The modified algorithm tree method in the geological data classification problem

Igor Povkhan [1] , Oksana Mulesa [1] , Olena Melnyk [1] , Vasyl Morokhovych [1]

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

Abstract

This study presents the development of an advanced algorithmic tree synthesis method predicated on a set configuration of initial data for the task of geological data recognition. The devised classification tree algorithm of the second type demonstrates precise classification of the complete training dataset, adhering to the established classification schema. It boasts high interpretability, a straightforward structure, and incorporates autonomous algorithms for classification and scheme recognition as vertices within a graphical framework. The refined construction methodology for the tree algorithm facilitates handling substantial volumes of discrete data across diverse categories, ensuring remarkable accuracy of the classification schema. Moreover, it judiciously utilises hardware resources during the creation of the definitive classification schema and supports the development of models with specified accuracy levels. The paper advocates a novel synthesis approach for recognition algorithms, drawing on a repository of extant algorithms and theoretical recognition methods. Employing the proposed second type tree algorithm, a suite of models has been constructed that adeptly classifies extensive arrays of geological data. The constructed models of classification trees have verified the absence of errors in both training and testing datasets, substantiating the efficacy of the second type tree method algorithm.

Keywords

Algorithmic tree, classifier, pattern recognition, feature, initial sample.

1. Introduction

Classification and image recognition represent critical problem domains within the sphere of artificial intelligence, notable for their extensive diversity, varying degrees of structural complexity, and significant applicability across numerous sectors of human economic and social endeavours. In disciplines such as geology, where the challenges of classification are tackled through sophisticated information systems, the importance and intensity of research in this area are well-documented [1-10]. These classification challenges demand the development and decomposition of mathematical models tailored to the specific systems under study. Presently, the field of artificial intelligence lacks a universally applicable approach capable of addressing the full spectrum of these complex problems. However, several broadly applicable theories and methodologies have emerged, with neural networks being particularly prominent due to their versatility in addressing a wide array of classification challenges [11-14]. In practical scenarios, specifically configured artificial neural networks often outperform traditional algorithms and established decision tree models, such as gradient boosting methods, especially in tasks involving unstructured data, discrete image sets, or textual content. Conversely, when dealing with structured datasets comprising large volumes of massive discrete data, which exhibit diverse feature spaces, decision tree-based methods and algorithms exhibit distinct advantages [15]. Generally, classification tree methodologies facilitate effective data processing across various magnitudes, presenting the input information in its inherent form. Numerous contemporary strategies and concepts are focused on developing recognition systems (RS) and classifications using logical/algorithmic classification tree models (LCT/ACT structures). The growing interest in tree-like graph-schematic representations of classifiers is driven by their numerous

CMIS-2024: Seventh International Workshop on Computer Modeling and Intelligent Systems, May 3, 2024, Zaporizhzhia, Ukraine

igor.povkhan@uzhnu.edu.ua (I. Povkhan); oksana.mulesa@uzhnu.edu.ua (O. Mulesa); olena.melnyk@uzhnu.edu.ua (O. Melnyk); morv77@ukr.net (V. Morokhovych)

0000-0002-1681-3466 (I. Povkhan); 0000-0002-6117-5846 (O. Mulesa); 0000-0001-7340-8451 (O. Melnyk); 0000-0002-4939-6566 (V. Morokhovych)

advantageous properties [16]. One promising area of application for the classification tree model, specifically within the realm of algorithmic trees, is in the classification of geological informations [22].

2. Formal problem statement

Let be $H_1, H_1, \dots, H_k$ the system of classes (images) defined on the set G consisting of objects x_i, $(i = 1, \dots, m)$. The nature of the division of the set G into the corresponding classes is specified using the following training sample (TS):

$$\Big((x_1, f_R(x_1)), (x_2, f_R(x_2)), \dots, (x_m, f_R(x_m))\Big). \quad (1)$$

Let us note that here $x_h \in G, f_R(x) \in \{0,1, \dots, k-1\}, (h = 1,2, \dots, m)$, $k -$ the number of TS classes, $m -$ the total number of TS objects, and $f_R(x)$ is some finitely significant function that determines the division of the set G into corresponding images. The ratio $f_R(x_h) = l$, $(l = 0,1, \dots, k-1)$ means that $x_h \in H_l$. We note that each TS of the form (1) can be according (with the help of some algorithm or representation method) to a wholly defined LCT, which matches $f_R(x_h)$ the objects of TS (1) with the value of the function x_h, $(h = 1, \dots, m)$, which specifies the partition R on the set G. Therefore, the task will be to build a structure of the classification tree (LCT/ACT), the structure of which would be optimal $f_R(x_j) \to opt$ with the initial data of TS.

3. Literature review

The current study delves into the theory of fixed-type decision trees, focusing on algorithm trees and the classification of discrete objects [14, 23, 25]. Notably, research [20] underscores that the classification rules and decision schemes, derived from any branching feature selection method or algorithm, manifest a tree-like logical structure. A typical decision tree classifier comprises an organized sequence of nodes, features, and attributes structured into layers or levels, each established during a specific phase of the classification tree synthesis [15]. A significant challenge identified in [18] is the effective construction of recognition tree structures, which can take the form of tree-like structures or algorithm graphs (ACT structures). Consequently, decision tree methodologies facilitate the creation of innovative classifiers based on a modular principle, utilizing well-known recognition algorithms [19-21]. The study [14] explores fundamental issues related to the generation of decision tree structures, particularly when features are low-informative, including their sets and combinations. Within the sphere of intelligent data analysis, the invariant capacity of LCT/ACT structures to execute one-dimensional branching allows for the analysis of the influence, importance, and quality of individual variables. This capability is essential for managing different types of variables as predicate sets. The persistent challenge with decision tree methods and structures is evaluating the quality and efficiency of the branches (generalized features) that serve as autonomous classification algorithms [15].Logical decision tree classification methods are prevalently employed in intelligent data analysis, aiming to synthesize operational models that predict the value of a target variable based on an initial set of data formatted as a structured training sample [19]. From an applied perspective, numerous methods and algorithms grounded in the decision tree concept are utilized for classification tasks; however, C4.5/C5.0 and CART have emerged as particularly popular. The C4.5/C5.0 methods employ a theoretical-informational criterion for node or vertex selection, whereas the CART algorithm relies on the Gini index, which assesses the relative distances between class distributions within the metric of the training sample [20, 21]. The set of methods and algorithms for branching feature selection (ACT structures) is based on optimally approximating the initial training set using a ranked series of classification algorithms [22]. A key issue within LCT/ACT methods, as discussed in [23], involves choosing an effective branching criterion—that is, selecting nodes, attributes, and features of discrete objects for LCT schemes and algorithms for ACT. These foundational issues are thoroughly examined in another paper [24], which addresses the qualitative evaluation and informativeness of individual discrete features, their sets, and fixed combinations, ultimately supporting the efficient implementation of a branching mechanism within the logical/algorithmic tree structure. Concerns regarding the convergence of the classification tree construction process, including the selection of stopping criteria for the synthesis of logical and algorithmic trees, remain significant [25]. The concept of classification trees accommodates the use of not only individual attributes and object features but also their combinations and sets as features, attributes, and nodes of the recognition tree structure. By adopting independent individual recognition algorithms (evaluated using training data) instead of object attributes as branches, a novel ACT structure is realized [21-24]. This research specifically targets the exploration of fixed-type ACT structures within the practical domain.

4. The general second type trees method algorithm

Let the initial TS of the general form (1) be given as a sequence of training pairs of known classification (power m) and some system (set) of independent and autonomous recognition (classification) algorithms for the initial TS $\alpha_1(x), \alpha_2(x), \dots, \alpha_n(x)$. Next, it is necessary to enter the following sets, which represent the breakdown of the data of TS by the corresponding classification algorithms a_i:

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

Note that to simplify explanations, each autonomous classification algorithm $\alpha_i(x)$ generates output values only within the binary set {0,1}, particularly $\alpha_i(x) = 1$ in the case of successful object classification x and $\alpha_i(x) = 0$ in the opposite case.

Note that the system of sets $G_{a_1,\dots,a_i}$ will represent a complete step-by-step division of the set G (with an increase in the size of i the involved classification algorithms), which is implemented by independent algorithms $\alpha_1, \alpha_2, \dots, \alpha_n$. Note that depending on the initial selection of a set of classification algorithms, $\alpha_1, \alpha_2, \dots, \alpha_n$ some of the sets $G_{a_1,\dots,a_i}$ may be empty (in case one or more algorithms are not suitable for approximating the current TS) [21].

At the next stage, we denote by the value $S_{a_1,\dots,a_n}$ the number of occurrences of those training pairs $(x_s, f_R(x_s))$, $(1 \le s \le m)$ which satisfy the basic condition of belonging, in the initial TS $x_s \in G_{a_1,\dots,a_i}$.

Accordingly, by the value $S^j_{a_1,\dots,a_i}$, $(j = 0,1, \dots, k-1)$ we denote the number of occurrences in the TS of those pairs $(x_s, f_R(x_s))$ $(s = 1,2, \dots, m)$, which satisfy the conditions $x_i \in G_{a_1,\dots,a_n}$and $f_R(x_s) = j$.

So, taking into account the above, what was said and by analogy with the methods of selection of sets of elementary features, the following values can be introduced, which should be considered as a certain criterion of branching in the structure of the ACT:

$$\delta_{a_1,\dots,a_i} = \frac{S_{a_1,\dots,a_i}}{m}, \psi^j_{a_1,\dots,a_i} = \frac{S^j_{a_1,\dots,a_i}}{S_{a_1,\dots,a_i}}, \rho_{a_1,\dots,a_i} = \max_j \psi^j_{a_1,\dots,a_i}. \tag{3}$$

Note that if the object $x_s \notin G_{a_1,\dots,a_i}$ is for all $s = 1, \dots, m$, then it is clear that $\delta_{a_1,\dots,a_i} = 0$ and $\psi^j_{a_1,\dots,a_i} = 0$ for $j = 0,1, \dots, k-1$.

Particularly the quantity $\delta_{a_1,\dots,a_i}$ characterizes the frequency of occurrences of members of the sequence $x_1, x_2, \dots, x_m$ (discrete objects) in the set $G_{a_1,\dots,a_i}$, and accordingly, the quantity $\psi^j_{a_1,\dots,a_i}$ characterizes the frequency of belonging to some object x of the image (class) H_j, provided that $x \in G_{a_1,\dots,a_i}$. It should be noted that the given condition is equivalent to the condition that in the sequence of algorithms $a_1, \dots, a_i$ there is such an algorithm a_y that $a_y(x) = 1$. Then the value $\delta_{a_1,\dots,a_i}$ characterizes the information efficiency of recognizing the belonging of some object x to one of the classes $H_0, H_1, \dots, H_{k-1}$ provided that $x \in G_{a_1,\dots,a_i}$.

At the next stage, a fundamental question arises again regarding the object's belonging to x classes $H_0, H_1, \dots, H_{k-1}$ (the question of forming a classification rule). It is clear that the object should be assigned x to the class H_j for which a simple relation is fulfilled:

$$\rho_{a_1,\dots,a_i} = \psi^j_{a_1,\dots,a_i}. \tag{4}$$

Note that here $\{0 \le j \le k-1\}$, and relation (4) represents a certain classification rule, and it is clear that the greater the value of the value of $\rho_{a_1,\dots,a_i}$, the higher the effectiveness of the rule.

Since the only information that represents the partitioning of images $H_0, H_1, \dots, H_{k-1}$ is the initial TS, then the class H_j is understood as the set of all training pairs $(x_s, f_R(x_s))$of TS that satisfy the ratio $f_R(x_i) = j$, that is, the condition of belonging.

It is clear that an algorithmic tree is not the only possible construction (structure) classification algorithm that can be organized in the form of a tree-like recognition model (several types of such structures can be proposed). Next, we will propose one scheme for organizing a set of classification and recognition algorithms $(\alpha_1, \alpha_2, \dots, \alpha_m)$ in the form of an ACT model, which we will call an algorithmic classification tree of the second type.

Let us note that a set of autonomous classification and recognition algorithms $(\alpha_1, \alpha_2, \dots, \alpha_m)$can act as a set of primary features (attributes) for an arbitrary discrete object of x_i some initial TS of the general form (1). Moreover, with regard to a fixed discrete object x_i of the initial TS, information about the appearance of the generalized feature (GF), which is built by the current classification algorithm, and information about the general possibility of recognizing this discrete object (presence of failure, incorrect classification, impossibility) will be necessary for this ACT scheme GF to describe this object, etc.).

Therefore, let each training pair $\left(x_i, f_R(x_i)\right)$ of the TS correspond to its training pair of the following form:

$$\left(x_i(\varphi(\alpha_1), \varphi(\alpha_2), \dots, \varphi(\alpha_m)), f_R(x_i)\right), \varphi\left(\alpha_j\right) \in \{0,1\} \text{ where.} \quad (5)$$

Moreover $\varphi\left(\alpha_j\right) = 1$, if this discrete object is approximated by some GF f_l, which is built by the α_j set classification algorithm $(\alpha_1, \alpha_2, \dots, \alpha_m)$ at the corresponding stage of ACT generation. Similarly $\varphi\left(\alpha_j\right) = 0$, if for a given discrete object the algorithm α_j did not build a suitable GF (which would ensure its approximation, classification), this situation also includes failures and classification errors (errors of the first and second kind).

By the algorithmic tree of the second type, we will understand some tree-like construction, the general view of which is presented in (Fig. 1), at the vertices of which there are appropriate labels (classification and recognition algorithms, α_j as well as sets of GFs that they generate at a specific step of the ACT construction procedure). Note that the logical tree of this construction belongs to the class of regular logical trees of full complexity (this logical tree will be equivalent to a logical function of four arguments, the arguments of which take values from the set $\{0,1\}$).

The following basic ACT scheme for synthesizing a tree of algorithms of the second type based on a branched selection of generalized features allows us to build ACT structures of arbitrary complexity and efficiency (Fig. 1).

Stage of initial selection and evaluation of independent classification algorithms. At the initial stage, it is necessary to select and evaluate the basic set (fixed set) of classification and recognition algorithms $(\alpha_1, \alpha_2, \dots, \alpha_m)$ from the initial algorithm library. Note that this procedure is performed based on the selected (fixed) performance criterion, followed by ranking – interactive or randomly. The performance criterion may vary depending on the type of act structure that is being built and cannot be changed during the classification tree synthesis process. The set of autonomous algorithms $(\alpha_1, \alpha_2, \dots, \alpha_m)$, as well as their total number in the set, are selected depending on the applied aspects of the problem and can be selected even on the basis of a complete search of the algorithm library (of course, with significant losses of hardware resources and processor time). At the initial stage of synthesis of the second type of ACT model, by selecting (ranking) a set of classification algorithms and their total number, the final structural complexity of the algorithm tree can be controlled.

Stage of synthesis of the algorithm tree structure and generalized features. At the next stage, the central task is to build a complete regular classification tree (fixed LCT structure), where the corresponding tiers of the structure contain the selected classification algorithms $(\alpha_1, \alpha_2, \dots, \alpha_m)$, fixed at the first stage of constructing classifier sets.

A special feature of the algorithm tree of the second type is that in the constructed classification tree structure (LCT structure), each vertex has two transitions to the next level, denoted by a value from the binary set $\{0,1\}$. This is why the structure of the algorithm tree is represented using a regular LCT construct. Based on this, all attributes (labels) of the same type (classification algorithms and generated generalized features) are located at each of the levels of this structure. In such a regular classification tree structure, nodes are independent algorithms (classifiers) $(\alpha_1, \alpha_2, \dots, \alpha_m)$. Generalized feature sets (GFs) of f_j are also generated during the synthesis step of the algorithms tree structure. Therefore, we can conclude that the algorithm tree generates a tree of generalized features.

The idea of the second stage of synthesis of the algorithm tree structure (ACT type II model) is the procedure for synthesizing a set of generalized features f_j (vertices of the generalized features tree) based on pre-selected sets of independent classification and recognition algorithms α_i. Note that the total number of GFs f_j generated by the corresponding classification algorithm depends on the initial parameters of the ACT model and synthesis parameters, the specifics of the application problem, and the resource constraints of the classification tree synthesis system.

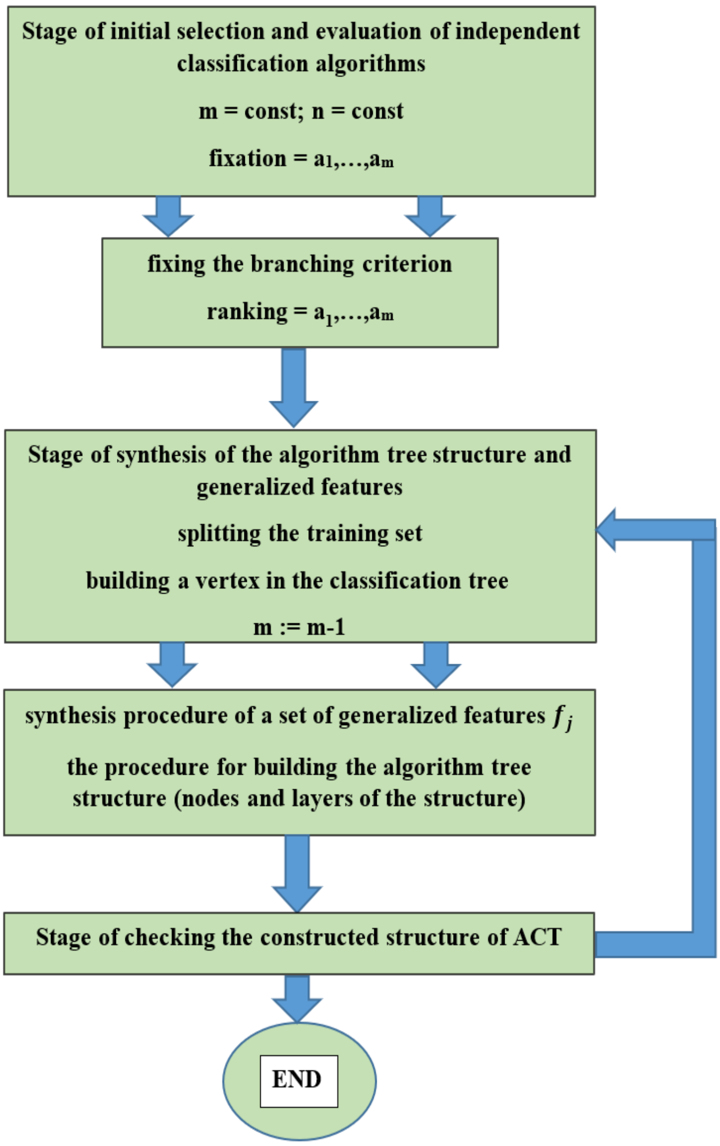

Figure 1: The general block - diagram of the second type tree method

At the end of the second stage, after the formation of a set of synthesized generalized features f_j for a given application problem is completed, they are located in the corresponding nodes, tiers of the tree of algorithms of the second type (the structure of the tree of generalized features is constructed).

Stage of checking the constructed structure of ACT. At the final stage of synthesizing the second type of algorithm tree, you need to check the constructed ACT model. For each element (object) of the test sample, the corresponding values of $\varphi(\alpha_j)$ are calculated. This value is calculated based on a set of previously constructed generalized features - for each node of the corresponding tree level. The constructed generalized features define the corresponding route (bounded classifier) in the structure of the tree of algorithms of the second type. For such a GFs structure, each of the nodes in the algorithm tree, in the event of a possible approximation of an object of unknown classification, increases the corresponding counter of the class belonging to it and leaves it unchanged in the event of a classification error or failure. This procedure allows you to make a final assessment of the effectiveness of the constructed tree of algorithms of the second type.

5. Experiments and results

The experimental validation of the proposed second type algorithm tree construction scheme underscores its capability to tune the complexity and accuracy of the resulting classification tree model. The model comprises various autonomous classification algorithms which, during the modeling process, evolve into a hierarchical structure of generalized features. The selection of an optimal model from the array of constructed Algorithmic Classification Trees (ACTs) for a specific task hinges on evaluating multiple parameters and the effectiveness of the model, which is typically assessed through techniques such as cross-validation against the training set (TS) data. An essential stage in this process involves identifying the most critical parameters of the model, such as the feature space size, the number of vertices, transitions, and algorithms. This step is crucial for estimating the ACT's error relative to the input data set, facilitating comparison, and aiding in the selection of a specific ACT model from the pre-defined ensemble.

Quality criteria of the constructed algorithm trees are paramount and depend on several factors including model error, the robustness of the initial TS data set, the size of the testing sample, and the dimensional characteristics of the problem (e.g., the number of model parameters). At the optimization stage of the constructed ACT model, priority is given to minimizing errors across the training and test datasets for each class defined by the initial conditions of the current applied problem.

A significant ongoing challenge is reducing the complexity and structural pruning of the ACT model. This reduction pertains to the overall count of functions (classifiers) and algorithms within the ACT framework, the total number of vertices (generalized features), and the number of transitions within the structure, as well as optimizing total memory usage and processing time of the information system. Consequently, the defining measure of the quality and efficiency of a constructed model, whether ACT or Logical Classification Trees (LCT), is determined by an overall integral quality indicator:

$$Q_{Main} = \frac{Fr_{All}}{V_{All} \cdot \sum_i p_i} \cdot e^{-\frac{Er_{All}}{M_{All}}}. \tag{6}$$

Note that in formula (6), the set of parameters p_i represents the most important characteristics of the constructed classification tree that is evaluated:

1) Er_{All} – the total number of errors of the ACT model on the data arrays of the initial test and training samples;
2) M_{All} – the total capacity (volume) of data arrays of training and test samples;
3) Fr_{All} – the number of vertices of the obtained ACT model with the resulting values f_R (recognition functions, i.e. leaves of the classification tree);
4) V_{All} – represents the total number of all types of vertices in the structure of the ACT model;
5) O_{Uz} – the total number of generalized features used in the classification tree model;
6) P_{All} – the total number of transitions between vertices in the structure of the constructed classification tree model;
7) N_{Alg} – the total number of different autonomous classification algorithms a_i used in the classification tree model.

Note that this integral indicator of the quality of the ACT model will take values from zero to one. The smaller it is, the worse the quality of the constructed classification tree will be, and the larger the indicator, the better the resulting model will be.

The Orion software complex was developed at the Uzhhorod National University based on classification tree methods to generate autonomous recognition systems. The algorithmic library of the system includes 18 recognition algorithms, among which tree schemes of algorithms of three types are implemented.

The primary task on which the effectiveness of algorithm tree methods was tested was the task of recognizing geological data – the task of separating oil-bearing and water-bearing strata. The initial parameters of this applied problem of geological data classification are presented in (Table 1).

Information about objects of two classes is presented in the TS. At the examination stage, the constructed classification system should effectively recognize objects of unknown classification relative to these two classes. Before starting work, the training sample was automatically checked for correctness - finding and removing errors of the first kind. The system implements a retraining and error correction scheme in the classification tree (REC algorithm).

The training sample of the presented problem consisted of 1342 objects, of which 761 were oil-bearing objects. The effectiveness of the constructed ACT model was evaluated on a test sample of 267 objects. The data from training and test samples were obtained based on geological exploration in the territory of the Transcarpathian region in the period from 2001 to 2019. A fragment of the main results of the above experiments, constructed models of LCT/ACT of various types, are presented in (Table 2).

Table 1

Initial parameters of the classification problem

Description of classes H_i tasks	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 $-H_i/M$
Oil-bearing layers (H_1)	(12/10)	1342	2	761 / 1342
Aquifers (H_2)	(12/10)	1342	2	581 / 1342

(Table 3) presents information on the classification models' generation time, the total number of vertices, and elementary and generalized features on the basic hardware configuration Intel i7-12700H. All constructed schemes of classification trees (structures of LCT/ACT) provided the necessary level of accuracy given by the task condition, speed, and consumption of the system's working memory.

Table 2

Comparison table of built ACT/LCT models for classification of geological data

Classification tree model No	Method of synthesis of classification tree structure	Integral feature of model quality Q_{Main}	The overall indicator of the structural complexity of the classification tree model S_{Main}	The number of errors and failures to classify the LCT/ACT model on the data set Er_{All}
No. 1	The method of full LCT based on the selection of elementary traits (extensive selection of features)	0.004786	121	7

No. 2	The LCT method with a one-time assessment of the importance of features	0.002271	144	12
No. 3	Limited method of construction of LCT	0.003193	9 7	16
No. 4	Algorithmic tree method (type I)	0.005287	52	10
No. 5	Algorithmic tree method (type II)	0.003033	64	8
No. 6	A limited method of building ACT	0.002654	55	14
No. 7	Algorithm tree based on hyperspheres	0.007221	31	6
No. 8	A tree of algorithms based on hyperparallelepipeds	0.004418	54	19
No. 9	A tree of algorithms based on hyperellipses	0.006476	30	8
No. 10	Algorithm tree based on hypercubes	0.006251	37	11

Table 3
General structural parameters of the constructed models of LCT/ACT

	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6	No. 7	No. 8	No. 9	No. 10
Total time of classification tree synthesis T_{All}	34 (s.)	21 (s.)	18 (s.)	65 (s.)	82 (s.)	55 (s.)	47 (s.)	56 (s.)	50 (s.)	98 (s.)
The number of tiers of the LCT/ACT structure R_{All}	12	10	9	26	30	23	21	24	22	34
The total number of attributes / vertices of the LCT/ACT structure V_{All}	102	91	86	234	244	212	198	223	207	219

The total number of elementary / generalized features in the structure of the classification tree O_{el}/O_{Uz}	56 (el.)	72 (el.)	40 (el.)	17 (g.)	41 (g.)	30 (g.)	18 (g.)	47 (g.)	21 (g.)	35 (g.)

Therefore, the algorithmic tree classification method proposed in the paper (second-type ACT methods) was compared with the complete LCT method and the limited method of selection of elementary features and showed a generally acceptable result.

6. Conclusion

The developed models of classification trees (ACT/LCT structures) have successfully met the requirements for quality and speed in geological data classification schemes while maintaining a compact structure (parameter S_{Main}). The sets of independent classification algorithms selected for generating GF groups also demonstrated their effectiveness within the scope of this applied problem. Notably, the models of ACT employing basic geometric classifiers were found to be the most effective upon critical evaluation. Furthermore, the composite ACT structures resulted in a relatively low number of classification errors in both the training and testing datasets. The full model of the second type ACT, based on geometric classifiers, showed promising results (Q_{Main}= 0,003033), largely due to the inclusion of a universal algorithm of hyperspheres in the scheme. In contrast, the structure of the first type ACT exhibited superior quality (Q_{Main}= 0,005287) compared to second type algorithm trees. This superiority is attributed to the more complex construction of the model (S_{Main}=52), which, consequently, required longer generation times. However, it is essential to consider the limitations of the selected geometric classifiers, which may not always provide effective approximation of the TS data. A notable drawback of the ACT models presented, identified during this task, is the relatively high time consumption during the synthesis stage of the classification tree models, especially when compared to the LCT structures. The time difference in constructing the first type of ACT models, which includes a step-by-step assessment of feature informativeness, was nearly 34% greater than that of LCT.

The scientific novelty lies in the fact that for the first time a modified method for constructing algorithm trees based on evaluating and ranking a set of autonomous recognition algorithms for generating a classification tree structure (ACT model) has been proposed.

The practical implications of these findings are significant. The proposed method for constructing ACT models (of the second type) enables the creation of economical and efficient classification models with specified accuracy. This method has been integrated into the algorithm library of the "ORION" system, addressing various applied classification challenges and demonstrating a high degree of versatility across a range of applications. The efficacy of the classification tree models and the associated software have been confirmed through practical applications. Looking ahead, future research could focus on the further development of ACT methods, including the introduction of new types and schemes of classification trees. Additionally, optimizing the software implementations of the proposed ACT method and its practical validation on a variety of precise classification and recognition tasks could provide valuable insights and enhancements to the field.

Acknowledgements

The presented work was carried out within the framework of projects: Expanding Opportunities of High Technologies for Higher Education Institutions (Skills2Scale) - HEI Initiative; Innovative teaching methods to support partnership relations - "InovEduc" - grant project No. CBC01008 of the Norwegian State Fund with the solidarity budget of the Slovak Republic within the framework of the SK08 cross-border cooperation program; Research work "Modeling and prediction of emergency situations in the Carpathian region and the countries of Central Eastern Europe", state registration number of the work - 0106V00285, work category - fundamental research, 01 Fundamental research on the most important problems of natural, social and humanitarian sciences.

References

[1] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer, Berlin, 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, Chapman and Hall/CRC, Boca Raton, 1984.
[4] A. Kintonova, M. Mussaif, G. Gabdreshov, Improvement of iris recognition technology for biometric identification of a person, Eastern-European Journal of Enterprise Technologies 6(2(120)) (2023) 60–69. doi: 10.15587/1729-4061.2022.269948.
[5] Y.V. Bodyanskiy, A.Y. Shafronenko, I.P. Pliss, Credibilistic fuzzy clustering based on evolutionary method of crazy cats, System Research and Information Technologies 3 (2021) 110–119. doi: 10.20535/SRIT.2308-8893.2021.3.09.
[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, IEEE, Los Alamitos, 2008, pp. 3044-3049. doi: 10.1109/ijcnn.2008.4634228.
[8] S.F. Jaman, M. Anshari, Facebook as marketing tools for organizations: Knowledge management analysis, in: S.F. Jaman, M. Anshari, Dynamic perspectives on globalization and sustainable business in Asia, IGI Global, Hershey, 2019, pp. 92-105. doi: 10.4018/978-1-5225-7095-0.ch007.
[9] V.E. Strilets, S.I. Shmatkov, M.L. Ugryumov et al, Methods of machine learning in the problems of system analysis and decision making, Karazin Kharkiv National University, Kharkiv, 2020.
[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, pp. 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. doi: 10.15588/1607-3274-2019-1-12.
[15] A. Shyshatskyi, Complex Methods of Processing Different Data in Intellectual Systems for Decision Support System, International Journal of Advanced Trends in Computer Science and Engineering 9(4) (2020) 5583–5590. doi: 10.30534/ijatcse/2020/206942020.

[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, pp. 1-4. doi:10.1109/INFOTEH48170.2020.9066297.
[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] I. Povkhan, O. Mulesa, O. Melnyk, Y. Bilak, V. Polishchuk, The Problem of Convergence of Classifiers Construction Procedure in the Schemes of Logical and Algorithmic Classification Trees, in: Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems, CMIS-2022, Volume 3137 of CEUR Workshop Proceedings, CEUR-WS, Zaporizhzhia, Ukraine, 2022, pp. 1-13.
[21] I. Povkhan, A constrained method of constructing the logic classification trees on the basis of elementary attribute selection, in: Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems, CMIS-2020, CEUR Workshop Proceedings, Volume 2608, 2020, pp. 843-857.
[22] 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, pp. 37-44.
[23] 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, IEEE, Los Alamitos, 2020, pp. 98-113.
[24] I. Povkhan, Classification models of flood-related events based on algorithmic trees, Eastern-European Journal of Enterprise Technologies, 6(4) (2020) 58-68. doi: 10.15587/1729-4061.2020.219525.
[25] 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.