The Quality Indicators of Decision Tree and Forest Based Models

Sergey Subbotin[1[0000-0001-5814-8268]]

[1]National University "Zaporizhzhia Polytechnic", Zhukovsky str., 64, Zaporizhzhia, 69063, Ukraine
subbotin@zntu.edu.ua

Abstract. The problem of quality model creation for models based on decision trees and forests is considered. The set of indicators characterizing properties of decision trees and forests is proposed. It allows to quantitatively evaluate such properties as diversity, equivalence, retraining, confidence in a decision-making, hierarchy, equifinality, generalization, nonlinearity, robustness, homogeneity, sensitivity to the input signals, plasticity, variability adaptability, symmetry, asymmetry, emergence (integrity), interpretability (logical transparency), learnability, and autonomy as for individually considered single tree model, as for an ensemble of tree models (forest).

Keywords: decision tree, forest, quality, property, model selection

1 Introduction

Automation of decision-making in applied tasks, as a rule, requires the construction of a decision-making model. To solve the problem of decision-making model constructing on the precedents, a wide class of computational intelligence methods has been proposed, including neural networks [1-5], neuro-fuzzy networks [6-9], decision and regression trees [10-17], forests of decision trees [18-22] and etc.

Usually, the quality of such models is characterized by the error function [1, 2]. As a result model is selected from several alternative obtained models, which has the smallest error. Note that for each class of methods, even for the same given training sample of observations, it is possible to obtain a wide range of different models with acceptable accuracy. At the same time, achieving the maximum accuracy (the smallest error) does not guarantee a high level of customer properties of the model.

Earlier in [23-26] author has proposed a set of indicators, applicable to models based on neural and neuro-fuzzy networks. However, most of these indicators are not applicable to the models based on decision trees and forests due to their paradigm difference from network models paradigm. Therefore, it is necessary to develop a quality model for decision trees and forests providing comparability of its indicators with the quality indicators of models based on neural and neuro-fuzzy networks proposed earlier in [23-25].

The properties of models can be affected not only by the structural parameters, but also by the properties of the training sample [27-31]. Therefore, it is necessary to take into account information about the properties of the sample when determining the quantitative indicators characterizing the properties of models.

The aim of this work was to create a quality model for models based on decision trees and forests as a set of quality indicators.

2 Formal problem statement

Let we have a training sample of observations in the form <x, y>, where $x = \{x^s\}$, $x^s=\{x^s_j\}$, y=$\{y^s\}$, s = 1, 2, ..., S, j = 1, 2, ..., N, x^s is a set of input values (descriptive) features of s-th sample instance, x^s_j is a value of j-th input feature of s-th sample instance, y is an output feature value for the s-th instance in a sample, S is a number of instances in the training sample, N is a number of input (descriptive) features characterizing the instances of training sample.

Then the problem of model building for the dependence $y=f(w, x)$ on a sample of observations <x, y> based on the decision tree *tree* consists in identifying the model structure f and the values of its parameters w that provide an acceptable value of the given quality functional of the model $F(x, y, f, w)$ [17].

The problem of model building based on a forest of decision trees, in turn, can be represented as a problem of obtaining a set of models $forest=\{tree_t\}$, $tree_t = f_t(w^t, x)$, that provide an acceptable value of a given quality functional $F(x, y, \{f_t\},\{w^t\})$, where t is a number of a tree in the forest, t = 1, 2, ..., T, T is a number of trees in the forest, f is a model structure of a t-th tree , w^t is a set of model parameter values of a t-th tree [20].

It is obviously, that the problem of the quality functional creation of models based on decision trees and forests requires the determination of a set of indicators $\{I_i\}$ that quantitatively describers the properties of the models.

3 Primary model characteristics

Along with the sample parameters described above, we will use such notation for the characteristics of the samples: <x_{test}, y_{test}> is a test sample, S_{test} is a number of instances in the test sample; N, N' are, respectively, the number of signs in the original set and in the reduced set of features; $f_{\max}, f_{\min}$ are, respectively, the maximum and minimum boundary values of a model output, $y^{\max}, y^{\min}$ are, respectively, the maximum and minimum boundary values of output feature.

The basic properties characterizing the model structure are defined as: M is a number of levels in a tree, N_n is a number of nodes in the model, N_η is a number of nodes in the η-th layer of a tree, $N_n^{\max}$ is a maximum possible number of nodes in a model, f_i is an i-th node function, $\Delta_{\min}$ is a smallest possible change of the real number, taking into account the bit grid of the computer, $\vartheta_{\alpha o}(j)$ is a complexity of the *j*-th node, which can be defined similarly to [250] in units of elementary operations of addition and multiplication, ϑ^{i*}_{aut} is a characteristic of autonomy of a formation of *i*-th element of a model structure (ϑ^{i*}_{aut} = 0, if the inclusion (or exclusion) of *i*-th node to the model determined only by the human; ϑ^{i*}_{aut} = 1, if the inclusion (or exclusion) of *i*-th node to the model is automatically defined by the training method; ϑ^{i*}_{aut} = 0.5, if the inclusion (or exclusion) of *i*-th node to the model can be determined by the human or training method), $\vartheta_{np}(i)$ is a characteristics of plasticity of *i*-th node of the tree, which is equal to the number of possible states of a node *i* (for leaf nodes containing singleton (not containing functions) $\vartheta_{np}(i)$ = 1, for the nodes with functions the $\vartheta_{np}(i)$ should be taken equal to the number of different functions that may contained in the node, for the rest nodes of the tree we should take $\vartheta_{np}(i)$ as a number of branches of the node), w_{ij} is a connection of *i*-th and *j*-th nodes (w_{ij} = 0, if *i*-th and *j*-th nodes are not connected, and w_{ij} = 1, if *i*-th and *j*-th nodes are connected).

Let introduce the notation for the description of a model parameters: $w_{\max}, w_{\min}$ are, accordingly, the maximum and minimum possible values of a model parameters, N_w is a number of adjustable parameters of model node, $N_w^{\max}$ is a maximum possible number of adjustable parameters of a model nodes, $w_{i,j}^{\max}, w_{i,j}^{\min}$ are, respectively, the maximum and minimum possible values of *j*-th parameter in *i*-th node, Δw is a smallest possible change in weights taking into account the size of the computer bit grid, w_j is a *j*-th model parameter, $N_w(i)$ is a number of parameters of *i*-th model node, $\Delta w_{i,j}$ is a minimum possible change of *j*-th parameter of *i*-th node taking into account the bit grid size of a computer, $\vartheta_{sp}(i)$ is a characteristics of a plasticity of parameters of *i*-th node ($\vartheta_{sp}(i)$ = 0 if the node has no adjustable parameters, otherwise set: $\vartheta_{sp}(i) = \sum_{j=1}^{N_w(i)} \text{round}((w_{i,j}^{\max} - w_{i,j}^{\min}) / \Delta w_{i,j})$),

We define the notation for describing the functioning of a model as: E_{tr}, E_{test} are, respectively, model errors for the training and test samples, $E(w)$ is a model error at a set of weights *w*.

The following notation will be used for the model training method parameters: N_{met} is a the number of training method parameters, N_{met}^{aut} is a number of parameters of a training method, which values of are determined automatically without a human intervention, $\vartheta_{aut}(w_i)$ is a characteristic of autonomy of *i*-th node parameter values setting ($\vartheta_{aut}(w_i)$ = 0, if parameter values set only by a human; $\vartheta_{aut}(w_i)$ = 1, if the values of the node parameters are determined automatically by the training method; $\vartheta_{aut}(w_i)$ = 0.5, if the values of the node parameters can be determined by the human and method).

For tree models in a forest we denote: $w_{f,\max}, w_{f,\min}$, accordingly, maximum and minimum possible values of parameters of a forest model, N_{w_f} is a number of parameters of a forest model without taking into account the number of parameters of its trees, $T^{\max}$ is a maximum possible number of trees in the forest.

If necessary to distinguish the use of indicators for the decision tree we will use notations in the form of I^t or I^{tree} (here *I* is an indicator, *t* is a tree number, *tree* is a tree symbol), and for the forest we will use notation of the form of I^{forest} (here *I* is an indicator, *forest* is a forest symbol).

4 Model Quality Indicators

Diversity is defined by a number of different states of a system. In accordance with the law of "Requisite variety" of W.R. Ashby, creating of a system able to decide a problem, which has certain known diversity (complexity), it is necessary to provide for a system an even greater diversity (knowledge of solving methods) than the diversity of the problem being addressed, or to ensure the ability of the system to create this diversity within itself (it would have methodology, could have developed a methodic or proposed new methods for solving the problem) [32].

The absolute indicator of the limiting diversity of the synthesized model based on the decision tree *tree,* by analogy with [24], is defined as (1):

$$I_{div}(tree) = \text{round}\left(\frac{w_{\max} - w_{\min}}{\Delta w}\right)^{N_w} \prod_{i=1}^{N_n^{\max}} (\vartheta_{np}(i)). \tag{1}$$

The absolute indicator of the limiting diversity of the synthesized model based on the forest of decision trees *forest* is defined as (2):

$$I_{div}(forest) = \text{round}\left(\frac{w_{f,\max} - w_{f,\min}}{\Delta w}\right)^{N_{w_f}} \prod_{t=1}^{T^{\max}} (I_{div}(tree_t)). \tag{2}$$

The greater the value of limiting diversity, the wider the range of models we can obtained on its basis.

By analogy with [24], we define the diversity indicators of the tree model *tree* and forest model *forest*:

– in relation to the training sample as (3) and (4):

$$I_{div}(tree, <x, y>) = \frac{I_{div}(tree)}{I_{div}(x, y)}; \tag{3}$$

$$I_{div}(forest, <x, y>) = \frac{I_{div}(forest)}{I_{div}(x, y)}; \tag{4}$$

– in relation to the population universe as (5) and (6):

$$I_{div}(tree, X, Y) = \frac{I_{div}(tree)}{I_{div}(X, Y)}; \tag{5}$$

$$I_{div}(forest, X, Y) = \frac{I_{div}(forest)}{I_{div}(X, Y)}. \tag{6}$$

The more the value of $I_{div}(tree, \langle x, y \rangle)$ for a single tree, and the value of $I_{div}(forest, \langle x, y \rangle)$ for the forest, the more will be the model potential for approximating the relationship represented by the sample. The smaller the value of the corresponding indicator for the model at an acceptable level of error E , the better the approximation of the sample is.

The more the value of $I_{div}(tree, X, Y)$ for a single tree, and the value of $I_{div}(forest, X, Y)$ for a forest, the more the model will be able to solve the given problem. However, if the relevant indicator is greater than one, or is near to one, then the model is too excessive for the problem solving.

The equivalence of models is determined as follows: two models are equivalent if they have the same sets of answers (they respond equally to the same input stimuli) [33].

The equivalence coefficient of trained models based on decision trees t_1 and t_2 for the sample $\langle x, y \rangle$ we defined as (7):

$$I_{eq}(t_1, t_2) = \exp\left(-\frac{1}{S} \sum_{s=1}^{S} \left(f_{t_1}(x^s) - f_{t_2}(x^s) \right)^2 \right). \tag{7}$$

The equivalence coefficient of trained forests models can be determined similarly to the above, replacing the calculated tree outputs with the corresponding forest outputs.

The values of the equivalence indicator will be in the range from zero to one: the more similar the responses of the models with the same input influences, the greater the value of the equivalence coefficient.

Retraining of a model for the training sample x relatively to the test sample $\langle x_{\text{test}}, y_{\text{test}} \rangle \neq \langle x, y \rangle$ may be defined as (in the given form with substitutions) [24]:

– for classification problems as (8):

$$\delta(x, x_{\text{test}}) = \frac{1}{S} \sum_{x=1}^{S} \left\{ 1 \middle| f(x^s) \neq y^s \right\} - \frac{1}{S_{\text{test}}} \sum_{x=1}^{S_{\text{test}}} \left\{ 1 \middle| f(x^s_{\text{test}}) \neq y^s_{\text{test}} \right\}; \tag{8}$$

– for the evaluation problems as (9):

$$\delta(x, x_{\text{test}}) = \frac{1}{S} \sum_{x=1}^{S} \left\{ 1 \middle| \delta \leq \left| f(x^s) - y^s \right| \right\} - \frac{1}{S_{\text{test}}} \sum_{x=1}^{S_{\text{test}}} \left\{ 1 \middle| \delta \leq \left| f(x^s_{\text{test}}) - y^s_{\text{test}} \right| \right\}, \tag{9}$$

where δ is an error threshold.

Since the error threshold for an instance in practice cannot always be set, as well as for greater universality and uniformity in solving various problems we define it as (10):

$$\delta(x, x_{\text{test}}) = -\frac{1}{S}\sum_{s=1}^{S}\exp\left(-\left(\frac{f(x^s)-y^s}{\max\limits_{s=1,2,...,S}(y^s)-\min\limits_{s=1,2,...,S}(y^s)}\right)^2\right)+ \\ +\frac{1}{S_{\text{test}}}\sum_{s=1}^{S_{\text{test}}}\exp\left(-\left(\frac{f(x^s_{\text{test}})-y^s_{\text{test.}}}{\max\limits_{s=1,2,...,S}(y^s_{\text{test}})-\min\limits_{s=1,2,...,S}(y^s_{\text{test}})}\right)^2\right). \quad (10)$$

These indicators can be used not only for a model based on a single tree, but also for a forest of decision trees, using as f the output value determined by the ensemble of forest trees.

The higher the value of the retraining indicator, the worse the approximating properties of the model for data that did not used in the training.

Confidence in a decision-making is a subjective assessment by the model of the made decision [34].

Regarding the value at the model output for the instance x^s fed to its inputs, we determine the confidence indicator of the model in the made decision (11):

$$I_{cert}(x^s) = \exp\left(-(f(x^s)-y^s)^2\right)\exp\left(-\sum_{j=1}^{N}(C_j^{u(x^s)}-x_j^s)^2\right), \quad (11)$$

where C_j^q is a value of the coordinate on j-th feature of q-th cluster center corresponding to the node of the tree, to which the recognized instance x^s hits. For instances that are not included in the training set, instead of y^s it is possible to substitute the value of the output feature associated with the center of the corresponding cluster.

It is possible to estimate the coordinates of the cluster centers for leaf nodes on the basis of a training sample:

– for the decision tree constructed on the basis of the training sample, determine the belonging of the training sample instances to leaf nodes;

– for every q-th leaf node u_q form a cluster $C^q = \{C_j^q\}$ of instances fallen into this node, $q = 1, 2, ..., Q$, where Q is a number of leaf nodes (clusters);

– for each j-th feature as the coordinate of the cluster center take the arithmetic average of the coordinates of the instances of the corresponding cluster according to the corresponding feature (12):

$$C_j^q = \frac{1}{S^q}\sum_{s=1}^{S}\{x_j^s \mid x^s \in u_q\},\ j = 1, 2, ..., N;\ q = 1, 2, ..., Q, \quad (12)$$

where S^q is a number of instances of the training sample that fell into the q-th node (cluster).

– determine the function $u(x^s)$ that maps the recognized instance x^s to the node number of the tree into which it fell.

The average confidence of the decision tree for a sample x is defined as (13):

$$I_{cert}(x) = \frac{1}{S}\sum_{s=1}^{S} I_{cert}(x^s). \tag{13}$$

Indicators of subjective confidence of model based on a decision tree will take values in the range from zero to one: the higher the value, the closer the properties of a recognized instance to formed cluster center templates, and the more the model is confident in the made decision.

The confidence of the forest of decision trees for the instance x^s is defined as (14):

$$I_{cert}^{forest}(x^s) = \underset{k}{\Re}\left\{I_{cert}^{forest}(x^s, k)\right\}, k = 1, 2, ..., K, \tag{14}$$

where

$$I_{cert}^{forest}(x^s, k) = \underset{t}{\aleph}\left\{I_{cert}^{t} \mid f_t(x^s) = k\right\}, t = 1, 2, ..., T, \tag{15}$$

f_t is an estimated value of the model output of t th tree, I_{cert}^{forest} is an indicator of confidence of forest models, $\underset{k}{\Re}$, $\underset{t}{\aleph}$ are symbols of operators, defining the confidence of the forest in the decision for the k-th class and the t-th tree, respectively. As such operators it is possible to use the minimum, maximum, arithmetic mean of the set of arguments.

The indicator of averaged confidence of forest for sample x is defined as (16):

$$I_{cert}^{forest}(x) = \frac{1}{S}\sum_{s=1}^{S} I_{cert}^{forest}(x^s). \tag{16}$$

Indicators of subjective confidence of the forest will take values in the range from zero to one: the higher the indicator value, the more the trees ensemble confident in made decision.

The hierarchical organization of the structure, the integrity and crushability of elements allows to build models of complex objects from simpler ones; the work of the hierarchical structure requires that the information element in each hierarchical level behave as a whole, but when moving from level to level it must be fragmented, and when moving from the upper hierarchical level to the lower, this fragmentation corresponds to the allocation of its constituent elements, and when moving from the lower level to the top, it corresponds to the inclusion of a certain part of this element in a more complex object [33].

The hierarchy of the model based on the decision tree is defined by analogy with [25] as (17):

$$I_h = \frac{2\sum_{\eta=1}^{M} \eta N_\eta}{M(M+1)N_n}, \; M \geq 1, \; N_n \geq 1. \tag{17}$$

The greater the I_h value, the greater the number of hierarchical levels in the model with respect to maximum possible number of levels for a given number of nodes N_n.

Estimate the maximum possible number of levels. Since the maximum number of levels in the tree will be at the minimum number of outcomes from nodes, then at each level there should be at least one node with two outcomes, and the rest of the nodes should be leafy. Moreover, the greatest number of levels will be achieved for the tree, where only one node at each level (except for the last) has two outcomes, and the rest are leafy and contain only two nodes at each level. Thus, for the deepest tree, the number of nodes of the highest level is 1 (root), for the lower level is 2 (leaves), for the remaining layers is 2, i.e. $1+2(M–1) = N_n$. From here we get: $M = 0.5(N_n–1)+1$. Therefore, for the decision tree we get (18):

$$I_h = \frac{2\sum_{\eta=1}^{M} \eta N_\eta}{0{,}25 N_n^{\,3} + N_n^{\,2} + 0{,}75 N_n}, \; M \geq 1, \; N_n \geq 1. \tag{18}$$

The hierarchy of the model based on the forest of decision trees is defined as (19):

$$I_h^{forest} = \max_{t=1,2,\ldots,T} \{I_h^t\}. \tag{19}$$

The elasticity for a function $y(x)$ on the variable x_j in [35] is defined as (20):

$$El_{x_j}(y) = \lim_{\Delta x_j \to 0} \frac{\frac{\Delta y}{y}}{\frac{\Delta x_j}{x_j}} = \left(\lim_{\Delta x_j \to 0} \frac{\Delta y}{\Delta x_j} \right) \frac{x_j}{y}, \tag{20}$$

where $\Delta y = \frac{y(x_j + \Delta x_j) - y(x_j)}{y(x_j)}$, $x_j > 0, y > 0$.

The relative elasticity indicator on the variable x_j of approximating function $y=f(x)$ realized by the model at output y trained on a training sample <x, y> is defined similarly to [24] as (21):

$$El_{x_j}(y) = \frac{1}{2S} \sum_{s=1}^{S} \left(\frac{\tilde{x}_j^s \left(\tilde{f}_{(\tilde{x}_j^s + \Delta x_j)} - \tilde{f}_{(\tilde{x}_j^s)} \right)}{\Delta x_j \left(\tilde{f}_{(\tilde{x}_j^s)} \right)^2} \right), \tag{21}$$

where $\widetilde{f}_{(\widetilde{x}_j^s)}$ is a calculated value at the output of the model when applying normalized values of the features of s-th instance to its inputs; $\widetilde{f}_{(\widetilde{x}_j^s+\Delta x_j)}$ is a value of the model output when applying normalized values of features of s-th instance to its inputs, and corrected normalized by Δx_j value of j-th feature of s -th instance to j-th input.

The larger the value of elasticity indicator, the more elastic is the model. This indicator applies both to a single tree model and for a forest model.

Equifinality is a regularity of functioning and development of the system, characterizing its ultimate capabilities [36].

Relative equifinality of a model based on a decision tree *tree* defined similarly to [24] as (22):

$$I_{eqf}(tree, < x, y >) = \frac{N_w}{N_w^{\max}} \frac{N_n}{N_n^{\max}} \exp\left(-\frac{1}{S}\sum_{s=1}^{S}(f(w, x^s) - y^s)^2\right). \qquad (22)$$

Relative equifinality will receive the largest value (top-limited by one) for those models which have reached the maximum possible size and the number of parameters during synthesis as well as the smallest error (bottom limited by zero) in the learning process.

For the forest of decision trees, we determine the relative equifinality indicator as (23):

$$I_{eqf}(forest, < x, y >) = \frac{T}{T^{\max}} \prod_{t=1}^{T} I_{eqf}(tree_t, < x, y >). \qquad (23)$$

Generalization is the model's ability to integrate partial data to determine patterns and prolongate results, that is, after training based on the training set, to give answers for test sample instances similar to the training sample but not included in it [37, 38].

The generalization indicator of the decision tree for the training $<x, y>$ and test $<x_{\text{test}}, y_{\text{test}}>$ samples is determined by analogy with [37, 38] as (24):

$$I_G = 1 - \exp\left(-\frac{1}{S_{\text{test}}}\sum_{p=1}^{S_{\text{test}}}\left\{\frac{(f(x^{p*}) - y_{\text{test}}^p)^2 \sum_{j=1}^{N}(x_j^s - x_{j\text{test}}^p)^2}{N(y^s - y_{\text{test}}^p)^2} \middle| (y_i^s - y_{i\text{test}}^p)^2 > 0\right\}\right), \qquad (24)$$

$$s = \arg\min_{t=1,2,\ldots,S} \sum_{j=1}^{N}(x_j^t - x_{j\text{test}}^p)^2,\ j = 1, 2, \ldots, N, p = 1, 2, \ldots, S_{\text{test}}.$$

In a similar way, the generalization indicator for the forest of decision trees will be determined.

Generalization indicator will take values in the range from zero to one, and will be the greater, the smaller the error of a model at instance recognition, and difference of recognized instance to nearest by features instance of a training sample is more.

The generalization indicator of the trained model is defined as (25):

$$I_{gen} = \frac{NS}{N_w N_n} \exp(-(E_{\text{tr}} - E_{\text{test}})^2). \tag{25}$$

If generalization indicator is significantly greater than one, then the model shows great ability to generalization, if the generalization indicator much smaller than one, then the model does not shows no generalizing properties.

Generalization indicator for a forest of decision trees is defined as (26):

$$I_{gen}^{forest} = \frac{NS}{\sum_{t=1}^{T} N_w^t N_n^t} \exp(-(E_{\text{tr}} - E_{\text{test}})^2). \tag{26}$$

The errors here are defined for the ensemble of trees.

Nonlinearity is a dependency that cannot be explained by a linear combination of variable inputs [39, 40].

The nonlinearity indicator for classification problems is defined similarly to [39] as (27):

$$I_{nl} = \frac{2}{S(S-1)} \sum_{s=1}^{S} \sum_{p=s+1}^{S} \left(\frac{\sum_{\ell=0}^{S} \left\{ 1 \middle| f\left(\frac{\ell x^p}{S} + \left(1 - \frac{\ell}{S}\right) x^s \right) \neq f(x^p) \right\}}{\sqrt{\sum_{j=1}^{N} \left(x_j^s - x_j^p\right)^2}} \right). \tag{27}$$

The nonlinearity indicator for estimation problems is defined as (28):

$$I_{nl} = \frac{2}{S(S-1)} \sum_{s=1}^{S} \sum_{p=s+1}^{S} \left(\frac{\sum_{\ell=0}^{S} \left| \frac{f(\ell x^p S^{-1} + (1 - \ell S^{-1}) x^s) - f(x^p)}{f_{\max} - f_{\min}} \right|}{\sqrt{\sum_{j=1}^{N} (x_j^s - x_j^p)^2}} \right). \tag{28}$$

The nonlinearity indicator of the classifier will take values in the range [0, 1]: the greater its value, the more nonlinear is a model. A disadvantage of this indicator is its applicability only for models with single output, and exclusively for classification problems.

The nonlinearity indicator for estimation problems is applicable also for classification problems.

In a similar way, the nonlinearity indicator for the forest of decision trees can be determined.

The indicator of compliance of nonlinearities of the sample and model is defined as (29):

$$\widetilde{I}_{nl} = \frac{I_{nl}(< x, y >)}{I_{nl}}, \tag{29}$$

where $I_{nl}(< x, y >)$ is a nonlinearity indicator of the sample, determined according to [41, 42].

If the indicator $\widetilde{I}_{nl}$ is equal to one, then we can conclude that the model corresponds to the sample in complexity. If the indicator $\widetilde{I}_{nl}$ is less than one, then the smaller its value, the greater the effect of retraining will be, and it would show possible redundancy of a model. If the indicator $\widetilde{I}_{nl}$ value exceeds one, then the model is not sufficient for good approximation (require additional training or change the model structure).

Robustness is a model property to reliably solve a problem when receiving incomplete and / or damaged data. In addition, the results must be consistent, even if some part of the model is damaged [43, 44].

The robustness of the model on the basis of a decision tree in relation to the input signals is defined as (30):

$$I_{Rb}^{x} = \min_{j=1,2,...,N} \left\{ \frac{\min(\tau_1, \tau_2) - \min\limits_{s=1,2,...,S} \{x_j^s\}}{\max\limits_{s=1,2,...,S} \{x_j^s\} - \min\limits_{s=1,2,...,S} \{x_j^s\}} \right\}, \tag{30}$$

where

$$\tau_2 = \min_{\frac{1}{S}\sum_{s=1}^{S}\left(f\left(x^{s*} \middle| \begin{array}{l} x_b^{s*} = x_b^s, \forall b \neq j, b=1,2,...,N \\ x_j^{s*} = x_j^s - \Delta x_j \end{array} \right) - y_i^s \right)^2 > \varepsilon} \{\Delta x_j\},$$

$$\Delta x_j = \min_{s=1,2,...,S}\{x_j^s\} + \delta_x \left(\max_{s=1,2,...,S}\{x_j^s\} - \min_{s=1,2,...,S}\{x_j^s\} \right),...,$$
$$\max_{s=1,2,...,S}\{x_j^s\} - \delta_x \left(\max_{s=1,2,...,S}\{x_j^s\} - \min_{s=1,2,...,S}\{x_j^s\} \right),$$

$\delta_x \in (0,1)$ is a constant that regulates the accuracy of determining a robustness indicator on model inputs.

The indicator is normalized smallest change in the input signal, resulting in a significant increase of model error.

The robustness of a model based on a decision tree with respect to weights (parameters) is defined as (31):

$$I_{Rb}^{w} = \min_{j=1,2,...,N_w} \left\{ \frac{\min(\tau_1, \tau_2) - w_{\min}}{w_{\max} - w_{\min}} \right\}, \tag{31}$$

$$\tau_1 = \min_{\frac{1}{S}\sum_{s=1}^{S}\left(f\left(x^s \middle| w_j = w_j + \Delta w\right) - y_i^s\right)^2 > \varepsilon} \{\Delta w_j\},$$

$$\tau_2 = \min_{\frac{1}{S}\sum_{s=1}^{S}\left(f\left(x^s \middle| w_j = w_j - \Delta w\right) - y_i^s\right)^2 > \varepsilon} \{\Delta w_j\},$$

$$\Delta w_j = w_{\min} + \delta_w(w_{\max} - w_{\min}),...,w_{\max} - \delta_w(w_{\max} - w_{\min}),$$

where $\delta_w \in (0, 1)$ is a constant regulating accuracy of determination of the robustness indicator on model parameters.

The indicator is a normalized least change in weight values, leading to significantly increase in model error.

The integral robustness indicator for a decision tree can be defined as (32):

$$I_{Rb} = I_{Rb}^{x} \cdot I_{Rb}^{w}. \tag{32}$$

The indicator I_{Rb} will take values in the range [0, 1]. The closer its value to zero, the lower the robustness of the model, the more sensitive the model to a change in input signals or parameter values. The closer the value of the indicator I_{Rb} to one, the greater the robustness of the model, the less sensitive the model to changes in input signals or parameter values.

In this way, robustness indicators for the forest of decision trees can be determined.

The homogeneity of the elements lies in the fact that models are built from many simple unified standard elements that perform elementary actions and are interconnected by various connections [45].

The homogeneity of the functions of the nodes of a decision tree is defined by analogy with [25] as (33):

$$I_{hn} = \frac{2\sum_{i=1}^{N_n}\sum_{j=i+1}^{N_n}\{1 \mid f_i \equiv f_j\}}{N_n(N_n - 1)}. \tag{33}$$

The homogeneity indicator will vary from zero to one: the more its value, the more uniform the corresponding elements of the model.

The homogeneity of the node functions of the of the forest trees is defined as (34):

$$I_{hn}^{forest} = \frac{\sum_{t=1}^{T} I_{hn}^{t} N_n^t (N_n^t - 1)}{\sum_{t=1}^{T} N_n^t (N_n^t - 1)}. \tag{34}$$

The sensitivity to the input signals is characterized by calculating the partial derivatives of the model error function [46–48]. However, this approach is computationally hard.

The averaged normalized indicator of the sensitivity of the output of the decision tree to a change in the input signal is defined as (35):

$$I_{tol} = \frac{1}{SN\left(y^{\max} - y^{\min}\right)} \sum_{s=1}^{S} \sum_{j=1}^{N} \max\left(\tau_1, \tau_2\right), \tag{35}$$

$$\tau_1 = \left(f\left(x^{s*} \middle| \begin{matrix} x_b^{s*} = x_b^s, b \neq j, b = 1,2,..., N, \\ x_j^{s*} = x_j^s + \Delta_{\min} \end{matrix} \right) - y_i^s \right)^2,$$

$$\tau_2 = \left(f\left(x^{s*} \middle| \begin{matrix} x_b^{s*} = x_b^s, b \neq j, b = 1,2,..., N, \\ x_j^{s*} = x_j^s - \Delta_{\min} \end{matrix} \right) - y_i^s \right)^2.$$

The value of the sensitivity indicator will be in the range [0, 1]. The higher the value of the sensitivity indicator, the stronger the model reacts to changes in the input signal, the greater are its categorization capabilities. However, too high sensitivity may indicate a weak model resistance to noise and interference in the input signal.

The average indicator of the sensitivity of the forest to changes in the input signal is defined as (36):

$$I_{tol}^{forest} = \frac{1}{T} \sum_{t=1}^{T} I_{tol}(tree_t). \tag{36}$$

Plasticity determines the complexity of the model's behavior, which is considered as a result of the interaction of many elements, each of which limits the action of others and is limited by others on the way to the formation of global observable behavior [49, 50]. As an analogue of neural plasticity, where neuron nodes are considered as plastic elements for neural network models, with respect to decision trees, we will consider the plasticity of nodes. As an analogue of synaptic plasticity (modification of the strength of the synaptic connection between nodes, implemented by the scales in neural network models) as applied to the decision tree, we will consider the plasticity of tunable parameters of tree nodes.

The relative indicator of plasticity of the nodes of the model by analogy with [25] is defined as (37):

$$I_{np} = \frac{\sum_{i=1}^{N_n} \vartheta_{np}(i)}{N_n^{\max} \vartheta_{np}^{\max}}. \tag{37}$$

The indicator I_{np} will take values in the range from zero to one: the greater its value, the higher the level of plasticity of the model nodes.

The relative plasticity indicator of the adjustable model parameters, by analogy with [25], is defined as (38):

$$I_{sp} = \frac{\sum_{i=1}^{N_n} \vartheta_{sp}(i)}{N_n^2 \text{round}\left(\frac{w_{\max} - w_{\min}}{\Delta w}\right)}. \quad (38)$$

The coefficient I_{sp} will take values in the range from zero to one: the bigger its value, the higher the level of plasticity of the model parameters.

The relative indicator of plasticity of a model is defined as (39) [25]:

$$I_{pl} = I_{np} I_{sp}. \quad (39)$$

The relative indicator of plasticity will take values in the range from zero to one: the greater its value, the higher the level of plasticity of the model and, therefore, it has better adaptive abilities.

For the forest of decision trees, we define the plasticity indicators as (40)–(42):

$$I_{np}^{forest} = \frac{1}{T} \sum_{t=1}^{T} I_{np}^{t}, \quad (40)$$

$$I_{sp}^{forest} = \frac{1}{T} \sum_{t=1}^{T} I_{sp}^{t}, \quad (41)$$

$$I_{pl}^{forest} = \frac{1}{T} \sum_{t=1}^{T} I_{pl}^{t}. \quad (42)$$

Variability is an ability to obtain several different models for approximating dependencies from the same data sample using the same method [51-53]. As applied to decision trees, the variability of models is determined by the choice of a feature for the root node and the order in which features are added for checks at other nodes, the method of determining threshold values in nodes, etc.

The absolute indicator of the variability of the model we define similarly to [25] as (43):

$$I_v = \prod_{\eta=1}^{M} \prod_{i=1}^{N_\eta} \vartheta_v(n(\eta, i)) \vartheta_v(w(\eta, i)), \quad (43)$$

where $\vartheta_v(n(\eta, i))$ is a variability of the verification of the *i*-th node of η-th layer of the model: $\vartheta_v(n(\eta, i))$ = 1, if a non-random feature hit into the node; $\vartheta_v(n(\eta, i))=N$, if a feature for checking in the node selected as random from all original feature set $\vartheta_v(n(\eta, i))= N^*$, if a feature for checking in the node is randomly selected from the set of N^* not yet considered features (44):

$$N^{*}_{(\eta,i)} = \left(\sum_{\mu=1}^{\eta-1} \sum_{j=1}^{N_\mu} 1 \right) + i - 1, \quad (44)$$

where $\vartheta_v(w(\eta,i))$ is a variability of determining the values of the parameters of i-th node of η-th model layer: $\vartheta_v(w(\eta,i))$ is equal to the number of parameters in the node that can be configured non-deterministically, if all parameters in the node depend on previous nodes, then $\vartheta_v(w(\eta,i)) = 1$.

The more the I_v value, the more different models can be obtained based on the corresponding paradigm.

The absolute indicator of the variability of the forest of decision trees is defined as (45):

$$I_v^{forest} = \prod_{t=1}^{T} I_v^t. \quad (45)$$

Noise resistance is the property of the model to provide the correct response to an input signal containing noise [54].

The indicator of the resistance of the trained model to additive noise in the input signal at the j-th input is defined similarly to [24] as (46):

$$I_{tol\,j}^{\ell} = \exp\left(-\frac{1}{2} \sum_{s=1}^{S} (\tau_1 + \tau_2) \right), \quad (46)$$

$$\tau_1 = (f(x^s) - f({}_{(j^+)}x^s))^2, \tau_2 = (f(x_j^s) - f({}_{(j^-)}x^s))^2,$$

$${}_{(j^+)}x_g^s = \begin{cases} x_g^s, g \neq j, g = 1,2,\ldots,N; \\ x_g^s + \ell\left(\max\limits_{p=1,2,\ldots,S} \{x_g^p\} - \min\limits_{p=1,2,\ldots,S} \{x_g^p\} \right), g = j, \end{cases}$$

$${}_{(j^-)}x_g^s = \begin{cases} x_g^s, g \neq j, g = 1,2,\ldots,N; \\ x_g^s - \ell\left(\max\limits_{p=1,2,\ldots,S} \{x_g^p\} - \min\limits_{p=1,2,\ldots,S} \{x_g^p\} \right), g = j, \end{cases}$$

where ℓ is a given noise level, $0 < \ell < 1$. In order to automate the process of ℓ setting it is proposed to use such formula (47):

$$\ell = \min_{j=1,2,\ldots,N} \left\{ \frac{\min\limits_{\substack{s=1,2,\ldots,S;\\ p=s+1,\ldots,S}} \left\{ x_j^s - x_j^p \middle\| y^s \neq y^p \right\}}{\left| \max\limits_{p=1,2,\ldots,S} \{x_j^p\} - \min\limits_{p=1,2,\ldots,S} \{x_j^p\} \right|} \right\}. \quad (47)$$

The greater the value of the indicator of model resistance to noise in the input signal on j-th input, the less important this input to the decision making.

This indicator is applicable both to a model based on a single decision tree and to a forest-based model.

Adaptability is a property of structures to dynamically and independently change their behavior in response to an input stimulus [55]. In relation to a model based on decision trees, adaptability is determined, first of all, by plasticity, which determines the resources for adaptation: the greater the plasticity, the more adaptive the properties of the model.

Plasticity is a necessary but insufficient prerequisite for adaptability. Along with plasticity, the adaptive properties of the model are influenced by the sensitivity of the model, which determines the strength of the reaction of the model to the minimum change in the values of its parameters.

The adaptability indicator of a model is defined as (48):

$$I_{adapt} = I_{pl} \cdot I_{tol}. \tag{48}$$

The larger the adaptability indicator value, the greater the possibility has models to adapt to a given task.

The adaptability indicator in this way can be determined for the forest of decision trees.

Symmetry reflects the proportionality in the arrangement of the parts of the whole in a space, the complete correspondence (in location, size) of one half of the whole to the other half [56]. In relation to decision trees, it is possible to determine the indicators of symmetry and asymmetry.

The indicator of symmetry of the structure of the decision tree is defined as (49):

$$I^{n}_{sym} = \frac{2}{N_n{}^2 - N_n} \sum_{i=1}^{N_n} \sum_{j=i+1}^{N_n} \left\{ 1 \middle| f_i \equiv f_j \right\}. \tag{49}$$

The greater the I^{n}_{sym} value, the greater the symmetry of the structure of the decision tree.

The asymmetry indicator of the model structure based on the decision tree as (50):

$$I^{n}_{asym} = 1 - I^{n}_{sym}. \tag{50}$$

The greater the value of I^{n}_{asym} the greater the asymmetry of the model structure.

The symmetry indicator of the nodes of the decision tree is defined as (51):

$$I^{w}_{sym} = \frac{1}{N_n{}^2} \sum_{i=1}^{N_n} \sum_{j=1}^{N_n} \left\{ 1 \middle| w_{ij} = w_{ji} \right\}. \tag{51}$$

The higher the I^{w}_{sym} value the more the symmetry of model connections.

The asymmetry indicator of model connections based on the decision tree is defined as (52):

$$I^{w}_{asym} = 1 - I^{w}_{sym}. \tag{52}$$

The larger the I^{w}_{asym} value the greater the asymmetry of the model connections.

The general indicator of the symmetry of a model based on a decision tree is defined as (53):

$$I_{sym} = I^{n}_{sym} I^{w}_{sym}. \tag{53}$$

The higher the value *of* I_{sym} the bigger the symmetry of the decision tree .

The general indicator of the asymmetry of a model based on a decision tree is defined as:

$$I_{asym} = 1 - I_{sym}. \tag{54}$$

The larger the I_{asym} the greater the asymmetry of the model.

For the forest of decision trees, the symmetry and asymmetry indicators can be determined as the average values of the corresponding indicators of the individual trees included in the forest.

Emergence (integrity) is a regularity that manifests itself in the system in the appearance of new properties in it, which are absent in its elements. The integrity property is associated with the purpose for which the system is created. Let C_o is an intrinsic complexity, which is the total complexity (content) of system elements without interconnecting them (in the case of pragmatic information, the total complexity of the elements that affect the achievement of the goal), C_v is a mutual complexity characterizing the degree of interconnection of elements in the system (i.e. the complexity of its schema or structure). The degree of system integrity in accordance with [24, 57, 58] is defined as (55):

$$I_{\alpha} = C_v / C_o. \tag{55}$$

The emergence of a model based on a decision tree is defined similarly to [24] as (56):

$$I_{\alpha} = \frac{\sum_{i=1}^{N_n} \sum_{j=1}^{N_n} \vartheta_{\alpha v}(i, j)}{\sum_{j=1}^{N_n} \vartheta_{\alpha o}(j)}. \tag{56}$$

The larger the I_{α} value, the more holistic is the model.

For a forest-based model, emergence is defined as (57):

$$I^{forest}_{\alpha} = \frac{T^2}{\sum_{t=1}^{T} I^{t}_{\alpha}}. \tag{57}$$

Interpretability (logical transparency) is a model property to be understandable for human perception and analysis [59, 60]. Obviously, a model is more interpretable if it is hierarchical, and the average number of node connections does not exceed 5-7 (this number is caused by the peculiarities of the human psyche). Since each node in the decision trees has only one input, it is necessary to consider mainly the number of outcomes from the node.

Heuristically we define interpretability through hierarchy and the number of nodes [25] as (58):

$$I_{\text{interp.}} = \frac{I_h}{\sum_{i=1}^{N_n}\sum_{j=1}^{N_n}\{w_{ij} | j \neq i\}}. \tag{58}$$

The level of model interpretability increases with increasing of $I_{\text{interp.}}$ value.

For the forest of decision trees, we define the interpretability as (59):

$$I_{\text{interp.}}^{forest} = \frac{1+\max\limits_{t=1,2,...,T}\{I_h^t\}}{2N_n^{\ 2}}. \tag{59}$$

Learnability is the property of a model to improve its work (to learn or adapt), using examples to turn it to solve a particular problem [61].

The decision tree model learning indicator similarly to [25] is defined as (60):

$$I_{lr} = \frac{I_{pl}L(tree)}{NSL}, \tag{60}$$

where L is the Lipschitz constant for the training sample [62] as (61):

$$L = \max_{\substack{s,p=1,2,....,S;\\ s\neq p}} \left\{\left\|y^s - y^p\right\| / \left\|x^s - x^p\right\|\right\}, \tag{61}$$

$L(tree)$ is a Lipschitz constant (complexity) of the model [63, 64], which, as applied to the binary decision tree, is estimated as (62):

$$L(tree) \leq 10^{\frac{N'}{K}}\sqrt{N'K}\prod_{\eta=0}^{N'/K}\left(\frac{N'}{2^{\eta}}\right). \tag{62}$$

The greater the value of the learnability indicator, the model has bigger potential for solving the problem of approximating dependence $y = f(x)$ given in tabular form.

For a forest of decision trees, we define the learnability indicator as (63):

$$I_{lr}^{forest} = \frac{\sum_{t=1}^{T} I_{pl}^t L(tree_t)}{NSL}. \tag{63}$$

Autonomy is an agent's ability to act without direct human intervention by control on its own actions and internal state. Autonomy also implies the possibility of learning based on experience [65].

Since the trained computational intelligence models, as a rule, in the process of their functioning in decision-making does not require human intervention, they are equally have a property of autonomous functioning. However, in the process of training the level of autonomy for different models and different training methods may vary considerably.

Therefore, we will consider further characteristics of model autonomy only in relation to the process of its learning. Since the ability to learn is determined by plasticity, the autonomy of learning (self-adaptivity) will be characterized by an indicator that depends on the plasticity characteristics of the model. On the other hand, the dependence of model learning from the human may be characterized by its influence (portion) on the formation of the structure and parameters of the model.

Combining these considerations, we obtain the indicator of autonomy of model training method (64):

$$I_{aut} = \frac{1 + N_{met}^{aut}}{1 + N_{met}} \cdot \frac{\sum_{i=1}^{N_n} \left(\vartheta_{aut}^{i*} \vartheta_{aut}(w_i) \right)}{N_n}. \tag{64}$$

As I_{aut} increases, the level of model autonomy in the training process increases.

For a forest of decision trees, the indicator I_{aut} can be defined as a smallest of the indicators of forest trees (65):

$$I_{aut}^{forest} = \min_{t=1,2,...,T} \{ I_{aut}^{t} \}. \tag{65}$$

5 Integral indicators of model quality

Information quality criteria is a family of integral indicators, depending on the model error E , the training sample volume S and the number of adjusted model parameters N_w. They include Hannan-Quinn Criterion [66], Bayesian Information Criterion [67], Akaike's Information Criterion [68], Corrected AIC [69], and Unbiased AIC [69]. A number of criteria in addition to the error, the sample size and the number of adjustable parameters also take into account the maximum possible number of adjustable parameters $N_w^{\max}$. They include a Minimum Description Length [70] , Shortest Data Description [71], Consistent AIC [69] and Mallow Criterion [72].

At model constructing and comparing it is usually assumed to be identical the sample size. Therefore, it is advisable to exclude the sample size from the comparison criteria. At the same time, various synthesized models may not use all of the features of the sample. Therefore the number of features used in the models N' should be seen as an important property of the models at their comparison.

On the basis of these considerations we can define integral information criterion of a model as (66):

$$IIC = \left(1 - \frac{N'}{NN_w^{\max}}\right)e^{-E}. \quad (66)$$

The *IIC* criterion will take values in the range from zero to one. The less its value the worse the model, and the bigger its value the better the model. Here, for different models, the error values E and the maximum number of adjustable parameters $N_w^{\max}$, as well as the number of used features N', may differ. At the same time the number of features in the original set N will not differ, but is used in the formula (66) for normalizing N'.

This indicator is applicable for comparing models based on single decision trees and for ensembles (forests) of decision trees. For the case of ensembles, N will remain unchanged, and the error E, the number of selected features N' and the number of adjustable parameters $N_w^{\max}$ will be determined for the entire ensemble of trees.

The effectiveness (quality) of a problem solving by the model is determined by the accuracy (error) of problem solving for the training and test data, simplicity, logical transparency and speed of the resulting model, as well as by the cost of model building (by hardware requirements, iteration and time spent of training method).

The generalized indicator of the model effectiveness based on the indicators proposed above is determined by analogy with [25] as (67):

$$I_{ef} = I_{gen} I_{lr} \exp(-E). \quad (67)$$

The indicator I_{ef} can be used as for comparing the models and methods of their synthesis, as for optimization of the model building process.

Similarly, the I_{ef} indicator can be determined for a forest of decision trees.

In this case, the error E and indicators I_{gen}, I_{lr} will be determined for the entire ensemble of trees.

An alternative generalized indicator of model efficiency may be defined by analogy with [25] as (68):

$$I_{ef}' = \frac{2}{\pi}\operatorname{arcsec}\left(\frac{N}{N'}\frac{N_w^{\max}}{N_w}\frac{N_n^{\max}}{N_n}\frac{NS}{N_w}\right)\exp(-E), \quad (68)$$
$$N' \geq 1, N_n^{\max} \geq 1, N_n^{\max} \geq 1.$$

Having resulted similar, we receive (69):

$$I_{ef}' = \frac{2}{\pi}\operatorname{arcsec}\left(\frac{N^2 S N_w^{\max} N_n^{\max}}{N' N_w^2 N_n}\right)\exp(-E), \quad (69)$$
$$N' \geq 1, N_n^{\max} \geq 1, N_n^{\max} \geq 1.$$

An alternative generalized efficiency indicator I_{ef}' may be used as for comparing models and methods for their synthesis, as to optimize the process of model building.

I_{ef}' indicator can be used also for models based on forest of decision-trees. For the case of ensembles, N and S will remain unchanged, and the error E, the number of selected features N', and the number of adjustable parameters $N_w^{\max}$ will be determined for the entire ensemble of trees.

6 Results and Discussion

The set of indicators proposed above is extensive and for its application in practice it is advisable to analyze the proposed indicators.

The Fig. 1 presents the classification of a set of proposed indicators characterizing the properties of models based on decision trees and forests.

Horizontally at Fig. 1, indicators are divided into groups according to the complexity of data compilation: sample (indicators characterize the properties of the sample and are model independent), tree model (indicators are defined for a single decision tree model), forest (indicators are defined for a set of decision tree models of a forest).

The more to the right the indicator is located horizontally at Fig. 1, the higher the level of complexity of data generalization by the model is required to determine it.

Vertically at Fig. 1, indicators are divided into groups according to the level of computational complexity relative to primary characteristics: basic properties (easily identifiable characteristics of data and models), primary indicators (indicators determined on the basis of basic properties), secondary indicators (indicators determined on based on primary indicators), and integrative indicators (indicators determined on the basis of indicators of previous levels).

The higher the level of the indicator, the more difficult it is to calculate it with respect to the primary properties of the data and models.

7 Conclusion

The problem of creation of a quality model for models based on decision trees and forests is solved.

The set of indicators characterizing properties of decision trees and forests is proposed. It allows to quantitatively evaluate such model properties as diversity, equivalence, retraining, confidence in a decision-making, hierarchy, equifinality, generalization, nonlinearity, robustness, homogeneity, sensitivity to the input signals, plasticity, variability adaptability, symmetry, asymmetry, emergence (integrity), interpretability (logical transparency), learnability, and autonomy as for individually considered (single) tree model, as for a an ensemble of tree models (forest).

The prospects of further study are to obtain estimates of the computational (time) and spatial (memory) complexity of calculating the proposed indicators, to conduct an experimental study of the set of the proposed indicators for assessing the properties of models in solving practical problems of diagnosis and automatic classification on features, to identify the relationships between different indicators of the properties of models based on decision trees and forests.

Level	Sample	Tree model			Forest model
Integrative indicators		I_{ef}	I_{ef}'	IIC	
Secondary indicators		I_{Rb} $I_{\text{interp.}}$ $I_{cert}(x)$ $I_{div}(tree, < x, y >)$ $I_{div}(tree, X, Y)$	$\widetilde{I}_{nl}$ I_{gen} I_{lr} I_{adapt}	I_{asym} I_{sym} I^{n}_{asym} I^{w}_{asym} I_{pl}	$I^{forest}_{cert}(x)$ $I^{forest}_{cert}(x^{s})$ I^{forest}_{pl} $I_{div}(forest, < x, y >)$ $I_{div}(forest, X, Y)$
Primary indicators	$I_{nl}(< x, y >)$ $I_{div}(x, y)$ $I_{div}(X, Y)$ L $I_{nl}(< x, y >)$ C^{q}_{j}	$I_{div}(tree)$ $I_{cert}(x^{s})$ I_{h} $I_{eqf}(tree, < x, y >)$ I^{x}_{Rb} I^{w}_{Rb} I_{tol} $I_{tol\,j}^{\ell}$	I_{G} I_{nl} I_{aut} $L(tree)$ I_{α} I_{v} I_{hn} $I_{eq}(t_{1}, t_{2})$	E I^{n}_{sym} I^{w}_{sym} I_{np} I_{sp} $El_{x_{j}}(y)$ $\delta(x, x_{\text{test}})$	$I_{div}(forest)$ $I^{forest}_{cert}(x^{s}, k)$ I^{forest}_{h} I^{forest}_{gen} I^{forest}_{aut} I^{forest}_{lr} $I^{forest}_{\text{interp.}}$ I^{forest}_{α} I^{forest}_{v} I^{forest}_{np} I^{forest}_{sp} I^{forest}_{hn} I^{forest}_{tol} $I_{eqf}(forest, < x, y >)$
Basic properties	${x^{s}}_{j}$ y^{s} S N $y^{\max}$ $y^{\min}$	N' M N_{n} $N^{\max}_{n}$ N_{η} f $f_{\max}$ $f_{\min}$ f_{i} $\Delta_{\min}$ $\vartheta_{\alpha o}(j)$ ϑ^{i*}_{aut} $\vartheta_{np}(i)$ $\vartheta_{sp}(i)$	w $w_{\max}$ $w_{\max}$ Δw N_{w} $N^{\max}_{w}$ w_{ij} $w^{\max}_{i,j}$ $w^{\min}_{i,j}$ w_{j} $N_{w}(i)$ $\Delta w_{i,j}$	$\vartheta_{v}(n(\eta, i))$ $\vartheta_{v}(w(\eta, i))$ N_{met} N^{aut}_{met} $\vartheta_{aut}(w_{i})$	$tree_{t}$ $w_{f,\max}$ $w_{f,\min}$ $N_{w_{f}}$ $T^{\max}$

Fig. 1. Analysis of decision tree and forest indicators

References

1. Haykin, S.: Neural Networks and Learning Machines. Pearson, London (2008).
2. Bishop, C.: Neural networks for pattern recognition. Oxford University Press, New York (1995).
3. Subbotin, S.: The special deep neural network for stationary signal spectra classification. In: Proceedings of 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET 2018), Slavske, pp. 123-128. IEEE, Los Alamitos (2018). doi: 10.1109/tcset.2018.8336170
4. Subbotin, S.: Neural network modeling of medications impact on the pressure of a patient with arterial hypertension. In: Proceedings of the International Conference on Information and Digital Technologies (IDT 2016), Zilina, pp. 249-260. IEEE, Los Alamitos (2016). doi: 10.1109/dt.2016.7557182
5. Subbotin, S.A.: The neural network model synthesis based on the fractal analysis. Optical Memory and Neural Networks (Information Optics), 26(4): 257-273 (2017). doi: 10.3103/s1060992x17040099
6. Kumar, K. V.: Neural networks and fuzzy logic. S. K. Kataria & Sons, New Delhi (2009).
7. Rutkowska, D.: Neuro-Fuzzy Architectures and Hybrid Learning. Studies in Fuzziness and Soft Computing, Springer, Berlin (2002).
8. Oliinyk, A.O., Zayko, T.A., Subbotin, S.O.: Synthesis of Neuro-Fuzzy Networks on the Basis of Association Rules. Cybernetics and Systems Analysis, 50(3): 348-357 (2014). doi: 10.1007/s10559-014-9623-7
9. 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), 22(2): 97-103 (2013). doi:10.3103/s1060992x13020082
10. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: Classification and regression trees. Chapman and Hall, Wadsworth, New York (1984).
11. Rabcan, J., Rusnak, P., Subbotin, S.: Classification by fuzzy decision trees inducted based on Cumulative Mutual Information. In: Proceedings of 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET 2018), pp. 208-212. IEEE, Los Alamitos (2018). doi: 10.1109/tcset.2018.8336188
12. Geurts, P., Irrthum, A., Wehenkel, L.: Supervised learning with decision tree-based methods in computational and systems biology. Molecular Biosystems, 5(12):1593–1605 (2009).
13. Rabcan, J., Levashenko, V., Zaitseva, E., Kvassay, M., Subbotin, S.: Application of Fuzzy Decision Tree for Signal Classification. IEEE Transactions on Industrial Informatics, 15(10): 5425-5434 (2019). doi: 10.1109/tii.2019.2904845
14. Kamiński, B., Jakubczyk, M., Szufel, P.: A framework for sensitivity analysis of decision trees. Central European Journal of Operations Research, 26 (1): 135–159 (2017). doi:10.1007/s10100-017-0479-6
15. Rabcan, J., Levashenko, V., Zaitseva, E., Kvassay, M., Subbotin, S.: Non-destructive diagnostic of aircraft engine blades by Fuzzy Decision Tree. Engineering Structures, 197 (2019). doi: 10.1016/j.engstruct.2019.109396
16. Quinlan, J. R.: Induction of decision trees. Machine learning, 1(1):81-106 (1986).
17. Subbotin, S., Kirsanova, E.: The regression tree model building based on a cluster-regression approximation for data-driven medicine. CEUR Workshop Proceedings, 2255: 155-169 (2018).
18. Breiman, L.: Random Forests. Machine Learning. 45 (1): 5–32 (2001). doi:10.1023/A:1010933404324

19. Denisko, D., Hoffman, M.: Classification and interaction in random forests. Proceedings of the National Academy of Sciences of the United States of America, 115 (8): 1690–1692 (2018). doi:10.1073/pnas.1800256115.
20. Subbotin, S.: A random forest model building using a priori information for diagnosis. CEUR Workshop Proceedings, 2353: 962-973 (2019).
21. Boulesteix, A.-L., Janitza, S., Kruppa, J., König, I. R.: Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2(6):493–507 (2012).
22. Lin, Y., Yongho, J. Random forests and adaptive nearest neighbors. Journal of the American Statistical Association. 101 (474): 578–590 (2006). doi:10.1198/016214505000001230
23. Subbotin, S.: Methods of data sample metrics evaluation based on fractal dimension for computational intelligence model buiding. In: Proceedings of 4th International Scientific-Practical Conference Problems of Infocommunications Science and Technology (PICS and T 2017), pp. 1-6 (2017). doi: 10.1109/infocommst.2017.8246136
24. Subbotin, S. A.: Models of criterions of comparison of neural networks and neuro-fuzzy networks in the problems of diagnosis and pattern classification. Scientific Reports of Donetsk National Technical University. Serie "Informatics, Cybernetics and Computers", 12 (165): 148-151 (2010).
25. Subbotin, S. A.: Analysis of properties and criterions of comparison of neural network models for solving diagnostics anf pattern recognition problems. Data Registration, Saving and Processing, 11 (3): 42–52 (2009).
26. Subbotin, S. A.: Metodics and criterions of comparison of models and algorithms of artificial neural network synthesis. Radio Electronics, Computer Science, Control, 2: 109–114 (2003).
27. Subbotin, S.A., Oliinyk, A.A.: The dimensionality reduction methods based on computational intelligence in problems of object classification and diagnosis. Advances in Intelligent Systems and Computing, 543: 11-19 (2017). doi: 10.1007/978-3-319-48923-0_2
28. Subbotin, S.: Quasi-relief method of informative features selection for classification. In: Proceedings of IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), pp. 318-321. IEEE, Los Alamitos (2018). doi: 10.1109/stc-csit.2018.8526627
29. Subbotin, S.A.: Methods of sampling based on exhaustive and evolutionary search. Automatic Control and Computer Sciences, 47(3): 113-121 (2013). doi: 10.3103/s0146411613030073
30. Subbotin, S., Oliinyk, A.: The sample and instance selection for data dimensionality reduction. Advances in Intelligent Systems and Computing, 543: 97-103 (2017). doi: 10.1007/978-3-319-48923-0_13
31. Subbotin, S.: The instance and feature selection for neural network based diagnosis of chronic obstructive bronchitis. Studies in Computational Intelligence, 606: 215-228 (2015). doi: 10.1007/978-3-319-19147-8_13
32. Ashby, W. R.: An Introduction to Cybernetics. Martino Fine Books, Eastford (2015).
33. Dopico, J. R., de la Calle, J. D., Sierra, A. P.: Encyclopedia of artificial intelligence. Information Science Reference, New York (2009).
34. Luger, G. F.: Artificial Intelligence. Structures and Strategies for Complex Problem Solving. Pearson Education, London (2011).
35. Nievergelt, Y.: The Concept of Elasticity in Economics. SIAM Review, 25 (2): 261–265 (1983). doi:10.1137/1025049.

36. Beven, K.J., Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems. Journal of Hydrology, 249: 11–29 (2001).
37. Pedrycz, W.: Fuzzy modelling: paradigms and practice. Springer, Berlin (1996).
38. Hoekstra, A.: Generalisation in feed forward neural classifiers. Technische Universiteit Delft, Delft (1998).
39. Hoekstra A., Duin, R.: On the nonlinearity of pattern classifiers. In: Proceedings of 13 International conference on Pattern recognition, Vienna, 25-29 August 1996, vol. 4, pp. 271–275. IEEE, Los Alamitos (1996)
40. Grossberg, S.: Nonlinear neural networks: principles, mechanisms, and architectures. Neural Networks, 1(1): 17-61 (1988).
41. Subbotin, S. A.: The training set quality measures for neural network learning. Optical Memory and Neural Networks (Information Optics), 19 (2): 126–139 (2010). doi: 10.3103/s1060992x10020037
42. Subbotin, S.A. The sample properties evaluation for pattern recognition and intelligent diagnosis. In: 10th International Conference on Digital Technologies (DT 2014), Zilina, pp. 321-332. IEEE, Los Alamitos (2014). doi: 10.1109/dt.2014.6868734
43. Weng T.-W, Zhang, H., Chen, P.-Yu, Yi, J., Su, D., Gao, Y., Hsieh C.-J., Daniel, L.: Evaluating the Robustness of Neural Networks: An Extreme Value Theory Approach. https://arxiv.org/pdf/1801.10578
44. Carlini, N., Wagner, D.: Towards Evaluating the Robustness of Neural Networks. In: 2017 IEEE Symposium on Security and Privacy (SP), vol. 1, pp. 39-57. IEEE, Los Alamitos (2017). DOI:10.1109/SP.2017.49
45. Krus, D.J., Blackman, H.S., Test reliability and homogeneity from perspective of the ordinal test theory. Applied Measurement in Education, 1: 79–88 (1988).
46. Alippi, C., Piuri, V., Sami, M.: Sensitivity to errors in artificial neural networks: a behavioral approach. IEEE transactions on circuits and systems – I: Fundamental theory and applications, 42 (6): 358–361 (1995).
47. Hashem, S.: Sensitivity analysis for feedforward artificial neural networks with differentiable activation functions. In: Proceedings of International Joint Conference on Neural Networks, Baltimore, 7-11 June 1992, vol. I., pp. 419–424. IEEE, Los Alamitos (1992).
48. Tao C.-W., Nguyen H.T., Yao J.T., Kreinovich V.: Sensitivity analysis of neural control. In: Proceedings of Fourth International Conference on Intelligent Technologies, Chiang Mai, 17-19 December 2003, pp. 478–482. Chiang Mai University, Chiang Mai (2003).
49. Gerrow, K. Synaptic stability and plasticity in a floating world. Current Opinion in Neurobiology, 20 (5): 631–639 (2010). doi:10.1016/j.conb.2010.06.010. PMID 20655734
50. Meyer, D., Bonhoeffer, T., Scheuss, V.: Balance and Stability of Synaptic Structures during Synaptic Plasticity. Neuron, 82 (2): 430–443 (2014).
51. Kick, D.R., Schulz, D.J.: Variability in neural networks. eLife, 7: e341532018 (2018). doi: 10.7554/eLife.34153
52. Norris, B.J., Wenning, A., Wright, T.M., Calabrese, R.L.: Constancy and variability in the output of a central pattern generator. Journal of Neuroscience, 31:4663–4674 (2011). doi: 10.1523/JNEUROSCI.5072-10.2011
53. Masquelier, T.: Neural variability, or lack thereof. Front. Comput. Neurosci., 7: 7 (2013). doi: 10.3389/fncom.2013.00007
54. Rusiecki, A., Kordos, M., Kamiński, T., Greń, K.: Training Neural Networks on Noisy Data. In: International Conference on Artificial Intelligence and Soft Computing (ICAISC 2014) pp. 131-142. Springer, Cham (2014).

55. Martín J.A., de Lope, J., Maravall, D.: Adaptation, Anticipation and Rationality in Natural and Artificial Systems: Computational Paradigms Mimicking Nature. Natural Computing, 8(4): 757-775 (2009).
56. Mainzer, K.: Symmetry And Complexity: The Spirit and Beauty of Nonlinear Science. World Scientific, Singapore (2005).
57. Bunge, M. A.: Emergence and Convergence: Qualitiative Novelty and the Unity of Knowledge, University of Toronto Press, Toronto (2003).
58. Wan, P.: Emergence à la Systems Theory: Epistemological Totalausschluss or Ontological Novelty? Philosophy of the Social Sciences, 41 (2): 178–210 (2011). doi:10.1177/0048393109350751
59. Japaridze, G., De Jongh, D. The logic of provability. in Buss, S., ed., Handbook of Proof Theory, pp. 476-546 North-Holland, Amsterdam (1998).
60. Molnar, Ch.: Interpretable Machine Learning. A Guide for Making Black Box Models Explainable. https://christophm.github.io/interpretable-ml-book/
61. Valiant, L.: A theory of the learnable. Communications of the ACM, 27 (11): 1134–1142 (1984). doi:10.1145/1968.1972.
62. Iuliano, E.: A Comparative Evaluation of Surrogate Models for Transonic Wing Shape Optimization. In: Andrés-Pérez, E., González, L.M., Periaux, J., Gauger, N.R., Quagliarella, D., Giannakoglou, K.C., eds., Evolutionary and Deterministic Methods for Design Optimization and Control With Applications to Industrial and Societal Problems, pp. 161-180. Springer, Cham (2019).
63. Scaman, K., Virmaux, A.: Lipschitz regularity of deep neural networks: analysis and efficient estimation. In: 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada. https://papers.nips.cc/paper/7640-lipschitz-regularity-of-deep-neural-networks-analysis-and-efficient-estimation.pdf
64. Lipschitz constant. Encyclopedia of Mathematics. http://www.encyclopediaofmath.org/index.php?title=Lipschitz_constant&oldid=30687
65. Russell, S., Norvig, P.: Artificial intelligence: a modern approach. Prentice Hall, Upper Saddle River (2009).
66. Hannan, E. J., Quinn, B. G.: The determination of the order of an autoregression. Journal of the Royal Statistical Society, Ser. B, 41(2): 190–195 (1979).
67. Schwarz, G. E.: Estimating the dimension of a model. Annals of Statistics, 6 (2): 461–464 (1978).
68. Akaike, H.: A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19 (6): 716–723 (1974).
69. Gheissari, N., Bab-Hadiashar, A.: Model selection criteria in computer vision: are they different? In: Proceedings of Digital Image Computing: Techniques and Applications, Sydney, 10-12 December 2003, pp. 185-194. CSIRO, Collingwood (2003).
70. Grünwald, P., Myung J., Pitt, M.: Advances in Minimum Description Length: Theory and Applications. MIT Press, Cambridge (2005).
71. Rissanen, J.: Modeling by shortest data description. Automatica, 14 (5): 465-471 (1978).
72. Mallows, C. L.: Some Comments on CP. Technometrics, 15 (4): 661–675 (1973). doi:10.2307/1267380