Quality comparison of various machine classification techniques

S. S. Sosinskaya \*, M.V. M.V.Porohova \*, I.S. Komarova\*

\* Irkutsk national research technical University, Irkutsk, Russia

e-mail of S.S.Sosinskaya [sosinskaya@mail.ru](mailto:sosinskaya@mail.ru)

e-mail of M.V.Porohova mariannaporohova@mail.ru

e-mail of I.S. Komarova [ifekomarov@gmail.com](https://e.mail.ru/compose/?mailto=mailto%3Alifekomarov@gmail.com) .

Abstract - The article discusses the approach to classification of machines described by a set of numerical attributes according to their types based on decision trees and their ensembles, the fuzzy inference system and support vector machine method. Based on the information stored in database tables, a single table of attributes has been constructed for all the machine models. The classification results are analyzed with these methods, and the most efficient is recommended according to the conformity percentage. An expert system developed on the base of a fuzzy inference system for the same set of machines, allowing to determine the type membership of the machine by answering the specific questions, is described. This expert system can be used on new machines not included in the predefined set. The software implementation is based on MATLAB system, which provides a variety of tools and methods for acquisition, analysis and visualization of data.

# INTRODUCTION

Intelligent data analysis is one of the most important areas of research in the modern world. The necessity of processing large arrays of data, over several years, Data Mining is being used, in particular, for classification and clustering.

Data classification comprises of defining rules and algorithms of their construction for the proper grouping of the sampled data. Each group (class, cluster) is characterized by certain features, which belong to all cluster components. The number of classes is provided in advance as a result of consideration of the conceptual framework of the subject.

Clustering data is the partitioning the sampled data on various groups so that components of the same group were similar according to a measuring technique of the distance between pairs of sample objects. During clustering, the number of groups is not set and determined within the program execution, in the case of classification the number of classes is provided in advance. In order to traverse the contents of large databases, it is necessary to spend a large amount of time. Therefore, it is advisable to implement a number of methods to automate routine operations of data classification.

The use of techniques of quality evaluation of the classification performance [and their analysis (whose? methods or results)] will allow us to make recommendations for the choice of the most effective method for further classification of newly acquired machines, and develop an expert system for this purpose.

# MACHINE CLASSIFICATION

## Data preparation

For an intelligent data analysis, a database of machine tools from website www.innovatics-tm.ru [] has been used. This database was chosen due to its large size (691 models, each of whichis characterized by 185 attributes), but also because of the difficulty to find open access data associated with a specific company.

Each machine model in the database belongs to a particular group. There are nine groups of machines:

* Turning;
* Drilling and boring;
* Grinding, polishing, lapping;
* Combined;
* Gear and thread processing;
* Milling;
* Planning, slotting, broaching;
* Splitting;
* Center processing.

Common features for all groups are the following:

* The largest diameter of work piece in mm
* The minimum rotation frequency
* The maximum rotation frequency
* The minimum engine power in kW
* The maximum engine power in kW
* The length of the work piece, mm
* The length of the machine, mm
* The width of the machine, mm
* The height of the machine, mm
* The machine weight, kg

In addition, each type of machine has own specific set of features. For example, machines of the turning group have the following characteristics:

* The maximum number of tools
* The diameter of the faceplate/spindle, mm
* The weight of the work piece set
* The flow, rotations in min.
* The caliper movement is longitudinal, mm per rotations
* The caliper movement is crossing, mm per rotation
* The diameter of the table, mm
* For the "drilling and boring" type of machines, the following set of features are defined:
* The maximum spindle stroke , mm
* The maximum distance from spindle to column of machine, mm
* The minimum and the maximum range of thread, mm
* The maximum boring diameter, mm
* The maximum torque on the spindle, mm

A single table of all the features in the amount of 185 is prepared, for the missing features of some models 0 is specified. The data in the table was normalized to [0, 1] to exclude the data of different scales. This table and the source data for all kinds of classification and analysis.

## B. Classification methods for analysis of a sample of machines

A class is a grouping that combines a selection of objects according to similarity of certain features. In the perception of the natural world, a classification of sensations is being carried on, i.e., splitting them into groups of similar but not identical phenomena.

The classification process consists of three stages: design a model, testing the model, and its use.

1. The design of the model consists of figuring out of its parameters based on the training set, containing input and output values.
2. Testing the model implies quality control of the constructed model on the basis of a test sets, whose output values is used only for assessing the quality of the model. If the precision model is valid, it is used to classify new objects.
3. The use of the model is performing the classification according to the chosen model for a sample containing only input features.

In this paper, the training and the test subsets contain a roughly equal number of objects of each class.

## Decision trees and their ensembles

Decision trees are the most common technique of implementing data mining. A decision tree is a graph whose nodes represent the conditions imposed on the values of input features, from which the output argument (a classification feature) is depended, and the leaves of the tree is the value of the classification attribute. The branches of the tree represent paths between intermediate nodes and leaves. In other words, a decision tree is a way of representing rules in a hierarchical structure, where each object corresponds to a single node that provides a decision.

The formation of the decision tree is divided into two phases: the construction of a binary tree and its reduction.

The stage of a binary tree construction is iterative. The most common used algorithm is CART (Classification and Regression Tree), where parent node splitting performed into exact two descendants. At each step of the construction, condition placed in a node divides a given training sample set into two parts: one for the condition being satisfied, and one for the other part, where the condition is not satisfied.

To select the optimal rule, the evaluation function of the quality of a split is used, e.g., the Gini index defined by the formula (1).

(1)

where Pi is a part of objects of class Ki in the sample S. To build a tree of minimum cost, the evaluation the errors of branches and leaves is carried out, then branches that increase the cost are trimmed.

When building a decision tree for machines, tree depth of minimum cost is less by 1 than source one. The decision Tree is shown in figure 1.

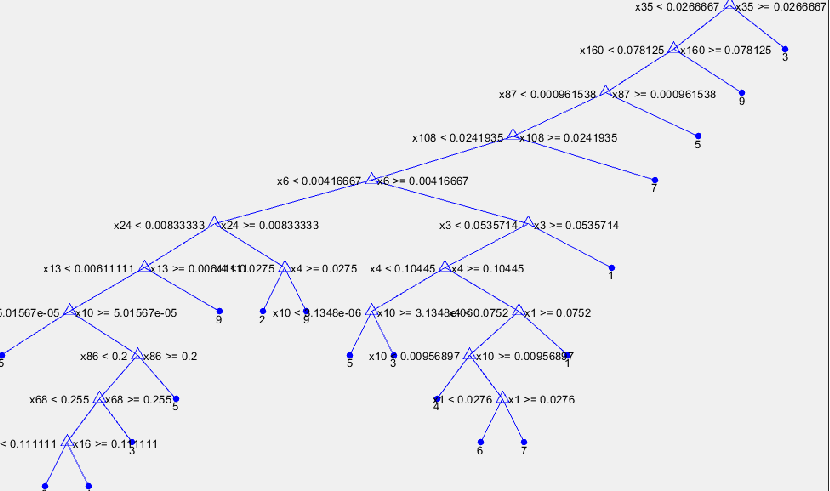


Figure 1. The decision tree for selection of machines

In the decision tree, there are only 18 features; other features, obviously, are less informative.

A classification allows one to define strong and weak models. A model is strong if it contains minimal errors of classification. A weak model, in contrast, contains many errors, i.e., it is not accurate or losing in reliability.

To improve the accuracy of a classification, ensembles of decision trees are been widely used. An ensemble of a classification is based on decision trees, it is a composition of multiple trees, which when combined produce the common forecast with an aggregating classifier. To build ensembles of classifiers, two main approaches are used: bagging and boosting []. The first one is a construction of a set of independent classification models with further decision-making by voting.

In contrast to bagging, boosting teaches each subsequent model using the error data of previous models. The idea of boosting is that the classifiers of an ensemble are built sequentially, and at each iteration, the correction is carried on for observations so that the subsequent classifier would make fewer mistakes on those observations, which often resulted to errors at the previous stages.

Using the language of mathematical logic, we express bagging as a process of improving with union operation, and boosting as an improving with intersection operation.

For the problem at hand, the construction of an ensemble of type bagging were carried out. The classification results of all used techniques are discussed in section III.

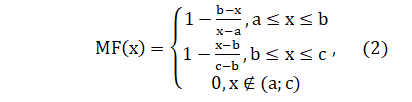
## Fuzzy inference system

Fuzzy inference system is a method of obtaining the output data about the object (for example, belonging to a specific class) based on the input data, which describe the information about the state of the object.

Let be the degree of membership of x to fuzzy set C, which defines a generalization of the concept of the characteristic function of crisp (ordinary) sets. A set y of ordered pairs of the form C={MFc(x)!x}, MFc(x) ∈ [0,1] is called a fuzzy set, denoted by the letter C.

The membership function equal to zero means the object does not belong to a class, its equality to one indicates full membership to the class.

In the article, triangular functions are used for a term membership expression. A triangular function has three parameters a, b, c, and its value is calculated according to the formula (2).



For descriptions of fuzzy sets, linguistic (fuzzy) variables and their terms are used. In order to determine the correspondence between the numerical value of the input variable of the fuzzy inference system and the corresponding value of the membership function for a term of linguistic variable a concept of of fuzzification is introduced. It is the procedure of figuring out the values of the membership functions of fuzzy sets (terms) on the basis of crisp input data.

For the implementation of fuzzy inference a database of rules is used, which is built out of terms of input and output variables; each rule has the following form:

IF a premise THEN a conclusion

In this case, the premise is a combination of terms of the input variables and conclusion is a term of the output variable. As the problem of classification is being solved, a knowledge base of type Sugeno is constructed, rule conclusions are represented with linear functions of the inputs by the formula (3).

(3)

The article proposes an approach to build the system of fuzzy inference over a decision tree: it is assumed that each branch of the tree from root to leaf is a rule; a test condition for an input variable is its node with a corresponding value; the fuzzy inference system includes only those input variables, which are referenced in the tree, and one output variable with the terms, corresponding to the leaves of the tree, i.e., the values of the classification attribute.

Input and output variables of the system of fuzzy inference constructed according to a decision tree is presented in figure 2.

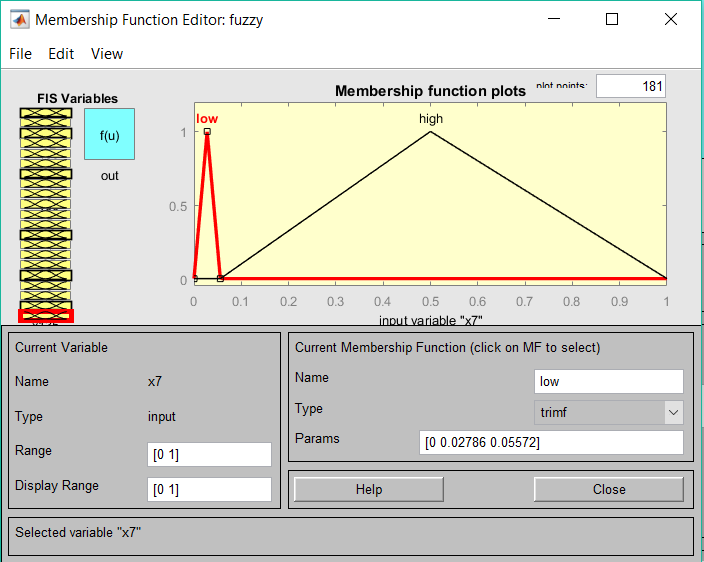


Figure 2.Input and output variables of a system of s fuzzy inference

A part of the rule set for this system is shown in figure 3.

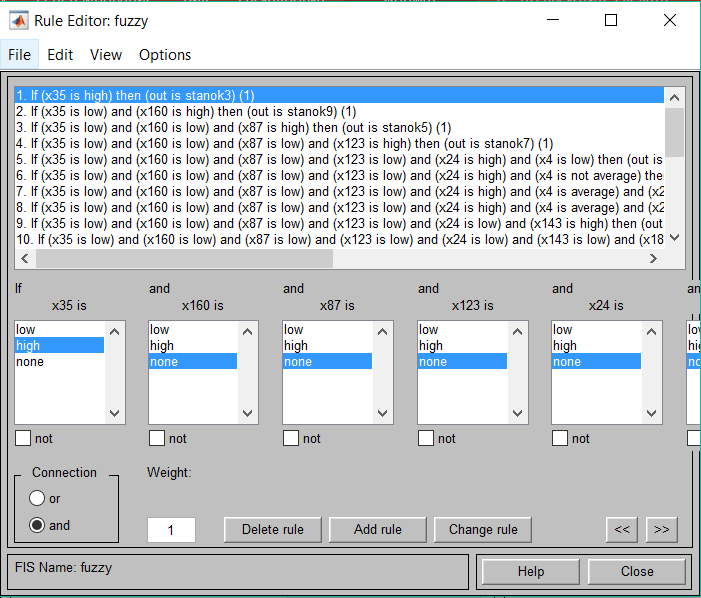


Figure 3. A fragment of the rule set

*E. Support vector machines*

The support vector machine (SVM) belongs to the group of boundary methods. Classes are defined using the borders of regions. This method initially is referred to binary classifiers, although there are techniques of implementing the method for problems of division of objects into several classes. The method is based on the concept of solution planes.

The support vector machine builds a classifier function, which is defined by the formula (4).

, (4)

where (,) is the scalar product, w is a normal vector to the separating hyperplane, b is an auxiliary parameter. Thus, the objects for which F(x) = 1 are in one class, and objects with F(x) = -1 are in the other. The choice of such function is not accidental: any hyperplane can be set in form for some w and b (Figure 2).

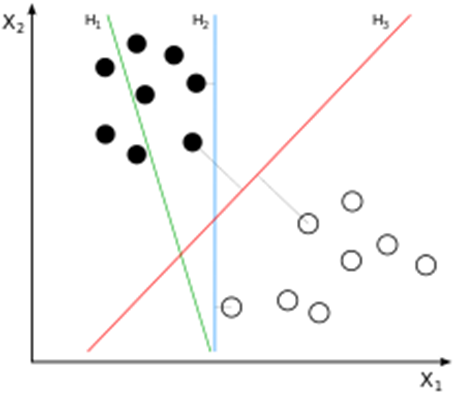


Figure 4. Illustration of SVM for two classes

In the case of multiple classes, the algorithm represents the problem of classification as problem of multiple split into two subclasses. A general approach for such transition include:

* Construction of binary classifiers which distinguish between
* one class from the rest (one-versus-all)
* one class from the other (one-against-one)

Classification of new objects using one-against-all is done by the "winner takes it all" strategy. Classifier having the highest value of the output function assigns the new object to a class. In the figure 5, the classification results of training and testing samples for an array of machines is shown.

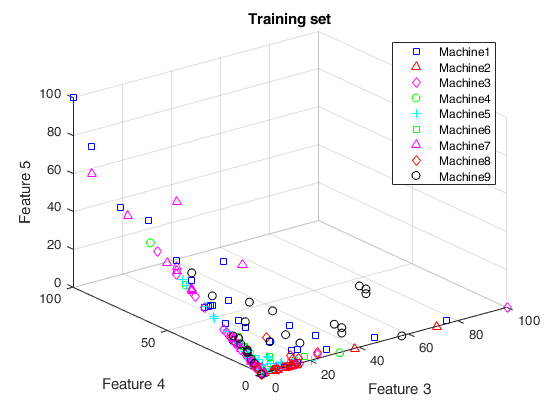


Figure 5. Visualization of classification results for the training samples of machines

*E. Expert system for the machine type determination*

The designed GUI application shows the questions for each input variable of the fuzzy inference system, and receives answer choices, which are terms of the input variables. By choosing answer terms, the user will receive the decision about the group for a machine corresponding to a chosen set of the values of the input variables. The GUI window of the expert system is shown in figure 7.

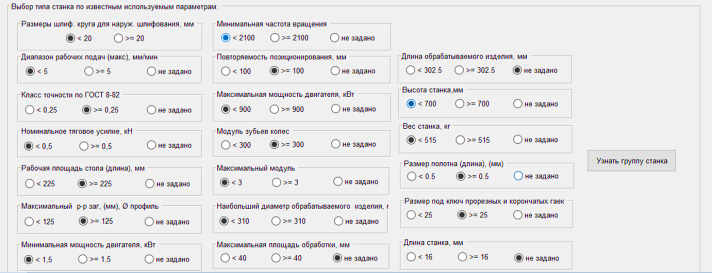


Figure 7. A GUI window of the expert system

*F.Reduction of the dimension of the feature space of the original sampling*

There are a large number of features in the original machine sampling. This may be a problem of overfitting arosen, when big volumes of feature data at some point, do not improve, but worsen the results. In such cases, a data reduction can br applied, i.e. decrease the numer of dimensions of feature space without significant loss of information. There are a number of methods for this purpose . This study uses the principal component analysis (PCA) . Calculation of the principal components can be represented as a calculation [eigenvectors](https://ru.wikipedia.org/wiki/%D0%A1%D0%BE%D0%B1%D1%81%D1%82%D0%B2%D0%B5%D0%BD%D0%BD%D1%8B%D0%B9_%D0%B2%D0%B5%D0%BA%D1%82%D0%BE%D1%80) and

[eigenvalues](https://ru.wikipedia.org/wiki/%D0%A1%D0%BE%D0%B1%D1%81%D1%82%D0%B2%D0%B5%D0%BD%D0%BD%D0%BE%D0%B5_%D0%B7%D0%BD%D0%B0%D1%87%D0%B5%D0%BD%D0%B8%D0%B5) of [the covariance matrix](https://ru.wikipedia.org/wiki/%D0%9A%D0%BE%D0%B2%D0%B0%D1%80%D0%B8%D0%B0%D1%86%D0%B8%D0%BE%D0%BD%D0%BD%D0%B0%D1%8F_%D0%BC%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0) for source data.

The principal component replaces the original matrix of observations with the product of two matrices: matrix of loads and matrix of scores. Matrix od scores is a projection of the original observations to the space of principal components. The order of the rows in these matrices correspond to the descending a percentage of the total variance explained by each principal component of. Instead of the original matrix of observations, we will use new the matrix of scores,. performing the classification of the previously discussed techniques for its first 10 columns. In figure 7, we visualizalised the classification results for a reduced training samples for array machines with support vector machine



Figure 7. Visualization of classification results for the reduced training samples of machines

1. CONCLUSION

To assess the quality of classification for the testing and training samples, we built a matrix of qualities comparing the number of correctly classified objects for each technique with a known number of objects of each class and looked for the average percentage of matches for each of the samples. The results of the comparison are presented in table 1.

TABLE 1. THE RESULTS OF DIFFERENT CLASSIFICATION TECHNIQUES

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Sampling | Decision  tree | | Ensembles  (Bagging) | | Support  vector machine | | |
| % of matches | Run time | % of matches | Run time | % of matches | | Run time |
| Training | 83.7 | 0.8159 | 99.66 | 2.4790 | 99.34 | 6.3813 | |
| Testing | 83.7 | 99.66 | 99.34 |

Analyzing this table, we see that the percentage of matches of training and testing samples for all techniques is the same. This is because each sample included half of the objects of each class.

The percentage of matches for decision tree is the lowest; ensembles of decision trees give a much higher percentage of matches. The support vector machine produces almost the same percentage of matches. However, the runtime of [the decision tree (m.b. SVM???)] is the lowest. Thus, for new machine classification, the user can choose by considering the importance of factors of training time or classification quality. However, all three considered methods produce a high percentage of matches. This allows us to recommend all three methods to be used for classification of new machines. Methods of fuzzy inference and boosting ensemble, which gave significantly worse results, had not been compared.

In the classification on the reduced sample set, the obtained results are shown in table 2.

TABLE 2 THE RESULTS OF VARIOUS CLASSIFICATION METHODS ON THE REDUCED SAMPLE SET

|  |  |  |  |
| --- | --- | --- | --- |
| Decision  tree  % of matches | Ensembles  (Bagging)  % of matches | Ensembles  Boosting  % of matches | Support  vector machine  %  of matches |
| **87.3** | **100** | **77.3** | **96.1** |

Analyzing this table, we see that the percentage of matches in the decision tree method has not changed, the same is for Bagging ensemble, which the percentage is not changed. For the support vectors machine, the percentage is slightly reduced. However, for the Boosting ensemble, which previously had a very low percentage of matches, in this case, it turned out to be much better due to the disappearance of the overfitting effect. Thus, in general, the reduction of feature set is productive. Further, obviously, an additional research using a variety of reduction and selsection methods should be carried on, selecting the most informative features.

##### References

[1] Allwein, E., R. Schapire, and Y. Singer. “Reducing multiclass to binary: A unifying approach for margin classiﬁers.” Journal of Machine Learning Research. Vol. 1, 2000, pp. 113–141.

[2] Fürnkranz, Johannes, “Round Robin Classification.” J. Mach. Learn. Res., Vol. 2, 2002, pp. 721–747.

[3] Escalera, S., O. Pujol, and P. Radeva. “On the decoding process in ternary error-correcting output codes.” IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 32, Issue 7, 2010, pp. 120–134.

[4] Loh, W.Y. “Regression Trees with Unbiased Variable Selection and Interaction Detection.” Statistica Sinica, Vol. 12, 2002, pp. 361–386.

[5] Breiman, L. “Random Forests.” Machine Learning. Vol. 45, pp. 5–32, 2001.

[6] Freund, Y. “A more robust boosting algorithm.” arXiv:0905.2138v1, 2009.

[7] Friedman, J. “Greedy function approximation: A gradient boosting machine.”Annals of Statistics, Vol. 29, No. 5, pp. 1189–1232, 2001.

[8] Friedman, J., T. Hastie, and R. Tibshirani. “Additive logistic regression: A statistical view of boosting.” Annals of Statistics, Vol. 28, No. 2, pp. 337–407, 2000.

[8] Hastie, T., R. Tibshirani, and J. Friedman. The Elements of Statistical Learning section edition, Springer, New York, 2008.

[9] Seiffert, C., T. Khoshgoftaar, J. Hulse, and A. Napolitano. “RUSBoost: Improving clasification performance when training data is skewed.” 19th International Conference on Pattern Recognition, pp. 1–4, 2008.

[10] Warmuth, M., J. Liao, and G. Ratsch. “Totally corrective boosting algorithms that maximize the margin.” Proc. 23rd Int’l. Conf. on Machine Learning, ACM, New York, pp. 1001–1008, 2006.

[11] Breiman, L. Random Forests. Machine Learning 45, pp. 5–32, 2001.

[12] Meinshausen, N. “Quantile Regression Forests.” Journal of Machine Learning Research, Vol. 7, 2006, pp. 983–999.

[13] Christianini, N., and J. C. Shawe-Taylor. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge, UK: Cambridge University Press, 2000.

[14] Fan, R.-E., P.-H. Chen, and C.-J. Lin. “Working set selection using second order information for training support vector machines.” Journal of Machine Learning Research, Vol. 6, 2005, pp. 1889–1918.

[15] Hastie, T., R. Tibshirani, and J. Friedman. The Elements of Statistical Learning, Second Edition. NY: Springer, 2008.

[16] Kecman V., T. -M. Huang, and M. Vogt. “Iterative Single Data Algorithm for Training Kernel Machines from Huge Data Sets: Theory and Performance.” Support Vector Machines: Theory and Applications. Edited by Lipo Wang, 255–274. Berlin: Springer-Verlag, 2005.

[17] Scholkopf, B., J. C. Platt, J. C. Shawe-Taylor, A. J. Smola, and R. C. Williamson. “Estimating the Support of a High-Dimensional Distribution.” Neural Comput., Vol. 13, Number 7, 2001, pp. 1443–1471.

[18] Scholkopf, B., and A. Smola. Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond, Adaptive Computation and Machine Learning. Cambridge, MA: The MIT Press, 2002.