

UDC 004.272.26: 004.93

Oliinyk A.¹, Subbotin S.², Lovkin V.³, Blagodariov O.⁴, Zaiko T.⁵

¹PhD., Associate Professor of Department of Software Tools, Zaporizhzhia National Technical University, Zaporizhzhia, Ukraine

²Dr.Sc, Head of Department of Software Tools, Zaporizhzhia National Technical University, Zaporizhzhia, Ukraine

³PhD, Associate Professor of Department of Software Tools, Zaporizhzhia National Technical University, Zaporizhzhia, Ukraine

⁴postgraduate student of Department of Software Tools, Zaporizhzhia National Technical University, Zaporizhzhia, Ukraine

⁵PhD., Associate Professor of Department of Software Tools, Zaporizhzhia National Technical University, Zaporizhzhia, Ukraine

THE SYSTEM OF CRITERIA FOR FEATURE INFORMATIVENESS ESTIMATION IN PATTERN RECOGNITION

Context. The task of automation of feature informativeness estimation process in diagnostics and pattern recognition problems is solved. The object of the research is the process of informative feature selection. The subject of the research are the criteria of feature informativeness estimation.

Objective. The research objective is to develop the system of criteria for feature informativeness estimation which enables to compute informativeness of interdependent feature sets.

Method. The system of criteria for feature informativeness estimation is proposed. The proposed system is based on the idea that feature significance is computed according to spatial location of observations of different classes (size of changing of output parameter). The developed criteria system enables to estimate individual and group feature informativeness in classification and regression problems in situations when initial data samples contain redundant and interdependent features as well as observations with missing values. The proposed criteria don't require to construct models based on the estimated feature combinations, in such a way considerably reducing time and computing costs for informative feature selection. Application of the proposed criteria for estimation and selection of informative features allows to reduce structural complexity of synthesized diagnosis and recognition models, to raise its interpretability and generalization ability due to removing of insignificant, interdependent and redundant features in diagnostics and pattern recognition problems.

Results. The software which implements the proposed system of criteria for feature informativeness estimation and allows to select informative features for synthesis of recognition models based on the given data samples has been developed.

Conclusions. The conducted experiments have confirmed operability of the proposed system of criteria for feature informativeness estimation and allow to recommend it for processing of data sets for pattern recognition in practice. The prospects for further researches may include the modification of the known feature selection methods and the development of new ones based on the proposed system of criteria for individual and group feature informativeness estimation.

Keywords: data sample, pattern recognition, feature selection, informativeness criterion, individual informativeness, group informativeness.

NOMENCLATURE

$\text{ceil}(A)$ is a function which gives the least integer that is greater than or equal to the given value A ;

M is a number of features in the sample of observations S ;

P is a set of features (attributes) of observations in the given sample;

P^* is a feature set which is estimated;

p_{qm} is a value of the m -th feature (attribute) of the q -th observation ($m = 1, 2, \dots, M$, $q = 1, 2, \dots, Q$);

$\Delta p_m(s_q; s_{q\text{oth}})$ is a quantity which computes distance along axis of feature p_m between an observation s_q and its nearest neighbor observation $s_{q\text{oth}}$ which belongs to the other class;

$\Delta p_m(s_q; s_{q\text{same}})$ is a quantity which computes distance along axis of feature p_m between an observation s_q and its nearest neighbor observation $s_{q\text{same}}$ which belongs to the same class;

$\rho(p_{qos,m}|t_q)$ is a probability of the situation when a feature p_m has value $p_{qos,m}$ on condition that output parameter T has value t_q ;

$\rho(p_m = p_{lm}|t_q)$ is a probability of the situation when a feature p_m has value p_{lm} from the l -th range of its values on condition that output parameter has value t_q ;

$\rho(p_m = p_{lm}|t_{qos})$ is a probability of the situation when a feature p_m has value p_{lm} from the l -th range of its values on condition that output parameter has value t_{qos} ;

Q is a number of observations in the given sample of observations S ;

S is a sample of observations (training sample);

$V(P^*)$ is an informativeness of a feature set P^* ;

$V(p_m)$ is an informativeness of a feature p_m ;

$V(p_{qm})$ is a partial individual informativeness of a feature p_m towards an observation $s_q \in S$;

$V(P^*, s_q)$ is a partial group informativeness of a feature set $P^* \subseteq P$ towards an observation $s_q \in S$.

t_q is a value of output parameter of the q -th observation;

T is a set of output parameter values.

INTRODUCTION

Construction of diagnosis and recognition models is connected with search of the most informative feature

combination, which characterizes researched objects, processes or systems [1–4]. The known feature selection methods [5–7] allow to extract combinations of informative features from initial data samples, removing insignificant and redundant characteristics. It simplifies process of diagnosis and recognition model synthesis and also improves its generalization and approximating abilities.

Feature selection methods generally use prognostication or classification error obtained by the model which was constructed using estimated data set as criterion of feature set informativeness estimation for searching of the most informative feature combination [3, 8–16]. Such approach needs significant computational and time costs of resources, because it is connected with computationally complex procedure of model synthesis which should be performed for every estimated feature set. Besides it feature selection result, which is a combination of features with the largest informativeness, depends on the type of model used for estimation.

Informational criteria (Information Gain, Gini Index, Entropy etc.) [3–6] don't require to perform computationally complex procedure of mathematical model synthesis for estimation of feature set informativeness. However such criteria suppose that features of initial data sample are independent [17, 18]. Therefore it is difficult to use such criteria in practice and it is unsuitable for situations when features in initial samples are interdependent and redundant. The described shortcomings cause actuality of the development of the criteria system for feature informativeness estimation, which is free from these drawbacks.

The research objective is to develop the system of criteria for feature informativeness estimation which enables to compute informativeness of interdependent feature sets.

1 PROBLEM STATEMENT

Suppose we have data sample $S = \langle P, T \rangle$, which consists of Q observations. Every observation is characterized by values of attributes $p_{q1}, p_{q2}, \dots, p_{qM}$ and output parameter t_q . Then the problem of informative feature selection can be ideally [1, 7] stated as searching for the feature combination P^* from the initial data sample $S = \langle P, T \rangle$ with minimum value of the given criterion of

feature set quality estimation: $V(P^*) = \min_{X \in XS} V(Xe)$.

2 REVIEW OF THE LITERATURE

As stated above, different criteria of individual and group feature significance estimation [1–8, 17–20] are known at present.

Pair correlation coefficient [7] is widely used for estimation of individual significance, when investigated feature and output parameter values are continuous. However such a criterion allows to estimate availability and closeness of the connection between two parameters only when it is linear.

Informational criterion [1, 6, 7] and criterion which is based on feature entropy computation use informational

approach for estimation of individual feature significance. Such criteria in contrast to pair correlation coefficient enable to estimate also closeness of nonlinear connection between features [6, 7]. However information theory is based on the assumption that system state probability values are known. For practical tasks solving, probabilities should be evaluated based on statistical data and are stochastic quantities. Therefore evaluated values can be considered as accurate only for input data samples $S = \langle P, T \rangle$ of infinitely large size [6, 7].

It is significant that criteria based on the informative approach suppose that features of the sample $S = \langle P, T \rangle$ are independent. That is why such criteria are hardly applied for solution of practical tasks, where training samples contain interdependent features [1, 6, 7].

In the papers [17, 18] it was proposed to compute feature significance based on the Relief method, which allows to estimate informativeness of interdependent features based on geometrical location of features in the sample $S = \langle P, T \rangle$. But such criteria allowed to estimate only individual significance of features and could not be used for estimation of feature set informativeness.

Criteria based on the informative approach (information-theoretical criterion, feature set entropy etc.) are applicable for evaluation of group feature informativeness [1, 7, 8]. However such criteria have the same disadvantages as criteria which are applied for individual informativeness estimation. Furthermore possibility of usage of such criteria is based on the assumption that patterns which define classes of the sample $S = \langle P, T \rangle$ are normally distributed.

Errors of models, which are synthesized using estimated feature set $P^* \subseteq P$, are often used for estimation of group informativeness in solving of feature selection problem. Such approach is characterized by significant computational and time costs of resources during feature selection. It is caused by high computational complexity of model synthesis procedure which should be performed for every estimated feature set $P^* \subseteq P$ and makes it difficult to use the known feature selection methods in practice [1, 7, 9].

Thus disadvantages of the known criteria of individual and group significance estimation cause actuality of the development of criteria system for feature informativeness estimation which should be free from the discovered drawbacks.

The described shortcomings cause actuality of the development of the criteria system for feature informativeness estimation, which is free from these drawbacks.

3 MATERIALS AND METHODS

As mentioned above, it is difficult to use informational criteria [1, 5, 7, 8] in practice for estimation of individual and group feature informativeness because such criteria suppose mutual independence of features of initial data sample. At the same time practicable data samples as a rule contain interdependent features and if such features are used for synthesis of diagnosis or recognition model, then model approximating and generalization properties and its

interpretability are getting worse and model structural complexity is increasing. Moreover possibility of usage of such criteria in practice is based on the assumption that patterns, which form classes of the sample $S = \langle P, T \rangle$, are normally distributed. In the developed criteria system it is proposed to estimate feature informativeness according to spatial location of observations of different classes (size of changing of output parameter). In contrast to criteria proposed in the papers [17, 18] and allowing to estimate individual feature informativeness for classification problem, the developed criteria system enables to estimate also group feature informativeness for classification and regression problems.

Let's define conceptions of individual and group informativeness.

Definition 1. Individual informativeness $V(p_m)$ of a feature p_m is a quantity which characterizes correlation between feature p_m and output parameter T .

Definition 2. Group informativeness $V(P^*)$ of a feature set $P^* \subseteq P$ is a quantity which characterizes correlation between feature set P^* and output parameter T .

Definition 3. Partial individual informativeness $V(p_{qm})$ of a feature p_m towards observation $s_q \in S$ is a quantity which characterizes quality of class separation (or quality of interval separation of output parameter T for discrete values of output parameter T) for observations which are the nearest to s_q along the axis of feature p_m .

Definition 4. Partial group informativeness $V(P^*, s_q)$ of a feature set $P^* \subseteq P$ towards observation $s_q \in S$ is a quantity which characterizes quality of class separation (or quality of interval separation of output parameter T for discrete values of output parameter T) for observations which are the nearest to s_q in feature space P^* .

For estimation of individual and group informativeness the following characteristics of training sample $S = \langle P, T \rangle$ should be taken into account: output parameter value type (discrete with the given number of values, for example, in the task of two-class or multiclass recognition; real values of output parameter), presence of missed values (incomplete or undefined data) in data sample.

Criterion $V(p_m)$ should be implemented for estimation of individual informativeness of features p_m in two-class recognition tasks. It is computed in the following way. At the beginning observations set $S' \subset S$ ($|S'| < |S|$) is randomly selected from the given data sample $S = \langle P, T \rangle$. Then partial individual informativeness $V(p_{qm})$ is calculated for each observation s_q from the set S' and for each feature p_m using expression (1):

$$V(p_{qm}) = \Delta p_m(s_q; s_{qoth}) - \Delta p_m(s_q; s_{qsame}), \quad (1)$$

where values of quantities $\Delta p_m(s_q; s_{qoth})$ and $\Delta p_m(s_q; s_{qsame})$ are calculated using expressions (2) and (3) accordingly:

$$\Delta p_m(s_q; s_{qoth}) = |p_{qm} - p_{qmoth}|, \quad (2)$$

$$\Delta p_m(s_q; s_{qsame}) = |p_{qm} - p_{qmsame}|. \quad (3)$$

Characteristics p_{qmoth} and p_{qmsame} represent values of the m -th feature for observations which are the nearest to s_q and belong to the other or the same class

correspondingly: $s_{qoth} = \min_{t_q \neq t_{qoth}} |p_{qm} - p_{qmoth}|$,

$s_{qsame} = \min_{t_q = t_{qsame}} |p_{qm} - p_{qmoth}|$. At that normalized values

of features p_{qm} are used for computation of values of quantities $\Delta p_m(s_q; s_{qoth})$ and $\Delta p_m(s_q; s_{qsame})$, allowing to reduce values of estimations $V(p_{qm})$ and $V(p_m)$ to the range $[-1; 1]$ which is easily interpretable.

Usage of formulas (1)–(3) for estimation of partial individual informativeness $V(p_{qm})$ is based on the assumption that greater values of feature informativeness correspond to situations with better sample division into classes. Hence the farther observation s_q of class t_q is situated on the feature axis p_m

from the nearest observation $s_{qoth} = \min_{t_q \neq t_{qoth}} |p_{qm} - p_{qmoth}|$ of the other class $t_{qoth} \neq t_q$, the greater is partial individual informativeness $V(p_{qm})$. Analogously, the farther observation s_q is situated on the feature axis p_m from the nearest observation

$s_{qsame} = \min_{t_q = t_{qsame}} |p_{qm} - p_{qmoth}|$ of the same class $t_{qsame} = t_q$, the lower is partial individual informativeness $V(p_{qm})$.

After partial individual informativeness $V(p_{qm})$ is estimated, individual informativeness values $V(p_m)$ are calculated for each feature p_m of each observation $s_q \in S' \subset S$ using expression (4):

$$V(p_m) = \frac{1}{|S'|} \sum_{s_q \in S'} V(p_{qm}). \quad (4)$$

Thus features with high values ($V(p_m) \rightarrow 1$) of individual informativeness $V(p_m)$ are considered as significant and informative towards output parameter T , and features with low values ($V(p_m) \rightarrow -1$) are considered as insignificant.

If input data sample $S = \langle P, T \rangle$ contains information about observations s_q , which have missed values of some features p_{qm} , then three approaches could be used.

Under the first approach observations with missed values of feature p_{qm} are not considered in computation of estimations of feature informativeness $V(p_m)$. This approach has simple realization. However information presented in training samples is lost under it. Besides it some data samples $S = \langle P, T \rangle$ (for example, in medical diagnostic tasks, where some measured characteristics can be missed for the majority of patients) can contain vast majority of observations s_q for which values of some features were not measured because of absence of necessity (for example, parameters which characterize some specific illness) or because of difficulties (for example, measurement of some technical diagnostics object parameters needs special methods of destructive quality evaluation which result in changing of the measured product state to «out-of-specification»). Usage of the first approach, where observations s_q with missed values are not considered, in such situations can give inadequate estimations of feature informativeness and as consequence incorrect results of feature selection procedure.

Under the second approach if there are observations s_q with missed values in the sample $S = \langle P, T \rangle$, then quantities $\Delta p_m(s_q; s_{q\text{oth}})$ and $\Delta p_m(s_q; s_{q\text{same}})$, which are used for evaluation of partial individual informativeness estimations $V(p_{qm})$, should be calculated using formula (5):

$$\Delta p_m(s_q; s_{q\text{os}}) = 1 - \frac{1}{N_{\text{int}}(p_m)}. \quad (5)$$

At that $s_{q\text{os}}$ is the nearest observation to observation s_q with the other or the same output parameter value. Formula (5) is used instead of formulas (2) and/or (3) in situations when value p_{qm} of feature p_m isn't defined for one observation from observations $s_q, s_{q\text{oth}}, s_{q\text{same}}$.

$N_{\text{int}}(p_m)$ is a number of different values (if feature p_m has the given limited number of discrete values) or number of intervals into which feature spread was divided (if feature p_m has real values from the given range).

Under the third approach quantity $\Delta p_m(s_q; s_{q\text{os}})$ is measured using estimated probability that observations s_q and $s_{q\text{os}}$ ($s_{q\text{oth}}$ or $s_{q\text{same}}$) have different values for the given feature p_m instead of using formula (5):

– if one observation (for example, s_q) has undefined value p_{qm} , then estimation $\Delta p_m(s_q; s_{q\text{os}})$ is calculated using expression (6):

$$\Delta p_m(s_q; s_{q\text{os}}) = 1 - \rho(p_{q\text{os},m} | t_q), \quad (6)$$

where probability $\rho(p_{q\text{os},m} | t_q)$ is calculated based on the sample data $S = \langle P, T \rangle$ using formula (7):

$$\rho(p_{q\text{os},m} | t_q) = \frac{N(s \in S : (p_m = p_{q\text{os},m}) \cap (T = t_q))}{N(t_q)}, \quad (7)$$

where $N(s \in S : (p_m = p_{q\text{os},m}) \cap (T = t_q))$ is a number of observations in data sample $S = \langle P, T \rangle$ with value of feature p_m equal to $p_{q\text{os},m}$ and with value of output parameter T equal to t_q ; $N(t_q)$ is a number of observations in data sample S with value of output parameter T equal to t_q ;

– if both observations s_q and $s_{q\text{os}}$ have undefined value of feature p_m , then estimation $\Delta p_m(s_q; s_{q\text{os}})$ is calculated using expression (8):

$$\Delta p_m(s_q; s_{q\text{os}}) = 1 - \sum_{l=1}^{N_{\text{int}}(p_m)} \rho(p_m = p_{lm} | t_q) \cdot \rho(p_m = p_{lm} | t_{q\text{os}}). \quad (8)$$

Formulas (6)–(8) take into account density of distribution of feature values p_m among observations of sample $S = \langle P, T \rangle$ and allow to take into account estimated probability that features p_m have the given values p_{lm} for feature informativeness estimation $V(p_m)$.

If it is necessary to select informative features for multiclass recognition tasks (output parameter T can have different values t_q from the set T' , $|T'| > 2$), then feature informativeness estimations $V(p_m)$ can be calculated using formulas (1)–(8). At that observation $s_{q\text{oth}}$ is selected as observation with minimal distance from observation s_q along axis of the estimated feature p_m and with the other class value T : $s_{q\text{oth}} = \min_{t_q \neq t_{q\text{oth}}} |p_{qm} - p_{q\text{moth}}|$.

Also for multiclass recognition tasks ($|T'| > 2$) average distance from the set S'_q to the nearest observations belonging to every possible class $t_l \in T', t_l \neq t_q$ should be used instead of distance to observation which is the nearest to the observation s_q and has the opposite value of class T . In that case partial individual informativeness $V(p_{qm})$ of feature p_m should be estimated based on the expression (9), which is average distance between observation s_q and its nearest neighbor observations with the opposite classes, weighted by class frequency of the sample $S = \langle P, T \rangle$:

$$\Delta p_m(s_q; S'_q) = \sum_{\substack{l=1 \\ t_l \neq t_q}}^{|T'|} \rho(t_l) \cdot \Delta p_m(s_q; s_l). \quad (9)$$

In the expression (9) quantity $\Delta p_m(s_q; s_l)$ determines distance along axis of feature p_m between observation s_q and its nearest neighbor observation $s_l \in S'_q$ with output parameter value $t_l \in T', t_l \neq t_q$, $\rho(t_l)$ is a probability that observations of sample $S = \langle P, T \rangle$ are characterized by class t_l (is estimated as ratio of number of sample observations $N(t_l)$ with class to overall number of observations in sample).

Then partial individual informativeness of feature can be calculated using formula (10):

$$V(p_{qm}) = \Delta p_m(s_q; S'_q) - \Delta p_m(s_q; s_{qsame}), \quad (10)$$

and individual informativeness $V(p_{qm})$ of feature p_m is calculated using formula (4).

It is proposed to use criterion $V(P^*)$ for estimation of group informativeness of feature set $P^* \subseteq P$ based on data sample $S = \langle P, T \rangle$. This criterion is computed in the following way. At the beginning observation set $S' \subset S$ ($|S'| < |S|$) is randomly selected from the given data sample $S = \langle P, T \rangle$ to reduce number of computing operations. After that partial group informativeness $V(P^*, s_q)$ for the estimated feature set $P^* \subseteq P$ is calculated for each observation $s_q \in S'$ using formula (11):

$$V(P^*, s_q) = \Delta P^*(s_q; s_{qoth}) - \Delta P^*(s_q; s_{qsame}), \quad (11)$$

where quantities $\Delta P^*(s_q; s_{qoth})$ and $\Delta P^*(s_q; s_{qsame})$ determine distance in feature space P^* between observation s_q and observations s_{qoth} and s_{qsame} which are the nearest to s_q and have the other and the same class values correspondingly. Distances $\Delta P^*(s_q; s_{qos})$ between observations s_q and s_{qos} (s_{qoth} or s_{qsame}) can be computed using formulas (12)–(14) which are metrics for evaluation of Euclidean, Hamming and Minkowski distances correspondingly [1, 6, 7] (at that ω is a parameter which is set by user):

$$\Delta P^*(s_q; s_{qos}) = \sqrt{\sum_{p_m \in P^*} (p_{qm} - p_{qos,m})^2}, \quad (12)$$

$$\Delta P^*(s_q; s_{qos}) = \sum_{p_m \in P^*} |p_{qm} - p_{qos,m}|, \quad (13)$$

$$\Delta P^*(s_q; s_{qos}) = \omega \sqrt{\sum_{p_m \in P^*} |p_{qm} - p_{qos,m}|^\omega}. \quad (14)$$

Normalized values of features p_{qm} are used for evaluation of quantities $\Delta P^*(s_q; s_{qos})$. Group informativeness $V(P^*)$ of feature set $P^* \subseteq P$ is calculated as average value of sum of partial informativenesses $V(P^*, s_q)$: $V(P^*) = \frac{1}{|S'|} \sum_{s_q \in S'} V(P^*, s_q)$. If some observations

s_q have missed values of features p_{qm} , then quantities $\Delta P^*(s_q; s_{qos})$ can be estimated by analogy with expressions (5)–(8) used for individual feature informativeness estimation.

It is proposed to use expressions presented above for estimation of individual $V(p_m)$ and group feature informativeness $V(P^*)$ in selection of informative features for regression problems (output parameter T can have different values t_q from spread $|T'| \in [t_{\min}; t_{\max}]$). At that initial spread $|T'| \in [t_{\min}; t_{\max}]$ of output parameter T should be divided into $N_{\text{int}}(T)$ intervals. Then estimations $V(p_m)$ or $V(P^*)$ should be evaluated using formulas (1)–(14).

It is important to consider how to choose number $N_{\text{int}}(T)$ of intervals into which initial spread $|T'| \in [t_{\min}; t_{\max}]$ of output parameter T should be divided. If expert set acceptable level of precision ε_n for applied problem with numerical data determined as sample $S = \langle P, T \rangle$, then interval number $N_{\text{int}}(T)$ can be evaluated using formula (15):

$$N_{\text{int}}(T) = \text{ceil}\left(\frac{1}{2\varepsilon_n}\right). \quad (15)$$

This formula allows to divide spread of output parameter T in such a way, that normalized width of every interval ΔT shouldn't exceed quantity ε_n : $\Delta T \leq \varepsilon_n$. At that value of output parameter t_q of estimated observation s_q is mapped onto interval $\Delta t_l : t_q \in [\Delta t_{l \min}; \Delta t_{l \max}]$ for diagnosis and recognition models construction and also for feature informativeness estimation.

Thus the developed criteria system for feature informativeness estimation proposes to evaluate feature significance according to spatial location of observations of different classes (size of changing of output parameter). The proposed criteria system enables to estimate individual

and group feature informativeness for classification and regression problems in situations when initial data samples contain redundant and interdependent features as well as observations with missing values. The proposed criteria don't require to construct models based on the estimated feature combinations, in such a way considerably reducing time and computing costs for informative feature selection. Application of the proposed criteria for estimation and selection of informative features allows to reduce structural complexity of synthesized diagnosis and recognition models, to raise its interpretability and generalization ability due to removing of insignificant, interdependent and redundant features for diagnostics and pattern recognition problems.

4 EXPERIMENTS

The proposed system of feature informativeness estimation for pattern recognition was integrated into the diagnosis model synthesis software system as corresponding module [21]. This module is implemented for informative feature selection in data sample reduction stage.

Numerical experiments for solving of informative feature selection problem for vehicle recognition [22] were held for investigation of efficiency of the proposed estimation criteria system application in practice.

Every vehicle was presented by images which were gotten from highway video cameras. Every image was in color mode of shades of gray. Whole training sample contained 10,000 images. Image areas where vehicle was situated were identified by recognition system. Areas which were obtained for every image in such a way were displayed into a matrix of 128x128. Object graphic information was encoded using 26 characteristics. Besides that for formation of training sample image data were classified by experts in 6 classes:

- state «not recognized» (class 0);
- motorcyclist image (class 1);
- passenger car image (class 2);
- truck image (class 3);
- bus image (class 4);
- image of minibus or minivan (class 5).

Software system of diagnosis model synthesis should perform synthesis of set which consists of 5 models to recognize vehicles of the given classes. Every model should recognize vehicles of corresponding type. For each model training sample was formed. Every training sample contains 10.000 observations which have values of 26 features and 1 output parameter (does observation belong to the considered class or no?).

It was necessary to estimate group informativeness of training sample feature subsets at the beginning of application of the considered mechanism of diagnosis model synthesis in data reduction process. Feature informativeness estimation criteria system proposed in the paper was applied for this purpose. Hamming distance was used for calculation of group informativeness. After that it was necessary to select subset of training sample features which were used by software system for recognition.

Feature selection stage for the other considered methods (Principal Component Analysis, Group Method of Data Handling, Canonical Method of Evolutionary Search, Method of alternately Adding and Removing of Features,

Multiagent Methods with Indirect and Direct Connection between Agents) was realized according to its special mechanism. Maximum power of feature sets obtained by these methods was limited to 12.

By application of the developed mathematical support and software the tasks of informative feature selection and diagnosis model synthesis were sequentially solved.

Recognition was realized using neural network of direct propagation which contains 3 neurons on the first layer and 1 neuron on the second layer, uses logistic sigmoid activation function of neuroelements and weighted sums as discriminant functions.

Numerical study of the developed software system application based on the proposed estimation criteria system and the traditional feature selection methods was held. For comparison of the study results the following comparison criteria set was developed:

- number *k* of features which were selected as informative and are in sample after reduction;
- recognition error *E*, which is computed as ratio of incorrectly recognized observations to the total number of observations in sample;
- operating time *T_c* which is needed by method to achieve an acceptable solution.

At that the first criterion is a relevant estimation for each specific version of informative feature selection problem (for example, when only passenger cars are recognized). When recognition results are averaged for several classes (for example, vehicles of several types), value of this criterion should be stricken off the results.

In recognition tasks it is important not only to evaluate overall recognition error level *E*, but also to evaluate recognition error levels for observations of corresponding class. So if there are two classes: 1 – observations belonging to images of passenger cars, and 2 – correspondingly observations which don't belong to images of passenger cars, then after recognition the following subsets can be formed: *C₁₁* which consists of observations belonging to the first class and recognized as belonging to the same class, *C₁₂* which consists of observations belonging to the first class and recognized as belonging to the second class, *C₂₁* which consists of observations belonging to the second class and recognized as belonging to the first class, *C₂₂* which consists of observations belonging to the second class and recognized as belonging to the same class. Then recognition error for observations of the first class *E₁* and the second class *E₂* can be calculated as following:

$$E_1 = \frac{|C_{12}|}{|C_{11}| + |C_{12}|}, \tag{16}$$

$$E_2 = \frac{|C_{21}|}{|C_{21}| + |C_{22}|}. \tag{17}$$

Probabilistic optimization realization of the majority of the researched methods causes necessity of results averaging, therefore search was realized 100 times during numerical research, and then averaged values of comparison criteria were calculated.

5 RESULTS

Table 1 presents computed values of comparison criteria for the proposed and known informative feature selection methods in vehicle recognition task. The results represent recognition of vehicles of passenger cars class.

Distribution of recognition error, depending on the number of features (feature subset dimensionality) selected for recognition, is presented in the Figure 1.

Estimations of group informativeness of subsets with corresponding numbers of features for the class of passenger cars are presented in the Figure 2.

Estimations of group informativeness of feature subsets, selected for recognition by the methods listed in the table 1, are presented in the Figure 3.

Distribution of recognition error over the classes defined by formulas (16) and (17) for passenger cars is presented in the Figure 4.

The recognition results obtained for application of all investigated methods for recognition of all types of vehicles are presented in the Figure 5.

Averaged comparison criteria values computed for the proposed (CSFIEFS) and the known informative feature selection methods in recognition of all vehicle classes are presented in the Table 2.

Table 1 – The results of application of informative feature selection methods for passenger cars recognition

№	Feature selection method	Values of comparison criteria		
		E	T_c	K
1	Principal Component Analysis (PCA)	0.0412	524	12
2	Group Method of Data Handling (GMDH)	0.0358	18578	11
3	Canonical Method of Evolutionary Search (CMES)	0.0219	23171	10
4	Method of alternately Adding and Removing of Features (MARF)	0.0471	2199	12
5	Multiagent Method with Indirect Connection between Agents (MMICA) (Ant Colony Optimization for Feature Selection)	0.0198	10318	10
6	Multiagent Method with Direct Connection between Agents (MMDCA) (Bee Colony Optimization for Feature Selection)	0.0191	10192	11
7	Criteria System of Features Informativeness Estimation for Feature Selection (CSFIEFS)	0.0172	712	10

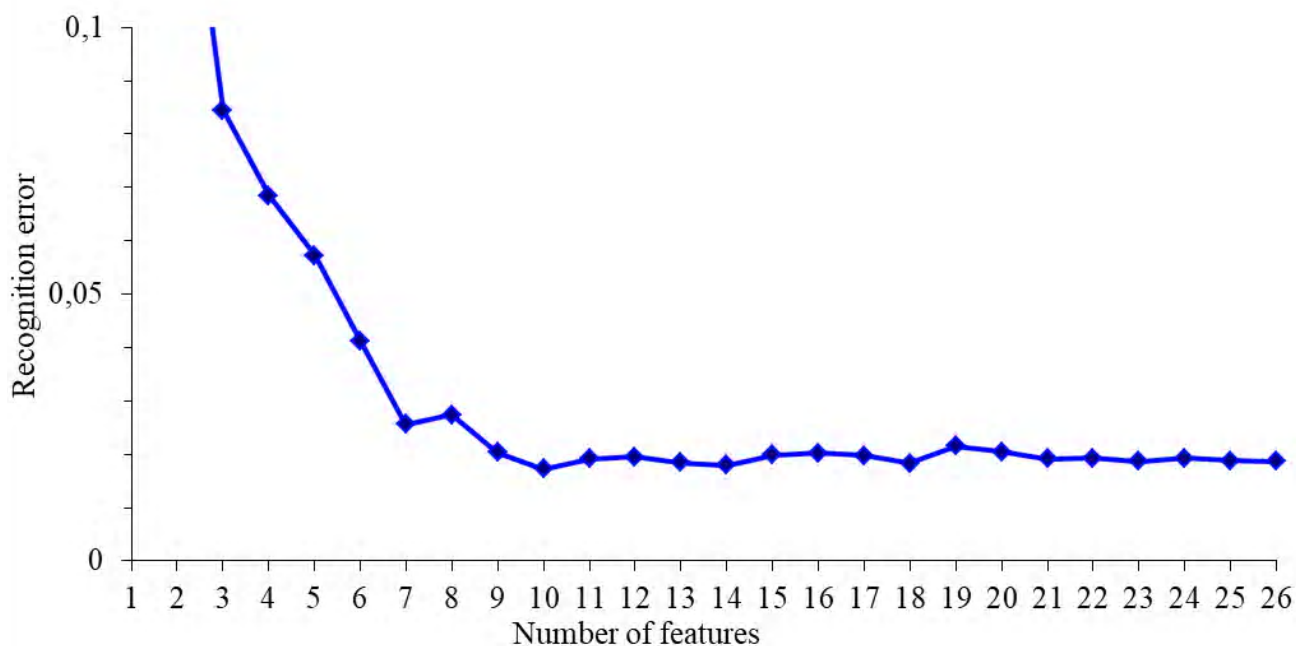


Figure 1 – Graph of dependence between recognition error and feature number

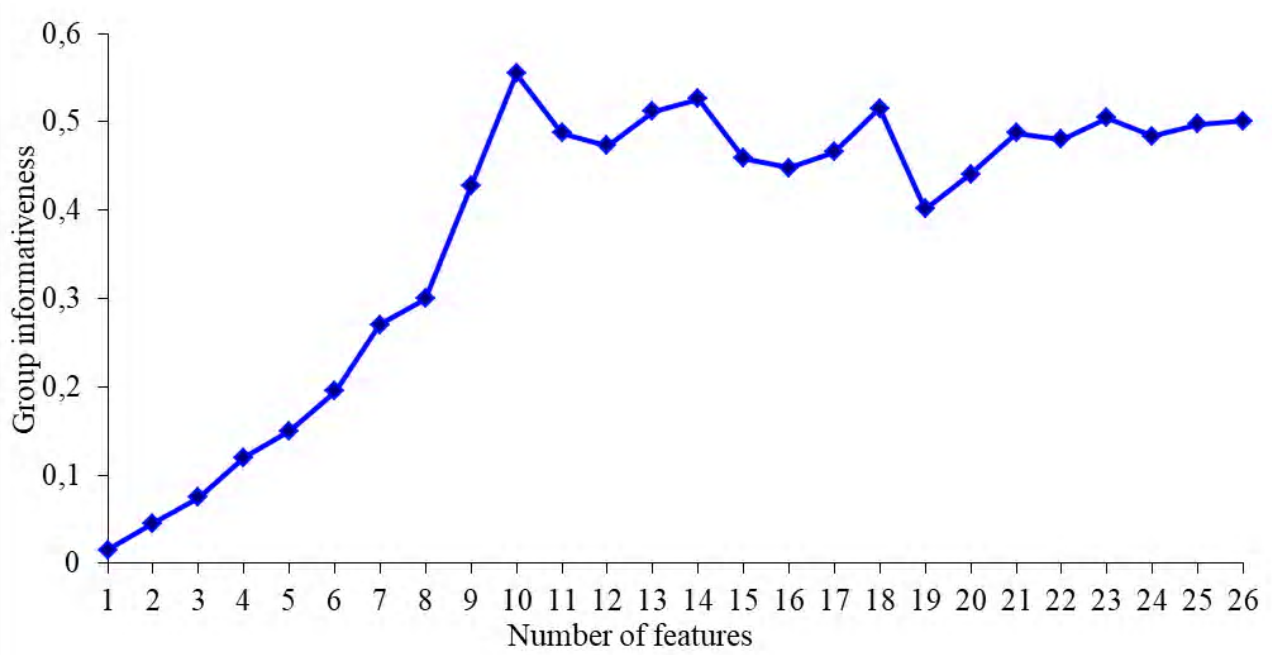


Figure 2 – Graph of dependence between group informativeness of feature subset and feature number

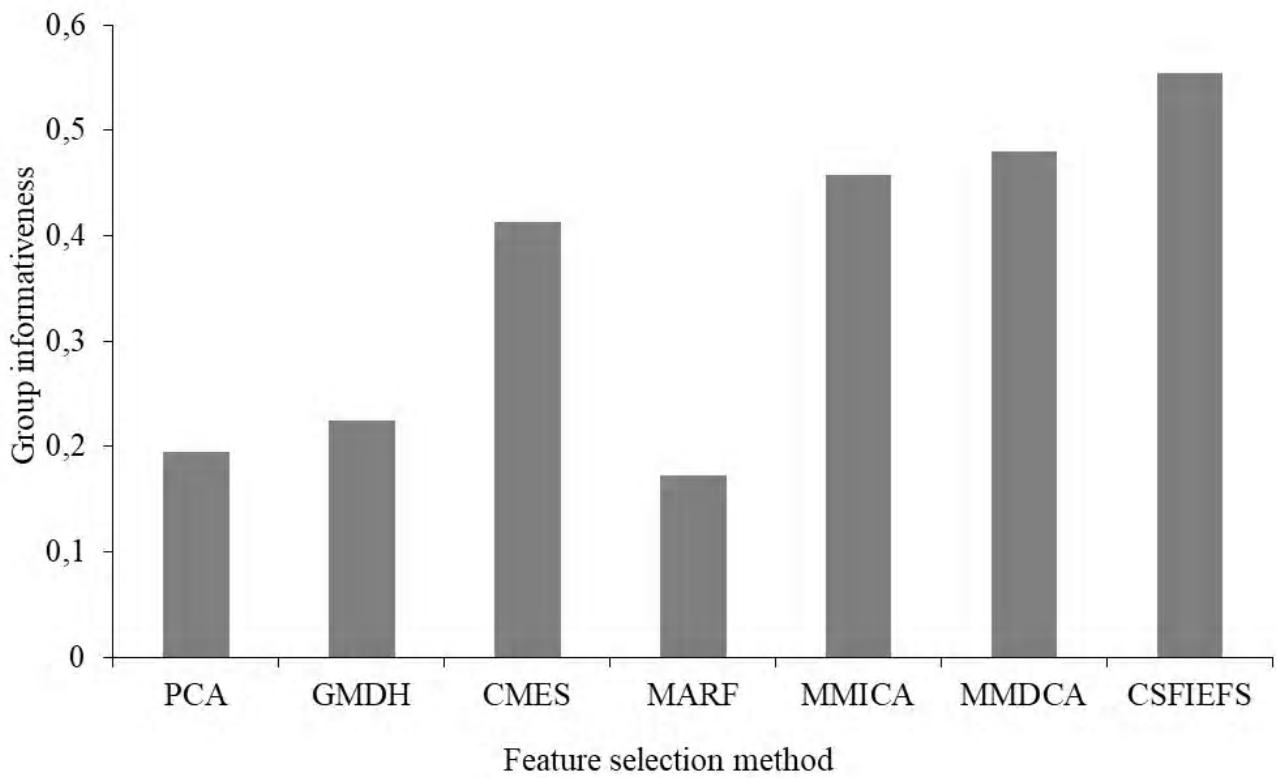


Figure 3 – Diagram of group informativeness of the selected feature subsets

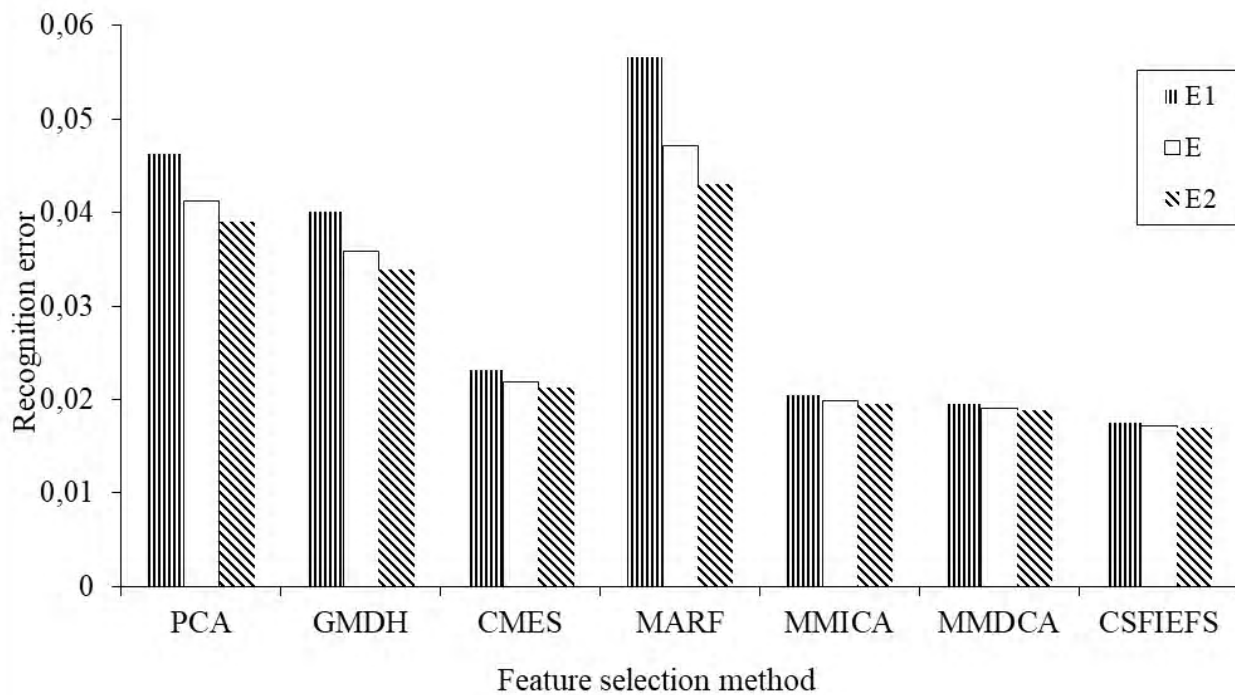


Figure 4 – Diagram of passenger cars recognition error distribution

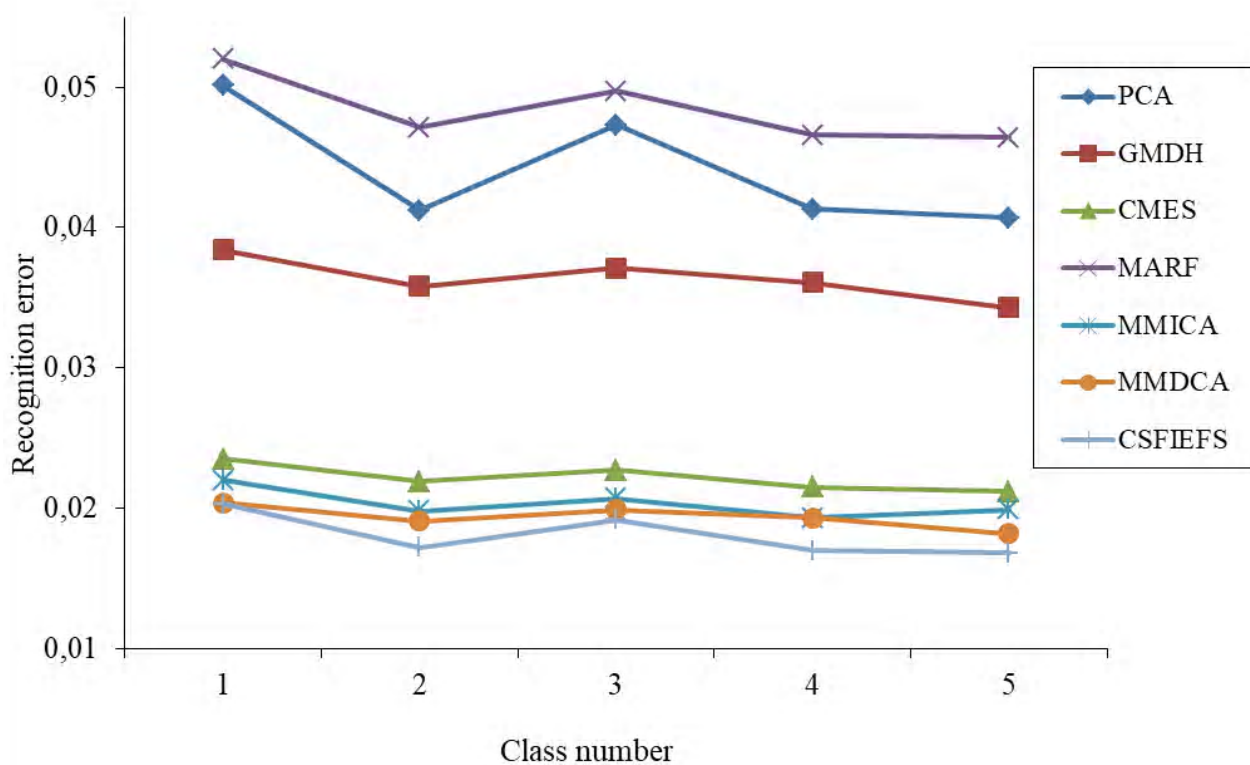


Figure 5 – Graph of dependence between recognition error obtained by the investigated feature selection methods and class number

Table 2 – The results of application of informative feature selection methods for vehicle recognition

№	Feature selection method	Values of comparison criteria	
		E	T_c
1	PCA	0.0441	618
2	GMDH	0.0363	18306
3	CMES	0.0222	23415
4	MARF	0.0484	2262
5	MMICA	0.0203	10549
6	MMDCA	0.0194	10337
7	CSFIEFS	0.0181	799

6 DISCUSSION

Comparison results, presented in the Tables 1 and 2, show that the lowest recognition error value of 0.0181 on average for all vehicles was obtained by the proposed method. Diagram, presented in the Figure 4, also shows acceptable recognition error level when it is divided into two components which correspond to every output feature value (0.0176 and 0.017).

At that PCA (618 sec) and MARF (2262 sec) operated considerably quicker (by a factor of 4.5 and more) than other traditional researched methods, however, its recognition error was the largest too. The proposed method showed speed (799 sec) comparable with these methods. It is caused by the fact that it didn't require model synthesis using data sample which is sufficiently computationally complex and long process. It allows to use the proposed feature informativeness estimation criteria system specifically under conditions of limited time and resources and also when feature selection is separated to the individual stage of decision making and its operating time gets additionally importance.

As shown in Table 1, the lowest number of features (10) for passenger cars recognition was selected by CMES, MMICA and CSFIEFS proposed in the paper. Every method reduced the sample by 61.5 %.

But at the same time as shown in Table 1 (recognition errors of 0.0219, 0.0198 and 0.0172), each method selected different informative feature subsets from the overall set, though these subsets had the same power. Data, represented in the Figure 3, show that estimated group informativeness of feature subsets selected by each method allowed to get estimation which correlates with recognition error. So it can be stated that criteria system proposed in the paper is informative and can be used for decision making.

CSFIEFS allowed to get the lowest recognition error (0.0172) among all feature subsets which were proposed by the researched methods and consist of 10 elements. It corresponds to effective feature set selection and to effective recognition problem solution.

Dependence, presented in the Figure 1, is made from feature subsets of different power. Every subset allowed to get the lowest recognition error among all subsets of the same power. This dependence shows that the most optimal solution is a subset which consists of 10 elements.

Graph, presented in the Figure 2, shows dependence between group informativeness and size of sets which were defined in a way described for the previous dependence. At that this graph shows that the proposed informativeness

criteria allow to estimate subsets in a way relevant to recognition error.

Thus the proposed criteria system for feature informativeness estimation in pattern recognition allows to efficiently solve the problem of informative feature selection, leading to effective solution of the pattern recognition task. At that in comparison with traditional informative feature selection approaches based on the error criteria this process has quicker realization, lower recourse requirements and provides the lowest recognition error.

CONCLUSIONS

In this paper the actual task of automation of feature informativeness estimation process in diagnostics and pattern recognition problems was solved.

Scientific novelty of the paper is in the proposed criteria system of feature informativeness estimation. The proposed criteria system is based on the idea that feature significance is computed according to the spatial location of observations of different classes (size of changing of output parameter). The developed criteria system enables to estimate individual and group feature informativeness in classification and regression problems in situations when input data samples contain redundant and interdependent features as well as observations with missing values. The proposed criteria don't require to construct models based on the estimated feature combinations, in such a way considerably reducing time and computing costs for informative feature selection. Application of the proposed criteria for estimation and selection of informative features allows to reduce structural complexity of synthesized diagnosis and recognition models, to raise its interpretability and generalization ability due to removing of insignificant, interdependent and redundant features for diagnostics and pattern recognition problems.

Practical significance of the paper consists in the solution of practical problems of pattern recognition. Experimental results showed that the proposed criteria system allowed to estimate individual and group informativeness of features and it could be used in practice for solving of practical tasks of diagnostics and pattern recognition.

ACKNOWLEDGMENTS

The work was performed as part of the research work "Methods and means of decision-making for data processing in intellectual recognition systems" (number of state registration 0117U003920) of software tools department of Zaporizhzhia National Technical University.

REFERENCES

1. Jensen R. Computational intelligence and feature selection: rough and fuzzy approaches / R. Jensen, Q. Shen. – Hoboken : John Wiley & Sons, 2008. – 339 p. DOI: 10.1002/9780470377888.
2. Mulaik S. A. Foundations of Factor Analysis / S. A. Mulaik. – Boca Raton, Florida: CRC Press. – 2009. – 548 p.
3. Lee J. A. Nonlinear dimensionality reduction / J. A. Lee, M. Verleysen. – New York : Springer, 2007. – 308 p. DOI: 10.1007/978-0-387-39351-3.
4. Bezdek J. C. Pattern Recognition with Fuzzy Objective Function Algorithms / J. C. Bezdek. – N.Y. : Plenum Press, 1981. – 272 p. DOI: 10.1007/978-1-4757-0450-1.
5. Hyvarinen A. Independent component analysis / A. Hyvarinen, J. Karhunen, E. Oja. – New York : John Wiley & Sons, 2001. – 481 p. DOI: 10.1002/0471221317.

6. Федотов Н. Г. Теория признаков распознавания образов на основе стохастической геометрии и функционального анализа / Н. Г. Федотов. – М. : Физматлит, 2010. – 304 с.
7. Guyon I. An introduction to variable and feature selection / I. Guyon, A. Elisseeff // *Journal of machine learning research*. – 2003. – № 3. – P. 1157–1182.
8. McLachlan G. Discriminant Analysis and Statistical Pattern Recognition / G. McLachlan. – New Jersey : John Wiley & Sons. – 2004. – 526 p. DOI: 10.1002/0471725293.
9. Parallel multiagent method of big data reduction for pattern recognition / [A. A. Oliinyk, S. Yu. Skrupsky, V. V. Shkarupylo, O. Blagodariv] // *Радіоелектроніка, інформатика, управління*. – 2017. – № 2. – С. 82–92.
10. Oliinyk A. Production rules extraction based on negative selection / A. Oliinyk // *Радіоелектроніка, інформатика, управління*. – 2016. – № 1. – С. 40–49. DOI: 10.15588/1607-3274-2016-1-5.
11. Oliinyk A. Parallel computing system resources planning for neuro-fuzzy models synthesis and big data processing/[A. Oliinyk, S. Skrupsky, S. Subbotin et al]// *Радіоелектроніка, інформатика, управління*. – 2016. – № 4. – С. 61–69. DOI: 10.15588/1607-3274-2016-4-8.
12. The model for estimation of computer system used resources while extracting production rules based on parallel computations / [A. A. Oliinyk, S. Yu. Skrupsky, V. V. Shkarupylo, S. A. Subbotin] / *Радіоелектроніка, інформатика, управління*. – 2017. – № 1. – С. 142–152. DOI: 10.15588/1607-3274-2017-1-16.
13. Subbotin S. The Sample and Instance Selection for Data Dimensionality Reduction / S. Subbotin, A. Oliinyk // *Recent Advances in Systems, Control and Information Technology. Advances in Intelligent Systems and Computing*. – 2017. – Vol. 543. – P. 97–103. DOI: 10.1007/978-3-319-48923-0_13.
14. Shitikova O. V. Method of Managing Uncertainty in Resource-Limited Settings / O. V. Shitikova, G. V. Tabunshchik // *Радіоелектроніка, інформатика, управління*. – 2015. – № 2. – P. 87–95. DOI: 10.15588/1607-3274-2015-2-11.
15. Tabunshchik G. V. Verification model of systems with limited resources / G. V. Tabunshchik, T. I. Kaplienko, O. V. Shitikova / *Радіоелектроніка, інформатика, управління*. – 2017. – № 4.
16. Bodyanskiy Ye. Hybrid adaptive wavelet-neuro-fuzzy system for chaotic time series identification / Ye. Bodyanskiy, O. Vynokurova // *Information Sciences*. – 2013. – Vol. 220. – P. 170–179. DOI: 10.1016/j.ins.2012.07.044.
17. Kononenko I. Estimating Attributes: Analysis And Extensions Of Relief / I. Kononenko // *Machine Learning : European Conference on Machine Learning ECML-94, Catania, 6–8 April 1994 : proceedings of the conference*. – Berlin : Springer, 1994. – P. 171–182. DOI: 10.1007/3-540-57868-4_57.
18. Kira K. A practical approach to feature selection / K. Kira, L. Rendell // *Machine Learning : International Conference on Machine Learning ML92, Aberdeen, 1–3 July 1992 : proceedings of the conference*. – New York: Morgan Kaufmann, 1992. – P. 249–256. DOI: 10.1016/B978-1-55860-247-2.50037-1.
19. Salfner F. A survey of online failure prediction methods / F. Salfner, M. Lenk, M. Malek // *ACM computing surveys*. – 2010. – Vol. 42, Issue 3. – P. 1–42. DOI: 10.1145/1670679.1670680.
20. Shin Y. C. Intelligent systems : modeling, optimization, and control / C. Y. Shin, C. Xu. – Boca Raton : CRC Press, 2009. – 456 p. DOI: 10.1201/9781420051773.
21. Oliinyk A. A. Information Technology of Diagnosis Model Synthesis Based on Parallel Computing / [A. A. Oliinyk, S. A. Subbotin, S. Yu. Skrupsky et al]// *Радіоелектроніка, інформатика, управління*. – 2017. – № 3. – С. 139–151.
22. Субботін С. О. Нейтеративні, еволюційні та мультиагентні методи синтезу нечіткологічних і нейромережних моделей: монографія / С. О. Субботін, А. О. Олійник, О. О. Олійник ; під заг. ред. С. О. Субботіна. – Запоріжжя : ЗНТУ, 2009. – 375 с.
Article was submitted 27.10.2017.
After revision 15.11.2017.

Олійник А. О.¹, Субботін С. О.², Львівкін В. М.³, Благодарьов О. Ю.⁴, Зайко Т. А.⁵

¹Канд.техн.наук, доцент кафедри програмних засобів, Запорізький національний технічний університет, Запоріжжя, Україна

²Д-р. техн. наук, завідувач кафедри програмних засобів, Запорізький національний технічний університет, Запоріжжя, Україна

³Канд.техн.наук, доцент кафедри програмних засобів, Запорізький національний технічний університет, Запоріжжя, Україна

⁴Аспірант кафедри програмних засобів, Запорізький національний технічний університет, Запоріжжя, Україна

⁵Канд.техн.наук, доцент кафедри програмних засобів, Запорізький національний технічний університет, Запоріжжя, Україна

СИСТЕМА КРИТЕРІЇВ ОЦІНЮВАННЯ ІНФОРМАТИВНОСТІ ОЗНАК ДЛЯ РОЗПІЗНАВАННЯ ОБРАЗІВ

Актуальність. Вирішено задачу автоматизації процесу оцінювання інформативності ознак при розв'язанні завдань діагностування та розпізнавання образів. Об'єкт дослідження – процес відбору інформативних ознак. Предмет дослідження – критерії оцінювання інформативності ознак.

Мета роботи полягає в створенні системи критеріїв оцінювання інформативності ознак, що дозволяє обчислювати інформативність наборів взаємозалежних ознак.

Метод. Запропоновано систему критеріїв оцінювання інформативності ознак. Запропонована система передбачає визначення значущості ознак виходячи з просторового розташування екземплярів різних класів (діапазонів зміни значень вихідного параметра). Розроблена система критеріїв дозволяє оцінювати індивідуальну і групову інформативність ознак при вирішенні задач класифікації і регресії в умовах, коли вихідні вибірки даних містять надлишкові і взаємозалежні ознаки, а також екземпляри з пропущеними значеннями. Запропоновані критерії не вимагають побудови моделей на основі оцінюваних комбінацій ознак, що істотно знижує часові і обчислювальні витрати в процесі відбору ознак. Використання запропонованих критеріїв для оцінювання та відбору інформативних ознак дозволяє при вирішенні завдань діагностування та розпізнавання образів знижувати структурну складність синтезованих діагностичних і розпізнавальних моделей, підвищувати їх інтепретовність і узагальнюючі властивості за рахунок виключення малозначущих, взаємозалежних і надлишкових ознак.

Результати. Розроблено програмне забезпечення, що реалізує запропоновану систему критеріїв оцінювання інформативності ознак і дозволяє виконувати відбір ознак для синтезу розпізнавальних моделей на основі заданих наборів даних.

Висновки. Проведені експерименти підтвердили працездатність запропонованої системи критеріїв оцінювання інформативності ознак і дозволяють рекомендувати її для використання на практиці при обробці масивів даних для розпізнавання образів. Перспективи подальших досліджень можуть полягати в модифікації існуючих і розробці нових методів відбору ознак на основі запропонованої системи критеріїв оцінювання індивідуальної і групової інформативності ознак.

Ключові слова: вибірка даних, розпізнавання образів, відбір ознак, критерій інформативності, індивідуальна інформативність, групово інформативність.

Олейник А. А.¹, Субботин С. А.², Левкин В. Н.³, Благодарев А. Ю.⁴, Зайко Т. А.⁵

¹Канд. техн. наук, доцент кафедры программных средств, Запорожский национальный технический университет, Запорожье, Украина

²Д-р. техн. наук, заведующий кафедрой программных средств, Запорожский национальный технический университет, Запорожье, Украина

³Канд. техн. наук, доцент кафедры программных средств, Запорожский национальный технический университет, Запорожье, Украина

⁴Аспирант кафедры программных средств, Запорожский национальный технический университет, Запорожье, Украина

⁵Канд. техн. наук, доцент кафедры программных средств, Запорожский национальный технический университет, Запорожье, Украина

СИСТЕМА КРИТЕРИЕВ ОЦЕНИВАНИЯ ИНФОРМАТИВНОСТИ ПРИЗНАКОВ ДЛЯ РАСПОЗНАВАНИЯ ОБРАЗОВ

Актуальность. Решена задача автоматизации процесса оценивания информативности признаков при решении задач диагностирования и распознавания образов. Объект исследования – процесс отбора информативных признаков. Предмет исследования – критерии оценивания информативности признаков.

Цель работы заключается в создании системы критериев оценивания информативности признаков, позволяющей вычислять информативность наборов взаимозависимых признаков.

Метод. Предложена система критериев оценивания информативности признаков. Предложенная система предполагает определение значимости признаков исходя из пространственного расположения экземпляров разных классов (диапазонов изменения значений выходного параметра). Разработанная система критериев позволяет оценивать индивидуальную и групповую информативность признаков при решении задач классификации и регрессии в условиях, когда исходные выборки данных содержат избыточные и взаимозависимые признаки, а также экземпляры с пропущенными значениями. Предложенные критерии не требуют построения моделей на основе оцениваемых комбинаций признаков, что существенно снижает временные и вычислительные затраты в процессе отбора информативных признаков. Использование предложенных критериев для оценивания и отбора информативных признаков позволяет при решении задач диагностирования и распознавания образов понижать структурную сложность синтезируемых диагностических и распознающих моделей, повышать их интерпретируемость (понимаемость человеком) и обобщающие способности за счет исключения малозначимых, взаимозависимых и избыточных признаков.

Результаты. Разработано программное обеспечение, которое реализует предложенную систему критериев оценивания информативности признаков и позволяет выполнять отбор информативных признаков для синтеза распознающих моделей на основе заданных наборов данных.

Выводы. Проведенные эксперименты подтвердили работоспособность предложенной системы критериев оценивания информативности признаков и позволяют рекомендовать ее для использования на практике при обработке массивов данных для распознавания образов. Перспективы дальнейших исследований могут заключаться в модификации существующих и разработки новых методов отбора признаков на основе разработанной системы критериев оценивания индивидуальной и групповой информативности признаков.

Ключевые слова: выборка данных, распознавание образов, отбор признаков, критерий информативности, индивидуальная информативность, групповая информативность.

REFERENCES

- Jensen R., Shen Q. Computational intelligence and feature selection: rough and fuzzy approaches. Hoboken, John Wiley & Sons, 2008, 339 p. DOI: 10.1002/9780470377888.
- Mulaik S. A. Foundations of Factor Analysis. Boca Raton, Florida, CRC Press, 2009, 548 p.
- Lee J. A., Verleysen M. Nonlinear dimensionality reduction. New York, Springer, 2007, 308 p. DOI: 10.1007/978-0-387-39351-3.
- Bezdek J. C. Pattern Recognition with Fuzzy Objective Function Algorithms. N.Y., Plenum Press, 1981, 272 p. DOI: 10.1007/978-1-4757-0450-1.
- Hyvarinen A., Karhunen J., Oja E. Independent component analysis. New York, John Wiley & Sons, 2001, 481 p. DOI: 10.1002/0471221317.
- Fedotov N. G. Teorija priznakov raspoznavanija obrazov na osnove stohasticheskoj geometrii i funkcional'nogo analiza. Moscow, Fizmatlit, 2010, 304 p. (In Russian).
- Guyon I., Elisseeff A. An introduction to variable and feature selection, *Journal of machine learning research*, 2003, No. 3, pp. 1157–1182.
- McLachlan G. Discriminant Analysis and Statistical Pattern Recognition. New Jersey, John Wiley & Sons, 2004, 526 p. DOI: 10.1002/0471725293.
- Oliinyk A. A., Skrupsky S. Yu., Shkarupylo V. V., Blagodariev O. Parallel multiagent method of big data reduction for pattern recognition, *Radio Electronics, Computer Science, Control*, No. 2. 2017, pp. 82–92.
- Oliinyk A. Production rules extraction based on negative selection, *Radio Electronics, Computer Science, Control*, 2016, Vol. 1, pp. 40–49. DOI: 10.15588/1607-3274-2016-1-5.
- Oliinyk A., Skrupsky S., Subbotin S., Blagodariev O., Gofman Ye. Parallel computing system resources planning for neuro-fuzzy models synthesis and big data processing, *Radio Electronics, Computer Science, Control*, 2016, Vol. 4, pp. 61–69. DOI: 10.15588/1607-3274-2016-4-8.
- Oliinyk A. A., Skrupsky S. Yu., Shkarupylo V. V., Subbotin S. A. The model for estimation of computer system used resources while extracting production rules based on parallel computations, *Radio Electronics, Computer Science, Control*, 2017, No. 1, pp. 142–152. DOI: 10.15588/1607-3274-2017-1-16.
- Subbotin S., Oliinyk A. The Sample and Instance Selection for Data Dimensionality Reduction, *Recent Advances in Systems, Control and Information Technology. Advances in Intelligent Systems and Computing*, 2017, Vol. 543, pp. 97–103. DOI: 10.1007/978-3-319-48923-0_13.
- Shitikova O. V., Tabunshchik G. V. Method of Managing Uncertainty in Resource-Limited Settings, *Radio Electronics, Computer Science, Control*, 2015, No. 2, pp. 87–95. DOI: 10.15588/1607-3274-2015-2-11.
- Tabunshchik G. V., Kaplienko T. I., Shitikova O. V. Verification model of systems with limited resources, *Radio Electronics, Computer Science, Control*, 2017, No. 4.
- Bodyanskiy Ye., Vynokurova O. Hybrid adaptive wavelet-neuro-fuzzy system for chaotic time series identification, *Information Sciences*, 2013, Vol. 220, pp. 170–179. DOI: 10.1016/j.ins.2012.07.044.
- Kononenko I. Estimating Attributes: Analysis And Extensions Of Relief, *Machine Learning : European Conference on Machine Learning ECML-94, Catania, 6–8 April 1994 : proceedings of the conference*. Berlin, Springer, 1994, pp. 171–182. DOI: 10.1007/3-540-57868-4_57.
- Kira K., Rendell L. A practical approach to feature selection, *Machine Learning : International Conference on Machine Learning ML92, Aberdeen, 1–3 July 1992 : proceedings of the conference*. New York, Morgan Kaufmann, 1992, pp. 249–256. DOI: 10.1016/B978-1-55860-247-2.50037-1.
- Salfner F., Lenk M., Malek M. A survey of online failure prediction methods, *ACM computing surveys*, 2010, Vol. 42, Issue 3, pp. 1–42. DOI: 10.1145/1670679.1670680.
- Shin Y. C. Intelligent systems : modeling, optimization, and control / C. Y. Shin, C. Xu. – Boca Raton, CRC Press, 2009, 456 p. DOI: 10.1201/9781420051773.
- Oliinyk A. A., Subbotin S. A., Skrupsky S. Yu., Lovkin V. M., Zaiko T. A. Information Technology of Diagnosis Model Synthesis Based on Parallel Computing, *Radio Electronics Computer Science Control*, 2017, No. 3, pp. 139–151.
- Subbotin S., Oliinyk A., Oliinyk O. Noniterative, evolutionary and multi-agent methods of fuzzy and neural network models synthesis : monograph. Zaporizhzhya, ZNTU, 2009, 375 p. (In Ukrainian).