Machine Learning Methods in Medicine Diagnostics Problem

Strilets Viktoriia[1[0000-0002-2475-1496]], Bakumenko Nina[1[0000-0003-3496-7167]], Donets Volodymyr[1[0000-0002-5963-9998]], Chernysh Serhii[2[0000-0002-1750-5158]], Ugryumov Mykhaylo[1[0000-0003-0902-2735]], Goncharova Tamara[3[0000-0003-3210-3867]]

[1] V. N. Karazin Kharkiv National University, Kharkiv, Ukraine
[2] National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine
[3] National University of Civil Defence of Ukraine, Kharkiv, Ukraine
striletsvictoria@gmail.com, n.bakumenko@karazin.ua, vovan.s.marsa@gmail.com, 91sergey@gmail.com, ugryumov.mykhaylo52@gmail.com, super-gusenichka@ukr.net

Abstract. Medical service improvement has always been a life topical problem. To decide it, we must continuously raise the competency of doctors on the one hand and it is necessary to develop new methods and approaches which could help take decisions concerning diagnostics (classification) of patient health conditions and concerning patient's further treatment.

At the paper the machine learning methods for patient health condition classification were considered. These methods were Naive Bayes Classifier, Linear Classifier, Support-vector machine, K-nearest Neighbor Classifier, Logistic Regression, Decision Tree Classifier, Random Forest Classifier, Ada Boost Classifier and Artificial Neural Network. A radial basis network was chosen from the variety of artificial neural system architecture to solve classification tasks.

The problem of patient health conditions classification was considered for two sets of laboratory research results: on liver diseases and on urological diseases.

Confusion matrixes and ROC-curves were taken to estimate classification quality of patient health conditions with above-mentioned methods.

Keywords: medicine diagnostic, machine learning, artificial neural network, ROC-curve, confusion matrix.

1 Classification Methods of the Complex Dynamic System State

Origin and evolution of errors in complex systems are the complex dynamic process. Experts cannot always predict their origin exactly and define the type of failure. However, bringing the system back to normal mode is paramount importance. Control and prediction over the state of complex dynamic systems help specialists take effective measures. Thus, great attention is paid to the classification of complex system states. Today, we have many publications describing methods of settling the problem

of complex system states classification. We'll consider the ones which are mostly frequently viewed.

Naïve Bayes Classifier is a simple probabilistic classifier, which relies on applying the Bayes theorem with 'naïve' assumption of mutual sign independence. It applies to the simplest models of the Bayes network. Naïve Bayes method developments started in the distant 1960s and are still a popular method of text categorization (e.g. a scientific text, a fiction literature text, spam etc.) [1]. This method is also applied for the automated medical diagnostics [2].

Advantages of the method:

- in the little data sets, it can achieve better results than other classifiers because of low tendency to retraining;
- linear scalability on the quantity of possible signs; a slight renovation on new educational data is also possible;
- the method can process overlooked data by retraining and forecasting;
- despite the fact that the assumption about sign independence is often false, the Bayes classifier can independently estimate each sign class; it makes it possible to avoid the problem of large dimension [3].

Disadvantages of the method:

- Naïve Bayes evidently assumes that all the signs are mutually independent, which is almost impossible in reality;
- if the variable has the textual data set category which was not observable in the learning data set, the model will set the probability 0 and won't be able to make forecasting;
- quality of work is sensitive to the class distribution representativeness in the total package.

Linear classifier is a method of machine learning, which takes decisions about the class by relying on the linear behavior combination values; they are usually represented as a vector. Subdiving into classes in the multidimensional space can be made by dividing with a straight or n-dimensional plane. These classifiers work well for practical problems, e.g. for document classification. Moreover, the method can achieve the non-liner classifier results within less time on learning and use [4].

Advantages of the method:

- linear classifier is often used when the speed of classification matters, since it is the fastest classifier, especially if the input vector is very big;
- fast method realization and low requirements for operative memory and the central processor.

Disadvantage of the method: linear character of the method doesn't make it possible to define the class exactly when it is impossible to clearly discriminate between the classes, as data distribution is usually mixed and demands non-linear separation.

Support-vector machine (SVM) is a controlled learning model with a tutor, and usually used for classification and regression analysis. The method was proposed by

Vapnik V. and Chervoneniks A. in 1963. Allowing for the set of educational examples, each of which earlier attributed to one of the two categories, the learning algorithm SVM builds a model, which can assign a new example to a specific category. The SVM model is a presentation of examples as points in space, represented so that the examples of separate categories are divided into discrete highest possible intervals. Then, new examples are represented in the same space; they are assumed to be in the category which is based on that side of the space where they belong [5].

SVM can be used for solving various real tasks:

- for categorization of a text and hypertext, since it diminishes the need in marked learning data [6];
- for image classification [7];
- for cursive identification [6];
- in biology and other sciences. It was used for protein classification and gave 90% classification correctness [6].

Advantages of the method:

- retraining problem is not so important as with other methods;
- SVM doesn't heavily depend on computer memory;
- SVM works rather effectively in the cases when the task dimension exceeds the number of examples.

Disadvantages of the method [8]:

- the methods are characterized by high calculation complexity. As compared to other simple methods (K-NN, Decision Tree, Naïve Bayes Classifier), the method requires more time for learning;
- the major problem is the choice of the most appropriate central function. Various central functions give different results for each data set;
- SVM has bad results in the case of noise present (target classes have no distinct partition boundary).

Relevance Vector Machine (RVM) is a method of machine learning, which uses a Bayes conclusion to obtain decisions on the principle of economy for regressive and probabilistic classification [9].

Advantages of the method:

- RVM has the identical functional form with SVM, but it provides probabilistic classification;
- Bayes RVM base makes it possible to avoid SVM independent parameter sets, which generally require post-optimization, based on cross check.

The main disadvantage of the method is that it employs a learning method, resembling expectance maximization, so it may give a local extremum. At the same time, standard algorithms on the basis of successive minimum, used in SVM, will find global extremum.

K-Nearest Neighbor Classifier (k-NN) is a nonparametric method used for classification and regression. In both cases, the input consists of k nearest educational examples in a function space. In B classification, k-NN output is a notion of class. The object is classified on the majority vote with its nearest neighbors. At this time, the object is assigned to the class which prevails over its nearest neighbors (k is a whole number, as a rule, small), If k equals 1, the object is simply assigns to the class of this nearest exclusive neighbor [10].

Neighbors are taken from the set of objects, for which the class (for k-NN classification) or object property value (for k-NN regression) is known. It is considered to be the learning algorithm set, though a distinct preparation step is not needed [10].

The specific feature of k-NN method is that it is not sensitive to the data local structure [10].

Advantages of the method:

- absence of the education step. The method saves the learning data set and learns only in real time forecasting. It makes the algorithm k-NN much faster than other ones which require learning, for example, SVM, Linear Regression etc.;
- new data are easy to add because the k-NN algorithm doesn't require preparation; this won't influence the method accuracy;
- k-NN is very simple to realize, it needs only two parameters: the number of classes k and the distance function (e.g. Euclidean, Manhattan etc.).

Disadvantages of the method:

- it works poorly with large data sets. In large data sets, the complexity of calculating the distance between the new point and each existing point is enormous; it worsens algorithm efficiency;
- it requires function scaling (standardization and normalization) before applying the k-NN algorithm to any data set. If we don't do this, k-NN may generate wrong forecasts;
- it is sensitive to noisy data, absence of values. It is necessary to inscribe omitted values or erase remainders manually.

Logistic regression is a statistical model, which uses the logistic function for modelling binary dependent variable in its basic form. The logistic regression measures interdependence between categorically dependent variable and one or several independent variables by evaluating probabilities with a logarithmic function [11].

Logistic regression can be considered as a special case of generalized linear model and, thus, similar to the linear regression. However, the logistic regression model is based on the assumptions of dependent and independent variable interdependence. The key differences between these models can be seen in the following two logistic regression peculiarities. Firstly, conditional distribution (y|x) is a Bernoulli distribution, but not gaussian, because the dependence curve is binary. Secondly, the forecast values are probable and, thus, are limited (0,1) through a logistic distribution function because logistic regression assumes concrete result probability, but not pure results.

Advantages of the method [12]:

- logistic regression works well when the data set is linearly separable (which is common for k-NN and Linear Regression);
- logistic regression has less tendency for retraining, but retraining can appear in big dimension data sets: regularization methods are generally used to solve this problem;
- logistic regression can forecast not only the final class, but show the interconnection between input data and a resulting class;
- logistic regression method is simple in realization, interpreting, and is effective in learning.

Disadvantages of the method [12]:

- the principal limitation is the assumption about linearity between the dependent variable and independent variables:
- if the quantity of observations is less than the quantity of variables, the logistic regression shouldn't be used because this can lead to overlearning:
- logistic regression can be used just to forecast discrete functions. Therefore, the dependent logistic regression variable is limited by a discrete number set.

Decision Tree Classifier [13] is the machine learning method, which uses a decision tree model for classification. The tree model, where the target variable can take a discrete value set, is also called the classification tree. In these tree structures, the leaves represent the class marks and the branches – the combination of signs leading to these class marks. The decision trees, where the target variable can take permanent values (real numbers, as a rule) are called the regression trees.

Advantages of the method:

- decision tree is simple to understand, is easy to represent graphically [13];
- it is capable of processing numerical and categorical data as well;
- it requires a small data preparation. Other methods often demand data normalization. Fictitious variables are not necessary here because the trees can work with qualitative forecasts;
- possibility for checking the models by statistic tests. It makes it possible to take into account the model reliability;
- a non-statistical approach, which doesn't foresee assumptions concerning learning data and forecast remainders e. g. no assumptions as to distribution, independence or constant dispersion;
- it works well with big data sets;
- the mirror of human decision taking is closer than other approaches [13]. This may be useful when modelling human decisions / behavior.

Disadvantages of the method:

- the trees can be very unstable. A little change of learning data can lead to the change of a tree and consequently of final forecasts [13];
- it is known that the problem of studying the decision optimal tree is NP-complete in several optimality aspects and even for simple conceptions [14]. Thus, the learn-

ing algorithms of the decision practical tree are based on the heuristic, such as a greedy algorithm, where the locally optimal decisions are taken at every unit. Such algorithms cannot guarantee obtaining optimal decisions on the whole decision tree. To decrease the locally optimum greedy algorithm, e. g. the double information distance tree was suggested [15];
- for the data including categorical variables with different level quantities, the decision tree information gain is biased in favor of big level attributes. However, the problem of biased choice is resolved, e.g. by a conditional conclusion approach.

Random Forest Classifier is an ensemble method of classification, regression, which works by constructing many decision trees during learning and withdrawal of the class which is a regime class (classification) or average forecast (regression) of separate trees [16, 17]. The first algorithm of random decision forests was created by Tin Kam Ho [16] on the basis other random subspace method [17], which, as formulated by Ho, is the means of realizing "stochastic discrimination" approach to the classification proposed by Eugene Kleinberg [18].

Advantages of the method:

- random forest is based on the stacking algorithm and uses the ensemble learning technics;
- random forest works well both with categorical and persistent variables;
- random forest can automatically process missing values;
- it doesn't demand function scaling (standardization and normalization), since this method demands the approach based on rules instead of calculating distance;
- the random forest algorithm is stable.

Disadvantages of the method:

- complexity. Random forest creates many trees (unlike only one tree in the case of a decision tree) and comprises their results: e.g. on default it creates 100 trees in the sklearn Python library. This method demands much more computing power and resources;
- longer learning time: random forests require much more time for preparation compared to decision trees, since they generate many trees (instead one tree in the case of a decision tree) and take decision by a majority of votes.

AdaBoost Classifier is the machine learning meta-algorithm formulated by Ioav Freindom and Robert Shapiro. It can be used with many other classification algorithms for productivity improvement. The output of other classification algorithms (weak classifiers) is assembled into a weighted sum, which is a finite output of the accelerated classifier. AdaBoost is adaptable: weak classifiers are adjusted in favor of those cases which were classified earlier. AdaBoost is not very sensitive to the data noise. In some tasks it can be less sensitive to the retraining problem, than other learning algorithms.

Each classification algorithm usually corresponds to some types of tasks better than others and, as a rule, has great number of various parameters and configurations, which must be corrected before achieving optimal data set productivity. AdaBoost,

alongside with decision trees as weak classifiers, are often called the best classifier [19]. When using decision trees, the information collected in every phase of the AdaBoost algorithm about the relative rigidity of each learning pattern is put to the tree building algorithm, so that the later trees, as a rule, are concentrated on more important examples to classify.

Advantages of the method:

- weak classifiers for cascading are easy to use;
- various classification algorithms can be used as weak classifiers;
- AdaBoost has high accuracy;
- AdaBoost isn't sensitive to the data noise.

Disadvantages of the method:

- quantity of AdaBoost iterations are also determined by the quantity of weak classifiers, which can be defined with cross check;
- data imbalance results in lower classification accuracy;
- learning takes longer time.

Artificial Neural Network (ANN) is a computing system, which is inspired by biological neural networks. Such systems 'learn' to decide tasks, considering examples and, as a rule, are not programmed to perform concrete tasks.

ANN is based on the combination of coupled units or packs called artificial neurons, which freely model neurons in the biological brain.

The primary aim of ANN approach consisted in solving problems as the human brain does. However, in due course, ANN application shifted to deciding variety of tasks, including computer vision, speech recognition, machine translation, filtering in the social network, game boards and video-games, medical diagnosis, and even in schools which are traditionally considered to be human activities (e.g. painting) [20].

Advantages of the method [21]:

- information storage on the whole network. Disappearance of some information fragments in one place don't impede the network function;
- ability to work with insufficient knowledge;
- fault tolerance: damage of one or several ANN cells won't impede the data output;
- possibility to learn: artificial neural networks study events and take decisions, using such events;
- possibility of parallel processing.

Disadvantages of the method [21]:

- estimation of proper network structure: there aren't concrete rules to define structure of artificial neural networks. The network structure is chosen by relying on the practical experience or trial-and-error method;
- ANN can work only with numerical data. The data must be converted into numerical values before introducing into ANN.

To solve the classifying problem of the patient's state the Radial Basis Function Network was chosen. In this Network a multiple logistic regression was used [22]. This allows classification by more than two classes. And as the learning algorithm the stochastic approximation algorithm with deep learning elements based on the ravine conjugate gradient method was used. This Radial Basis Function Network was independently implemented by the authors in the "ROD&IDS®" computer decision support system, designed to solve the problem of diagnosing, classifying and optimizing systems and processes.

2 Problem Statement

Let the condition multidimensional matrix be known $X = \{x_{i,j}\}$, $(i = 1..I, j = 1..J)$, where I is the quantity of checked patients and J is the quantity of state characteristics (variable) to be measured. The majority of the examined methods require normalizing input data; centering and normalizing are done according to the formula

$$x^0_{i,j} = \left(x_{i,j} - \left\langle X_j \right\rangle\right) \Big/ \sigma_j ,$$

where $<X_j>$ is the average of j-state attribute, σ_j is its quadratic mean deviation. The task of building a classification model of the patient state: the vector function is given by a set of learning couples $\left(\vec{X}^{(0)}, \vec{d}\right)_p$, p=1..P, with input dimension vectors H_0 and output dimension H_{k+1}. It is necessary to build the mathematical vector function $\vec{Y}^{(K+1)}\left(\vec{X}^{(0)}\right)$ for the input data approximation.

We formulate the classification problem. Let $\vec{X}^*$ be the variable vector, which describes the state of a patient and M – multitude of scenarios (possible state classes). According to the values of $\vec{X}^*$ vector, the current state is related to one of the multitudes R_m, where m=0..M-1. It is necessary to find such m-scenario, for which the maximal distribution density of the conditional appearance probability in m-scenario:

$$\exists! m^* \in C_m\left(\rho\left(\vec{X}^*_m \middle| R_m\right)\right)(m = 0..M-1): \rho\left(\vec{X}^*_m \middle| R_m\right) \to \max ,$$

where $C_m\left(\rho\left(\vec{X}^*_m \middle| R_m\right)\right)$ is the multitude of m-indices of distribution density of the conditional appearance probability in m-scenario.

Let us consider the medical-biological system. The final state of patients and a set of parameters describing it are characteristic for each medical treatment stage. Take the hypothesis that the state of a patient is definitely defined by this set of parameters. Therefore, the task of checking health state is reduced to the task of classification of patient status variables. Let's examine the application of above-mentioned methods for solving the task of patient state variable classification. We'll estimate and compare the quality of classification by these methods.

3 Methods of Estimating Classification Quality

A confusion matrix is used in machine learning to solve classification problems for productivity visualization and algorithm work quality (usually learning with a teacher) [23]. Each line of the fault matrix is a copy in a forecast class and each column is a copy in a real class 9 (or vice versa). The matrix name comes from the fact that it helps vividly see whether the resultant classes are mixed or not, i.e. whether the one class is defined as the other.

We consider fault matrix building for the problem of binary classification. Let the classification result be designated as positive (p) and negative (n). The binary classifier has four possible results. If the classification result is p and the actual meaning is p, then the result is called real positive (TP). If the classification result is p and the actual meaning is n, then the result is called confusion (fault) positive (FP). Similarly, the result is called real negative (TN), if the classification result and real meaning are n, and it is called confusion negative (FN) if the classification result is n, but the real meaning is p.

Suppose we carried out an experiment for P positive copies and N negative cases. The classification results can be summarized in in the fault matrix, shown in Fig. 1.

ROC curves are also used to estimate classification quality. An ROC curve is a diagram which helps estimate binary classification quality. It is defined by the correlation between the quantity of objects from the total amount of sign media classified as true sign media (classification algorithm sensitivity) and the number of objects from the total amount of sign media with no sign, classified mistakenly as sign media [24].

		True condition				
	Total population	Condition positive	Condition negative	Prevalence $= \frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	Accuracy (ACC) $= \frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$	
Predicted condition	Predicted condition positive	**True positive**	**False positive**, Type I error	Positive predictive value (PPV), Precision $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$	False discovery rate (FDR) $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$	
	Predicted condition negative	**False negative**, Type II error	**True negative**	False omission rate (FOR) $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$	Negative predictive value (NPV) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$	
		True positive rate (TPR), Recall, Sensitivity, probability of detection, Power $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Positive likelihood ratio (LR+) $= \frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) $= \frac{\text{LR+}}{\text{LR−}}$	F_1 score $= 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$
		False negative rate (FNR), Miss rate $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$	Specificity (SPC), Selectivity, True negative rate (TNR) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	Negative likelihood ratio (LR−) $= \frac{\text{FNR}}{\text{TNR}}$		

Fig. 1. Possible meanings defined by the confusion matrix.

ROC quantitative interpretation gives an AUC indicator; it is the area limited by a ROC curve and the axis, which equals to fault positive classifications. The higher AUC indicator the better a classifier works. The value less than 0.5 shows that the

classifier acts vice-versa: in the case of positive classifications it calls them negative, and the negative classification is represented as positive [23]. There exist many classifications of ROC curves for classification estimations according to more than 2 classes and also the ones which with a diagram help estimate the drawbacks of the current classification modes. Fig. 2 shows two diagrams, which characterize work of two classification algorithms.

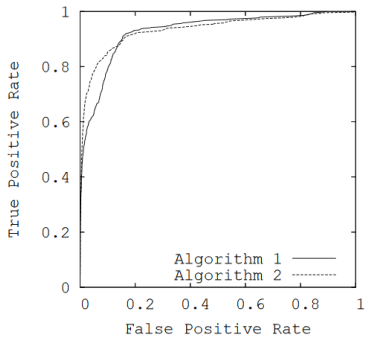


Fig. 2. ROC-curves which compare work of two algorithms.

The diagram clearly shows which class was better recognized as apposite, which is suitable for model classification adjustment.

4 Methods Comparison for Medical-Biological System State Classification

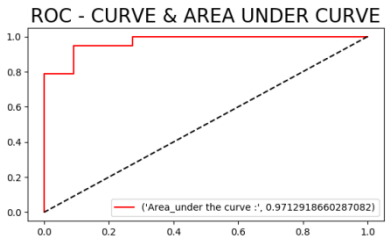
Let's examine a patient within a period of medical treatment. To make the diagnosis more accurate, we formulate the problem of patient state classification: to define the current state of the patient (healthy or sick) according to the laboratory research records and primary health examination. The problem was solved for two data sets on liver disease and urological disease. These data provided by the Department of Infectious, Pediatric and Oncological Urology, Kharkiv Medical Academy of Postgraduate Education.

The urological disease sampling contained information for 40 patients. These data were divided into learning (30) patients and testing (10 patients). The information for one patient consisted of 47 estimated characteristics with the values of three types: real, Boolean and enumerated numbers.

The liver disease sampling consisted of the information for 590 patients. Learning sampling was taken from 420 patients, testing sampling – from 170 patients. The information for one patient consisted of 10 estimated characteristics with the values of three types: real, Boolean and enumerated numbers.

To solve classification problems, we used Naïve Bayes Classifier, K-nearest Neighbor Classifier, Logistic Regression, Random Forest Classifier, Ada Boost Classifier and Radial Basis Function Network.

Learning and testing problems made it possible to error matrices and ROC-curves for classification quality analysis. For example, on fig. 3 and 4 the error matrices and ROC-curves of the Naïve Bayes Classifier results are presented.

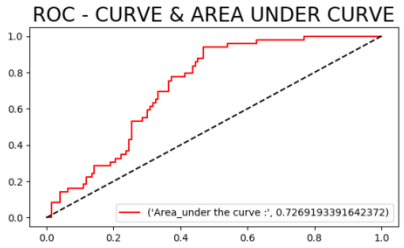


Classifier results / Truth data	Class 1	Class 2	Classification overall	Producer Accuracy (Precision)
Class 1	11	0	11	100%
Class 2	8	11	19	57.895%
Truth overall	19	11	30	
User Accuracy (Recall)	57.895%	100%		

Overall accuracy (OA): 73.333%

Fig. 3. ROC-curve and confusion matrix based on the method Naïve Bayes Classifier for patients with urological diseases.

For the first data set on urological diseases, Logistic Regression, AdaBoost Classifier and Radial Basis Function Network give 100% classification accuracy with ROC AUC=1, the Random Forest Classifier method – 96.6% classification accuracy and ROC AUC=1, Naïve Bayes Classifier – 73.3% classification accuracy and ROC AUC=0.97, K-nearest Neighbor Classifier – 80% classification accuracy with ROC AUC=0.82.



Classifier results / Truth data	Class 1	Class 2	Classification overall	Producer Accuracy (Precision)
Class 1	58	68	126	46.032%
Class 2	2	47	49	95.918%
Truth overall	60	115	175	
User Accuracy (Recall)	96.667%	40.87%		

Overall accuracy (OA): 60%

Fig. 4. ROC-curve and confusion matrix based on the method of Naïve Bayes Classifier for patients with liver diseases.

The second data set for liver diseases cardinally differs from the first in dimensions (4 times less attributes, but 18 times more recordings). For classification, we used the same methods. The obtained results are: Naïve Bayes Classifier method – 60% classification accuracy and ROC AUC=0.73, K-nearest Neighbor Classifier – 81.7% classification accuracy and ROC AUC=0.898, Logistic Regression – 80.98% classification accuracy and ROC AUC=0.787, Random Forest Classifier – 98,86% classification accuracy and ROC AUC=0.99, Ada Boost Classifier – 85.7% classification accuracy

and ROC AUC=0.94, Radial Basis Function Network – 80.56% classification accuracy and ROC AUC=0.801.

Thus, the most qualitative data classification about the state of patients' status was given by Random Forest Classifier method, it showed high accuracy and ROC AUC indicator for both data sets.

5 Conclusions

Diagnosing complex dynamic system states, e. g. medical-biological system (patient), faces the problems of system state classification. The work studied methods of deciding classification tasks, such as Naïve Bayes Classifier, K-nearest Neighbor Classifier, Logistic Regression, Random Forest Classifier, Ada Boost Classifier and Artificial Neural Network (Radial Basis Function Network architecture). Confusion matrices and ROC-curves were taken for quality classification estimation.

The Radial Basis Function Network differs from the classical one in that it uses multivariate logistic regression and a recurrent learning algorithm with deep learning elements. The Network application allows to not depend on the data type and expert opinion during making decisions.

As an example, we considered two data sets, which characterized the state of patients with liver and urological diseases. As a result, all the methods gave classification accuracy more than 80% except for Naïve Bayes Classifier. Radial Basis Function Network showed the best classification quality with 100% accuracy for urological diseases and the Random Forest Classifier method showed the best classification quality with 98.86% for liver diseases.

Further, we are planning to test the methods which showed the best classification for other data sets with different dimensions. The authors are also working out a modification of the method by using Radial Basis Function Network to improve its accuracy for various input data.

References

1. Maron, M.E.: Automatic Indexing: An Experimental Inquiry. Journal of the ACM 8(3), 404–417 (1961).
2. Rish, I.: An empirical study of the naive Bayes classifier. IJCAI Workshop on Empirical Methods in AI (2001).
3. Niculescu-Mizil, A., Caruana, R.: Predicting good probabilities with supervised learning. ICML (2005).
4. Guo-Xun Yua, Chia-Hua Ho, Chih-Jen Lin: Recent Advances of Large-Scale Linear Classification. Proc. IEEE, 100(9) (2012).
5. Cortes, C., Vapnik, V.N. Support-vector networks. Machine Learning, 20(3), 273–297 (1995).
6. Pradhan, Sameer S., et al.: Shallow semantic parsing using support vector machines. Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004 (2004).

7. Barghout, L.: Spatial-Taxon Information Granules as Used in for Image Segmentation. Granular Computing and Decision-Making. Springer International Publishing, 285–318 (2015).
8. Divya, T.: A survey on Data Mining approaches for Healthcare. International Journal of Bio-Science and Bio-Technology, 5(5), 241-266 (2013).
9. Tipping, M.E.: Sparse Bayesian Learning and the Relevance Vector Machine. Journal of Machine Learning Research, 1, 211–244 (2001).
10. Altman, N.S.: An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician, 46(3), 175–185 (1992).
11. Rodríguez, G.: Lecture Notes on Generalized Linear Models. Chapter 3, P. 45. (2007).
12. Kumar, N.: Advantages and Disadvantages of Logistic Regression in Machine Learning. The Professionals Point (2019).
13. Gareth, J., Witten, D., Hastie, T., Tibshirani, R.: An Introduction to Statistical Learning. Springer, New York, P. 315. (2015).
14. Hyafil, L., Rivest, R.L.: Constructing Optimal Binary Decision Trees is NP-complete. Information Processing Letters, 5(1), 15–17 (1976).
15. Ben-Gal I., Dana A., Shkolnik N.: Efficient Construction of Decision Trees by the Dual Information Distance Method. Quality Technology & Quantitative Management, 11(1), 133–147 (2014).
16. Ho, Tin Kam: Random Decision Forests. Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, pp. 278–282. (1995).
17. Ho, T.K.: The Random Subspace Method for Constructing Decision Forests. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(8), 832–844 (1998).
18. Kleinberg, E.: Stochastic Discrimination. Annals of Mathematics and Artificial Intelligence, 1(1–4), 207–239 (1990).
19. Kégl, B.: The return of AdaBoost.MH: multi-class Hamming trees, (2013).
20. Bethge, M., Ecker, A.S., Gatys, L.A.: A Neural Algorithm of Artistic Style (2015).
21. Schmidhuber, J.: Deep Learning in Neural Networks: An Overview. Neural Networks, 61, 85–117 (2015).
22. Strilets, V., Bakumenko, N., Chernysh, S. ets. Application of the c-means fuzzy clustering method for the patient's state recognition problems in the medicine monitoring system. Intelligent Systems and Computing Integrated Computer Technologies, pp. 173-185 (2020).
23. Stehman, S.V.: Selecting and interpreting measures of thematic classification accuracy. Remote Sensing of Environment, 62(1), pp. 77–89 (1997).
24. David, M.W.: Powers. Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. Journal of Machine Learning Technologies, 2(1), 37–63 (2011).