by a generalized feature f_1, and accordingly, the value n_0 is the number of correct answers (classifications), which are realized by a generalized feature f_0.

By the number of correct answers we mean the number of those training pairs $(x_i, f_R(x_i))$ in the initial training sample of the type (1), for which the relation of equality is fulfilled $f_R(x_i) = f_1(x_i)$.

Since $n_0^0 + n_0^1 = n_0$ and $n_1^0 + n_1^1 = n_1$, then we will have the following:

$$\rho = \max(n_0^0, n_1^0) + \max(n_0^1, n_1^1) \geq n_0. \tag{2}$$

Therefore, when choosing a feature φ_1 the number of correct answers at least does not decrease. The number of errors given by the generalized algorithm f_1, will be equal to:

$$m - n = n_1 - (\rho - n_0) \leq n_1. \tag{3}$$

Let us note that (3) follows from (2). Let us enter the value $\lambda_1 = \dfrac{n_1}{m - \rho}$ and call it the quality of elementary feature φ_1 relating to primary IS, similarly we determine λ_n of feature φ_n relating to primary IS $(n = 1,2,3,...)$.

With the power of some constructed generalized feature (GF) or set GF (for a fixed step of ACT scheme) we will call the number of training pairs $(x_i, f_R(x_i))$ of primary IS that look like (1), which is approximated (correctly classified) by this generalized feature (sequence of generalized features).

It is important for ACT schemes that at step-by-step division of IS on two samples H_0 and H_1 (and so on) part of the sample will be completely covered by the current classification algorithm (generalized feature or their set) - that is, we will have a case of strong separation of array IS classes. Therefore, we can assume that the complexity of the final ACT scheme (the total number of steps to build a tree) largely depends on the procedure of initial evaluation and selection of a set of independent classification algorithms a_i, their initial parameters, parameters of set GF f_i, which they generate for each step of the ACT scheme.

Then, for the ACT scheme, it is essential to consider the general complexity of the procedure for constructing a classification tree in the condition of weak separation of the primary IS classes. A single GF is generated with the power of one unit per verticle of the tree. Under high separation

conditions, when a problem and practical feasibility have not set the number limit on GF and its power, it is possible to build them.

In the first stage, we consider a case of weak class division with restrictions on the GF sets being built by the ACT scheme. Let us note that the procedure for constructing an algorithmic tree has certain features in terms of step-by-step approximation of the primary IS by sequence GF. Let at every step of the construction of some ACT model, select for work one fixed classification algorithm from a set of selected algorithms $(a_1, a_2, ..., a_n)$. Moreover, the classification tree can be built by one algorithm and sequence GF, which he is generating.

Therefore, after fulfilling the n steps of the classification tree constructing procedure the ACT structure represents some scheme s_n (generalized second-order feature, which is built from a set of GF synthesized by classification algorithms), which is the most effective approximation of the primary IS of general view (1) a set of independent classification algorithms and their GF. In particular s_n will represent some tree-like scheme (GFT structure), which consists of n vertices, that is, in the design of the scheme s_n enters only n classification algorithms (GF – conditioned that for each step of the tree constructing procedure no more than one generalized feature of the minimum power per unit is generated) from the initial set.

5. Experiments and results

In the next stage for the LCT structure let us make an assumption – quality λ_n of elementary features φ_n relative to the array of primary IS not less than some number y, where $y > 1$.

Let us analyze the complexity of the procedure for classification tree construction under this condition $(y > 1)$, to do this, let us estimate the number of steps for which this process (procedure) will implement full recognition of the initial training sample array.

To be certain, let us consider the following scheme of classification tree construction (Fig. 2).

Let n_1 – an unconditional number of errors within primary IS. Elementary feature φ_1^1 separated IS in two samples: H_0 and H_1. Let h_0 and h_1 accordingly is an unconditional number of errors in the samples H_0 and H_1. Feature φ_1^2 separate the set H_0 in two sets H_{00} and H_{01}. Let h_{00} and h_{01} - is an unconditional number of errors in the samples H_{00} and H_{01}. Similarly, we define sets H_{10}, H_{11} and quantities h_{10} and h_{11} for the elementary feature φ_2^3.
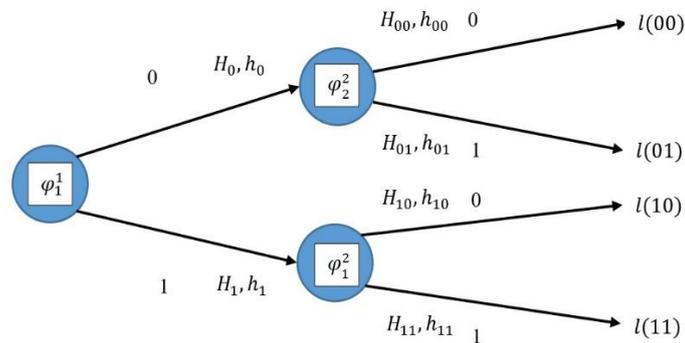

Figure 2: Scheme of division into subsets in the structure of the classification tree.

From the initial condition $(y > 1)$ it follows:

$$\begin{cases} h_0 + h_1 \le \dfrac{1}{y} * n_1 \\[2mm] h_{00} + h_{01} \le \dfrac{1}{y} * h_0 \;. \\[2mm] h_{10} + h_{11} \le \dfrac{1}{y} * h_1 \end{cases} \qquad (4)$$

From (4) we receive the following:

$$h_{00} + h_{01} + h_{10} + h_{11} \le \frac{1}{y^2} * n_1 \;. \qquad (5)$$

Let us make the following assumptions in this regard: $h_0 \ge 1$, $h_1 \ge 1$, $h_{00} \ge 1$, $h_{01} \ge 1$, $h_{10} \ge 1$ and $h_{11} \ge 1$. From here we will have the following:

$$2^1 \le \frac{1}{y} * n_1, \; 2^2 \le \frac{1}{y^2} * n_1 \;. \qquad (6)$$

Similarly for the set of features $\varphi_1^i, \varphi_2^i, \ldots$, located on $i-$ th tier of the logical tree, we will have:

$$2^i \le \frac{1}{y^i} * n_1 \; \text{ or } \; (2y)^i \le n_1 \;. \qquad (7)$$

Hence we can conclude that the process of constructing a classification tree will continue until there the structure of will have m tiers (levels), where m has the following form:

$$m = R(\frac{\log_2 n_1}{1 + \log_2 y}) \;. \qquad (8)$$

Under $R(x)$ means rounding the number x to the nearest integer, which exceeds x. For example $Q(1.2) = 2, Q(3.7) = 4, Q(4.1) = 5$.

Hence the classification tree, which has m full tiers (that is, the case when on $i-$ th tier there are vertices), has vertices – thus recognizing the primary IS with condition that $(y > 1)$ by means of full LCT takes place no more than in steps, where m is calculated using an expression (8).

For the case of structure ACT, we can conclude that the sequence of constructed schemes $s_1, s_2, \ldots, s_j$ (generalized features of the second order) coincides with the primary IS with the view (1), not more than in M steps (where $M-$ total power of primary IS), even if at each step only one GF is generated, while the power of each is not more than one unit.

Some classification algorithm that will be selected (fixed) on $n-$s th step in the procedure of building the ACT model (to generate the appropriate GF), let us denote with a_n, and it is clear, that this algorithm a_n corresponds to some scheme s_n, which consists of algorithms $a_1, a_2, \ldots, a_{n-1}$ and ends with this attribute (vertice of classification tree - ACT model). For example, on Fig. 3 some ACT model is shown, in which a fixed scheme s_2 (vertice of the classification tree under construction) corresponds to the sequence of steps (schemes) $\{s_1\}$, and scheme s_M – sequential path $\{s_1, s_2, \ldots, s_{M-1}\}$.

Therefore, for the ACT model it is possible to make the following conclusion: scheme s_n (in the structure of the classification tree) divides IS on n non-empty parts (subsets) that do not intersect, and because in the primary IS in total there are M training pairs, that is why the scheme s_M will completely divide (approximates) the primary IS (that fully recognizes the sample at the condition that it generates one GF at each step with the power of one unit). So, if on every $n-$ th step of the ACT construction scheme the generated GF (selected by classification algorithm a_n) weakly separates the set of primary IS, in this case, the process of classification tree construction coincides with the primary IS and finishes not more than in M steps, where $M-$ is the quantity of all training

pairs of primary IS. In the following stage of research, it is important to consider the case of strong separation of classes in primary IS, when there are no set restrictions on algorithms a_i regarding generation GF (power of constructed GF is limited only by the practical possibility of the classification algorithm itself a_i and structural parameters of IS). Let with $P(f_j)$ we denote the total power (approximation ability) of corresponding GF f_j, $(1 \leq j \leq s)$, where $s-$ is the quality of GF in constructed ACT scheme.
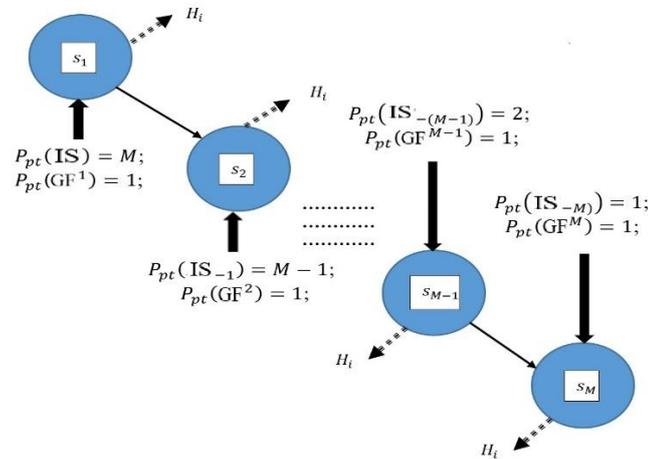

Figure 3: Example of ACT structure with GF as vertices.

Further on some step $r, (1 \leq r \leq M)$ of the ACT scheme, a sequence of generalized features is constructed $f_1, ..., f_r$ with their corresponding values $P(f_i) = z_i$, where $(1 \leq z \leq M), (1 \leq i \leq r)$, $M-$ general power IS, and among them there are values $z^{\max}$ and $z^{\min}$, which are for them, respectively, the maximum and minimum (relatively to the current step of ACT scheme). Then in this case the ACT scheme (model) will be constructed in t steps, where the value t is determined by the ratio (9).

$$t \leq 2 * \frac{P_{pt}(IS)}{z^{\max} + z^{\min}} = \frac{2M}{z^{\max} + z^{\min}}. \tag{9}$$

We notice that in the case when by the condition of the practical problem the power restrictions of synthesized GF are imposed on the ACT scheme under construction (not exceeding the appropriate value P) – classification tree scheme (ACT model) will be built in t steps, where the value t is determined by the ratio (10).

$$t \leq \frac{M}{P}. \tag{10}$$

At strict restrictions of the ACT scheme on one generated GF (where according to the condition $P(f_j) = 1, (1 \leq i \leq t)$, that is in case of weak classes separation of the current problem, the classification tree scheme (model ACT) will be built in t steps, where the value $t \leq M$.

In the next stage, for simplification, let us take the classification problems for which sets of structures LCT/ACT have been built from researches [19-24,30]. The initial parameters of these practical problems are presented in the Table 1. So in IS, the information regarding separation into two classes has been represented. At the examination stage, the constructed classification system should effectively recognize objects of unknown classification concerning these two classes. Let us note that the training and test sample was automatically checked for correctness at the initial stage. (searching and deleting the identical objects of different affiliations - errors of the first and second type).

Here (in Table 2) is represented the evaluation of the constructed structures of classification trees (LCT/ACT) of problems from Table 1. To assess the quality of constructed classifiers (classification schemes) we used an integrated feature of the quality of the classification tree Q_{Main} from resear [30].

Convergence of the structure synthesis LCT/ACT procedure is estimated on the basis of quantitative features – the total number of iterations S_{Main} and the number of tiers in the classification tree structure L_{Kol} .

Table 1
Initial parameters of classification problems

Type of classification problem	The dimension of the feature space N	The power of data array of the primary IS – M	The total number of classes by data splitting IS – l	Relation of objects of different classes IS – $(H_1 / H_2 / ... / H_i)$
The problem of geological data classification (Z1)	22	1250	2	756/494
The problem of chemical analysis of the quality of hydrocarbon fuels (Z2)	14	4863	6	823/648/1412/918/583/764
The problem of classification of flood situations in the Tisza river basin of Transcarpathian region (Z3)	18	6118	3	76/108/5934
The problem of classification of flood situations in the Uzh river basin of Transcarpathian region (observation post №1) (Z4)	18	4252	3	73/102/4107
The problem of classification of flood situations in the Uzh river basin of Transcarpathian region (observation post №2) (Z5)	18	4139	3	68/97/3974

Integrated feature of the classification tree quality Q_{Main} displays the basic parameters (characteristics) of classification trees and can be used as an optimality criterion in the evaluation procedure of an arbitrary tree-like recognition scheme [20]. Let us note that the main idea of classification tree methods based on autonomous algorithms in their structure lies in a step-by-step approximation by the selected set of algorithms the data set of primary IS [22].

Table 2

Comparative table of structure classification schemes LCT/ACT

№ of problem	Method of synthesis of classification tree structure	Integral feature of model quality Q_{Main}	Convergence of the classification tree structure (number of iterations) S_{Main}	Number of tiers in structure LCT/ACT L_{Kol}
Z1	Method of complete LCT based on the selection of elementary features	0,004789	79	22
Z1	LCT model with a one-time assessment of the importance of the features	0,002263	102	16
Z2	Limited LCT construction method	0,003244	91	17
Z2	Algorithmic tree method (type I)	0,005119	46	9
Z3	Algorithmic tree method (type II)	0,002941	72	15
Z3	Method of extensive selection of features (step-by-step assessment)	0,003612	84	13
Z4	Algorithm tree (type I)	0,005054	43	10
Z5	Algorithm tree (type II)	0,002813	75	16

The obtained structures of classification trees (ACT/LCT models) are from one side are characterized by high versatility in terms of practical problems and relatively compact structure of the

model itself, but from the other side it requires significant hardware costs for storing generalized features and initial assessment of the fixed classification algorithms quality according to data of IS comparing to the neural network concept [31-35]. Therefore in comparison with the ACT concept LCT method has a high speed of classification schemes, relatively insignificant hardware costs for storage and operation of the tree structure itself, and high quality of discrete objects classification.

Let us note that all ACT / LCT models and ACT / LCT structures were built in the software application "Orion III". Based on the classification tree method and the principle of modularity, Uzhhorod National University has developed a software application "Orion III" to generate autonomous recognition systems. The algorithmic library of the system has 15 recognition algorithms, including algorithmic implementations of both the LCT structure and the ACT structure [23].

From Table 2 we can see that the methods of algorithm trees (of two types) have shown a high rate of convergence for constructing the classification tree structure on represented IS compared to the LCT schemes. Consider that the first type of the ACT structure shows a good result in terms of structural complexity (number of tiers, vertices, generalized features) of constructed classification model in comparison with logical classification trees and the tree of algorithms of the second type. In general, we can conclude about the rapid convergence of ACT structures compared to LCT models and advantage due to this the structural complexity of the constructed classification tree and the informational capacity of generalized feature sets.

6. Conclusion

Therefore, taking into consideration all the above in the research, we can assume the following points:

On the condition of weak class separation in the LCT case, if on every $n-$ th step the selected elementary feature φ_n weakly separates the set (subset) of objects of primary IS, then, in this case, the process of classification tree construction coincides relating to the primary IS and terminates no more than in $m-1$ steps, where $m-$ the quantity of all training pairs of primary IS.

Classification tree (of LCT structure) on condition of strong classes separation of primary IS sets of objects, which has m full tiers, levels (that is the case, when on $i-$ th tier vertices are located), has vertices – so array recognition of primary IS on condition that $(y > 1)$ by means of full LCT takes place no more than in steps, where m is calculated by means of expression $m = R(\dfrac{\log_2 n_1}{1 + \log_2 y})$.

A general number of all final vertices of logical structure (sheets of recognition tree) of constructed classification scheme will unambiguously determine the final power of classification tree method scheme (models LCT/ACT).

Power of some GF (the set of constructed GF) for the fixed step of the ACT method scheme is considered the general quantity of training pairs $(x_i, f_R(x_i))$ of primary IS (a subset of primary IS) that look like (1), which are approximated (correctly classified) by given generalized feature (sequence of generalized features).

In case of weak class separation of primary IS for the ACT scheme the process of classification tree construction coincides relating to IS data set and terminates in no more than M steps, where $M-$ is the quantity of all primary IS training pairs.

In the case of strong classes, separation of primary IS for the ACT scheme, when the power of constructed GF (or set of GF) is limited only by the practical possibility of the classification algorithm itself a_i and initial parameters of IS, the ACT scheme (model) will be constructed in t steps, where the value t is defined by the ratio (9).

7. Acknowledgments

8. References

[1] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Berlin, Springer, 2008.

[2] J.R. Quinlan, Induction of Decision Trees, Machine Learning 1 (1986) 81-10.

[3] L.L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and regression trees, Boca Raton, Chapman and Hall/CRC, 1984.

[4] M. Lupei, A. Mitsa, V. Repariuk, V. Sharkan, Identification of authorship of Ukrainian-language texts of journalistic style using neural networks, Eastern-European Journal of Enterprise Technologies 1 (2 (103)) (2020) 30-36. doi: 10.15587/1729-4061.2020.195041.

[5] M. Jafari, H. Xu, Intelligent Control for Unmanned Aerial Systems with System Uncertainties and Disturbances Using Artificial Neural Network, Drones 3 (2018) 24-36.

[6] M. Miyakawa, Criteria for selecting a variable in the construction of efficient decision trees, IEEE Transactions on Computers 38(1) (1989) 130-141.

[7] H. Koskimaki, I. Juutilainen, P. Laurinen, J. Roning, Two-level clustering approach to training data instance selection: a case study for the steel industry, in: Proceedings of the International Joint Conference on Neural Networks, IJCNN 2008, part of the IEEE World Congress on Computational Intelligence, IEEE, Los Alamitos. (2008) 3044-3049. doi: 10.1109/ijcnn.2008.4634228.

[8] S.F. Jaman, M. Anshari, Facebook as marketing tools for organizations: Knowledge management analysis, in: Dynamic perspectives on globalization and sustainable business in Asia, IGI Global, Hershey. (2019) 92-105. doi: 10.4018/978-1-5225-7095-0.ch007.

[9] W. Sirichotedumrong, T. Maekawa, Y. Kinoshita, H. Kiya. Privacy-preserving deep neural networks with pixel-based image encryption considering data augmentation in the encrypted domain, in: Proceedings of the 26th IEEE International Conference on Image Processing, ICIP 2019, IEEE, Los Alamitos. (2019) 674-678. doi: 10.48550/arXiv.1905.01827.

[10] R.L. De Mántaras, A distance-based attribute selection measure for decision tree induction, Machine learning 6(1) (1991) 81-92.

[11] K. Karimi, H. Hamilton, Generation and Interpretation of Temporal Decision Rules, International Journal of Computer Information Systems and Industrial Management Applications 3 (2011) 314-323.

[12] B. Kamiński, M. Jakubczyk, P. Szufel, A framework for sensitivity analysis of decision trees, Central European Journal of Operations Research 26 (1) (2017) 135-159.

[13] H. Deng, G. Runger, E. Tuv, Bias of importance measures for multi-valued attributes and solutions, in: Proceedings of the 21st International Conference on Artificial Neural Networks, volume 2 of ICANN 2011, Springer-Verlag, Berlin. (2011) 293-300. doi: 0.1007/978-3-642-21738-8_38.

[14] S.A. Subbotin, Construction of decision trees for the case of low-information features, Radio Electronics, Computer Science, Control 1 (2019) 121-130.

[15] M. Jafari, H. Xu. Intelligent Control for Unmanned Aerial Systems with System Uncertainties and Disturbances Using Artificial Neural Network. Drones. 3 (2018) 24–36.

[16] A. Painsky, S. Rosset, Cross-validated variable selection in tree-based methods improves predictive performance, IEEE Transactions on Pattern Analysis and Machine Intelligence 39(11) (2017) 2142-2153. doi:10.1109/tpami.2016.2636831.

[17] D. Imamovic, E. Babovic, N. Bijedic, Prediction of mortality in patients with cardiovascular disease using data mining methods, in: Proceedings of the 19th International Symposium INFOTEH-JAHORINA, INFOTEH 2020, IEEE, Los Alamitos. (2020) 1-4.

[18] S.B. Kotsiantis, Supervised Machine Learning: A Review of Classification Techniques,

Informatica 31 (2007) 249-268.

[19] Y.I. Zhuravlev, V.V. Nikiforov, Recognition algorithms based on the calculation of estimates, Cybernetics 3 (1971) 1-11.

[20] Y.A. Vasilenko, E.Y. Vasilenko, I.F. Povkhan, Branched feature selection method in mathematical modeling of multi-level image recognition systems, Artificial Intelligence 7 (2003) 246-249.

[21] I. Povkhan, 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), 2608 (2020) 843-857.

[22] Y.A. Vasilenko, E.Y. Vasilenko, I.F. Povkhan, Conceptual basis of image recognition systems based on the branched feature selection method, European Journal of Enterprise Technologies 7(1) (2004) 13-15.

[23] I. Povkhan, M. Lupei, The algorithmic classification trees, in: Proceedings of the IEEE Third International Conference on Data Stream Mining & Processing, DSMP 2020, IEEE, Los Alamitos, (2020) 37-44.

[24] I. Povkhan, M. Lupei, M. Kliap, V. Laver, The issue of efficient generation of generalized features in algorithmic classification tree methods, , in: Proceedings of the International Conference on Data Stream Mining and Processing, DSMP 2020, Springer, Cham. (2020) 98-113.

[25] I. Povkhan, Classification models of flood-related events based on algorithmic trees, Eastern-European Journal of Enterprise Technologies 6(4) (2020) 58-68. doi: https://doi.org/10.15587/1729-4061.2020.219525.

[26] J. Rabcan, V. Levashenko, E. Zaitseva, M. Kvassay, S. Subbotin, Application of Fuzzy Decision Tree for Signal Classification, IEEE Transactions on Industrial Informatics 15(10) (2019) 5425-5434. doi: 10.1109/tii.2019.2904845.

[27] P.E. Utgoff, Incremental induction of decision trees, Machine learning 4(2) (1989) 161-186. doi:10.1023/A:1022699900025.

[28] L. Hyafil, R.L. Rivest, Constructing optimal binary decision trees is NP-complete, Information Processing Letters 5(1) (1976) 15-17.

[29] H. Wang, M. Hong, Online ad effectiveness evaluation with a two-stage method using a Gaussian filter and decision tree approach, Electronic Commerce Research and Applications 1(35) (2019) Article 100852. doi: 10.1016/j.elerap.2019.100852.

[30] I.L. Kaftannikov, A.V. Parasich, Decision Tree's Features of Application in Classification Problems, Bulletin of the South Ural State University, Ser. Computer Technologies, Automatic Control, Radio Electronics 15(3) (2015) 26-32. doi: 10.14529/ctcr150304.

[31] I.F. Povhan, Logical recognition tree construction on the basis a step-to-step elementary attribute selection, Radio Electronics, Computer Science, Control 2 (2020) 95-106.

[32] S.L. Lin, H.W. Huang, Improving Deep Learning for Forecasting Accuracy in Financial Data, Discrete Dynamics in Nature and Society Volume 2020 (2020) Article ID 5803407. doi: 10.1155/2020/5803407.

[33] R. Srikant, R. Agrawal, Mining generalized association rules, Future Generation Computer Systems,13(2) (1997) 161-180.

[34] Ameisen E, Building Machine Learning Powered Applications: Going from Idea to Product, O'Reilly Media, California. (2020).

[35] M. Sewak, Deep Reinforcement Learning, Frontiers of Artificial Intelligence, Springer, Berlin, 2020. doi: 10.1007/978-981-13-8285-7.

[36] J.B. Harley, D. Sparkman, Machine learning and NDE: Past, present, and future, AIP Conference Proceedings, 2102(1) (2019). doi:10.1063/1.5099819.