A study of fuzzy logic ensemble system performance on face recognition problem

A Polyakova¹, L Lipinskiy¹

¹ Reshetnev Siberian State Aerospace University
31 Krasnoyarskiy Rabochiy Ave., 660037, Krasnoyarsk, Russia

E-mail: polyakova_nasty@mail.ru

Abstract. Some problems are difficult to solve by using a single intelligent information technology (IIT). The ensemble of the various data mining (DM) techniques is a set of models which are able to solve the problem by itself, but the combination of which allows increasing the efficiency of the system as a whole. Using the IIT ensembles can improve the reliability and efficiency of the final decision, since it emphasizes on the diversity of its components. The new method of the intellectual informational technology ensemble design is considered in this paper. It is based on the fuzzy logic and is designed to solve the classification and regression problems. The ensemble consists of several data mining algorithms: artificial neural network, support vector machine and decision trees. These algorithms and their ensemble have been tested by solving the face recognition problems. Principal components analysis (PCA) is used for feature selection.

1. Introduction

Nowadays informational technologies (IT) find their application in different aspects of society. The power of IT and modern computational technologies is an important tool for solving complex application problems.

Problems for which the intellectual intelligent technologies are used have the following characteristic features: the incompleteness and inaccuracy of initial data, the high computational complexity of the results and their formalization.

Such data analysis problems are difficult to solve using a separate technology. For example, when it is impossible to obtain a formalized mathematical description for the data mining problem, one of the good solutions could be to solve the problem with an artificial neural network (ANN) [1, 2]. When you need to make a decision based on the experience of experts in the subject area of the problem, it is better to use a fuzzy logic system (FLS) to extract the expert experience [3].

The advantage of using support vectors machines is the lack of necessity of the pre-study behavior data [4]. The method of support vectors machine is based on solving a quadratic programming problem, which always has a unique solution.

Ultimately, the use of ensembles of IIT can significantly improve the efficiency of solving applied problems in many areas of human activity.

The ensemble may consist of different methods or of the same one but with the special settings each. For example, if the ensemble consists of several ANN, as described in [5], the separate neural network has a number of layers, number of neurons and other parameters may also vary.
The ensemble can be formed with various methods. In [6], the ensemble is made by means of genetic programming (GP), in [7, 8] the three-stage evolutionary approach is used and in [9] the ensemble based on a C4.5 algorithm.

The new scheme for the ensemble decision design based on fuzzy logic is proposed in this paper.

### 2. Methodology

The authors propose a ensemble approach based on fuzzy logic system to solve the classification and regression problems (Fig. 1.).

Suppose we have a sample \((\bar{x}_i, y_i^j), i = 1,2, ..., n\), where \(n\) - sample size; \(\bar{x}_i = (x_{i,1}, x_{i,2}, ..., x_{i,m})^T\) - set of independent variables, \(m\) - the number of variables in the problem; \(y_i^j = f(\bar{x}_i)\) - corresponding to \(\bar{x}_i\) dependent variable, \(j = 1, v\) - a finite number of classes (names, labels). It is required to effectively design the fuzzy controller capable of forming algorithms ensembles.

Fuzzy controller makes a decision of the algorithms choice depending on the proximity of the test object to the objects in the training sample and the accuracy of the classifier on the nearest subject.

The fuzzy controller has to provide the minimal solution error by the efficient distribution of computational resources between algorithms.

**Figure 1.** The ensemble approach based on fuzzy logic system to solve classification problems.

The sample \((\bar{x}_i, y_i)\) should be divided into 2 parts: training and testing. The task is to synthesize the critical rule based on the train sample that distinguish the classes with the least possible error, i.e. to recognize the new, not encountered earlier objects. The test sample is needed to assess the quality of the decision rule.

The scheme of the proposed approach is presented below:

1. All members of the ensemble are trained using the train sample \(X_{\text{learn}}\).
2. Formed an ordered list of algorithms as follows:
   2.1. Determine \(t=1\), and \(\text{Ensemble} = \emptyset, \varepsilon > 0\) - accuracy (threshold), \(N\) - the maximum number of algorithms in the ensemble, \(X = X_{\text{learn}}\).
   2.2. \(\text{Algorithm}_t\) is the algorithms with the greatest proportion of correctly classified objects on \(X\). For the regression it is the algorithm with the least error on \(X\). \(\text{Ensemble} = \text{Ensemble} \cup \text{Algorithm}_t\).
   2.3. A new set is formed \(X = X \setminus \{x_i; \text{Algorithm}_t(x_i) = y_i\}\). Determine \(t = t+1\).
   2.4. The error is calculated: \(\text{Error (Ensemble)} = \frac{|X|}{|X_{\text{learn}}|} * 100\%\).

   For classification problems \(|X|\) - a set of incorrectly classified objects. For regression problems \(|X|\) - a set of objects for which an error value is less than the threshold, where the threshold set by the user.

Repeat 2.2. - 2.4. as long as the \(X \neq \emptyset\), or \(\text{Error} > \varepsilon\), or \(|\text{Ensemble}| < N\).
In this scheme, the learning algorithm is performed on the entire training set $X_{\text{learn}}$. The ensemble is constructed of the methods that have shown the greatest efficiency in solving the classification problem (regression). Decision making algorithm is based on fuzzy controller.

The rule base and the following linguistic variables are formed to construct the fuzzy controller:

1. Linguistic variable №1 (Fig. 2.): The proximity of the object in training sample to the test sample. Term-set = {low, medium, far}. Basic variable: the distance in the Euclidean metric.

2. The linguistic variable №2 (Fig. 3.): classifier error (regression error) on a sample object. Term-set = {low, medium, high}. Basic variable: the distance in the Euclidean metric.

3. Linguistic variable №3 (Fig. 4.): the credibility of the classifier (regression algorithm). Term-set = {low, medium, high}. Basic variable: a dimensionless quantity in the range from 0 to 1, where 0 - complete lack of confidence in the algorithm, and 1 - absolute trust.

The output of a fuzzy controller is the degree of confidence to the classifier (regression algorithm). The qualifier (the regression algorithm) with the higher credibility is selected.

Suppose there are $N$ classification algorithms (regression), $i = 1, N$. The rule base is as follows:
• IF the Algorithm has a low error AND the closest object from its training sample is close to the test object, the degree of confidence to the Algorithm is high
• IF the Algorithm has a medium error AND the closest object from its training sample is medium to the test object, the degree of confidence to the Algorithm is average

Fuzzy credibility output is calculated for each algorithm. The membership functions for linguistic variables have a triangular form.

The developed decision making algorithm using fuzzy controller is the following:

1. The object $\bar{x}_{i\text{test}}$ from the test sample is an input to the ensemble.
2. Calculates the value of linguistic variables:
   a. Calculate $\rho(\bar{x}_{i\text{test}}, x_j), \forall x_j \in X_{learn}$. Object $x_j$, for which $\min_{x_j \in X_{learn}} \rho(\bar{x}_{i\text{test}}, x_j)$ is said to be $x^*$. Calculate the function membership $\mu_1(\min_{x_j \in X_{learn}} \rho(\bar{x}_{i\text{test}}, x_j))$ for the linguistic variable "The proximity of the object of training sample to the test sample".
   b. Calculate $(y^*_k(x^*) - y_{i^k})^2 \forall j (j = 1, v - a finite number of labels)$. Object $y_{i^k}$, for which $\min_{y_{j\in Y}}(y^*_k(x^*) - y_{i^k})^2$ is called $y_{i^k}^*$. Calculate the function membership $\mu_2(\min_{y_{j\in Y}}(y^*_k(x^*) - y_{i^k})^2)$ for linguistic variable "The success of the classifier (regression algorithm) on a sample object".
3. On the basis of linguistic variables $\mu_1^k, \mu_2^k$ the $k$ (where $k = 1, N$ - number of algorithms in the ensemble) is calculated as the degree of confidence (degree of membership) to each algorithm with center of gravity method $\mu_k = F(\mu_1^k, \mu_2^k)$.

On the basis of the values determined by the membership the $k^{th}$ algorithm in rule with the maximum degree of confidence is determined: $k^* = arg(\max_k(\mu_k))$.

The traditional classification methods work with objects, the parameters of which are given exclusively in the clear form, making them difficult to work with fuzzy objects. Currently for the classification of such objects the methods based on fuzzy logic are being actively developed. These methods form a classes with blurred boundaries and the object can simultaneously belong to several of them with varying degrees of membership.

When an object belongs to the different clusters equally or does not belong to one, the algorithm may be wrong. This problem could be successfully solved using fuzzy classification. In this case, the fuzzy inference is performed by using the classifiers or regression algorithms combination. Instead of a clear answer to the question "what class does the object have?", it determines the level of a membership to a particular class. Thus, the statement "object A belongs to the class 1 with a probability of 90%, Class 2 - 10%" is true and more convenient.

With the help of a software system that implements the algorithm described above on the basis of the formation of the ensemble fuzzy controller, the practical face recognition problems for their images has been resolved.

3. Face recognition problem

The automated human-machine system, working in the dialog mode is called a dialogue system. In the dialog mode, such system responds to all commands from the user, and also appeals to him for confirmation of the execution of specific actions or for obtaining additional information. [10]

Currently, dialog systems are applied in the most different fields of human activity. As a result of human-machine communication many parameters can be determined, such as age, sex, emotions, and other characteristics of the person. [11]
Pattern recognition problem is the main task of data mining. The human face has a central role in social interaction, it is not surprising that the automatic processing of information for the face is an important and a very active area of research. Face detection is a special case of pattern recognition problem [12].

Solving face recognition problem performed in many computer vision systems. The most common systems are the biometric verification and identification systems. Application areas of such systems are biometric scanners and passport, system of technical vision of robots, etc.

Face recognition based on the image is done in two stages:
1. Face positioning on the image;
2. Personality identification based on face images.

The complexity of this problem is determined by some factors:
1. Variability of the image from lighting (directivity, the type and number of light sources);
2. Different spatial positions (relative to the camera) and the scales of faces to their images;
3. The high variability of individuals because of the anatomical features of human variability and type of appearance (for a period of time).

Regarding this, in order to obtain reliable results for face recognition algorithms to their image, it is necessary to perform several requirements for images [13]:
1. All images must be converted from an RGB (color model) to grayscale;
2. Pictures should be centered;
3. Images must be the same size.

The classical scheme of the face recognition algorithm is shown in figure 5:

- Figure 5. Stages of the classical scheme of the face recognition algorithm.

Extracting feature vectors is based on the image. Each pixel of the analyzed image becomes a component of the vector, transforming the black and white image into a vector space [14] (Fig. 6.).

- Figure 6. Extraction the feature vector of the image dimension m × k.

Using of all set of numerical attributes in the process of solving the classification problem can significantly slow down the algorithm and reduce the accuracy of the solution. Therefore it is important to reduce the dimension in the process of problems solving with the feature extraction techniques. In order to select the most suitable characteristics, statistical methods such as factor analysis, as well as more complex, for example, based on genetic algorithms (GA) may be used.
In [15, 16] the procedure of extraction of informative features based on adaptive multicriteria genetic algorithm is considered, its efficiency in combination with different classification models is researched.

In this paper, the principal component analysis (PCA) is used for feature selection. Principal component analysis is based on determining the minimum number of factors which make the greatest contribution to the dispersion of the data.

4. Experimental results

Three image databases was used for the experimental research:

1. AT & T Facedatabase [17];
2. Faces95 [18].

AT & T Facedatabase contains images of 40 individual speakers of 10 images each. Images of persons are represented in the resolution of 92 × 112 pixels.

For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement).

Database of Faces95 contains images of 74 individual speakers of 20 images each. Images of persons are represented in the resolution of 180 × 200 pixels.

As the methods for the ensemble the following algorithms presented in the software system RapidMiner Studio were selected [19]:

1. Artificial neural network (ANN);
2. The method of decision trees (WJ48);
3. Support Vector Machines (SVM).

The principal component analysis is used to reduce dimensionality. Splitting into the test and training samples was made in a random way in the ratio of 70/30. The training of all the algorithms included in the ensemble was done on all training set. All tasks algorithms were run 10 times.

All the results were averaged. Also was the T-Test operator was applied (significance level - 0.05) by Student's criterion. A comparative analysis of the effectiveness of each method individually with the ensemble, which includes all 3 algorithm, was done.

The criterion of effectiveness is classification accuracy - the proportion of correctly recognized items.

Components number decreased from 10304 to 399 after using of the PCA method for the first problem. The results are shown in Table 1.

| T-test          | 77.42% | 61.83% | 77.42% |
|-----------------|--------|--------|--------|
| ANN             |        |        |        |
| SVM             | 0.00493818 | 0.000159513 | 0.00493818 |
| WJ48            | 0.00493818 | 0.000159513 | 0.00493818 |
| Ensemble        |        |        |        |
| ANN             | 3.02454E-05 | -       |        |
| SVM             | 3.02454E-05 | -       |        |
| WJ48            | 3.02454E-05 | -       |        |

According to Table 1 the ensemble and artificial neural network have shown the best result. The analysis by Student's criterion shows that the artificial neural network and ensemble have statistically significant distinctions in comparison with all other algorithms, but statistically not significant among themselves.

A comparative analysis at 10, 50 and 100 components was improved. The results are shown in Tables 2 - 4.
According to the results which were presented in the tables 2-4, we can conclude that with 10 components the ensemble has the best mean precision value, while the difference from the other algorithms are statistically significant.

The ensemble has the same mean precision value as neural network. Comparison of the results obtained by the ensemble and the neural network demonstrated that differences are not statistically significant.

In paper [20] the recognition results of the methods Eigenfaces and N-PCA for database Faces95 are presented (with various train and test percentage of database are presented).

According to the results which were presented in the tables 2-4, we can conclude that with 10 components the ensemble has the best mean precision value, while the difference from the other algorithms are statistically significant.

The ensemble has the same mean precision value as neural network. Comparison of the results obtained by the ensemble and the neural network demonstrated that differences are not statistically significant.

In paper [20] the recognition results of the methods Eigenfaces and N-PCA for database Faces95 are presented (with various train and test percentage of database are presented).

Table 2. Comparative analysis of efficiency for task №1 (number of components – 10).

| T-test | 91.501% | 62.748% | 91.833% |
|--------|---------|---------|---------|
| ANN    | 82.498% | SVM     | 0.000351169 |
|        | 91.501% | ANN     | 3.49373E-07 |
|        | 62.748% | WJ48    | 2.41443E-07 |

Table 3. Comparative analysis of efficiency for task №1 (number of components – 50).

| T-test | 95.916% | 63.418% | 95.916% |
|--------|---------|---------|---------|
| ANN    | 93.333% | SVM     | 0.122178256 |
|        | 95.916% | ANN     | 1.5052E-07 |
|        | 63.418% | WJ48    | 1.50376E-07 |

Table 4. Comparative analysis of efficiency for task №1 (number of components – 100).

| T-test | 95.966% | 62.167% | 95.666% |
|--------|---------|---------|---------|
| ANN    | 93.667% | SVM     | 0.000486134 |
|        | 95.666% | ANN     | 2.30975E-08 |
|        | 62.167% | WJ48    | 2.31149E-08 |

Table 5. Comparative analysis of efficiency for task №1 for some algorithms.

| Accuracy | SVM 93.667% | ANN 95.966% | WJ48 62.167% | Ensemble 95.666% | Eigenfaces 92.50% | N-PCA 93.75% |

According to the results which were presented in the table 5, we can conclude that the ensemble and artificial neural network has the best mean precision value.
Components number decreased from 36000 to 432 after using of the PCA method for the second problem. The results are shown in Table 6.

**Table 6.** Comparative analysis of efficiency for task №2 (all components).

| T-test | 83.262% | 71.203% | 83.2638% |
|--------|---------|---------|----------|
| ANN    | 76.065% | SVM     | 3.60889E-08 | 0.001098475 | 3.5693E-08 |
| WJ48   | 83.262% | ANN     | 6.53248E-07 | 0.022045801 |
| Ensemble | 71.203% | WJ48    | 6.5058E-07 |

According to Table 6, the ensemble have shown the best result. The analysis by Student's criterion shows that the ensemble have statistically significant distinctions in comparison with all other algorithms.

Also let's conduct comparative analysis for 10, 50 and 100 components. The results are shown in Tables 7 - 9.

**Table 7.** Comparative analysis of efficiency for task №2 (number of components – 10).

| T-test | 87.708% | 73.611% | 89.398% |
|--------|---------|---------|----------|
| ANN    | 63.008% | SVM     | 3.56777E-10 | 3.22703E-09 | 8.83219E-11 |
| WJ48   | 87.708% | ANN     | 8.0621E-09 | 0.000747201 |
| Ensemble | 73.611% | WJ48    | 5.39682E-10 |

**Table 8.** Comparative analysis of efficiency for task №2 (number of components – 50).

| T-test | 91.921% | 72.917% | 92.269% |
|--------|---------|---------|----------|
| ANN    | 87.222% | SVM     | 3.30767E-06 | 7.5287E-09 | 8.23075E-07 |
| WJ48   | 91.921% | ANN     | 2.48827E-09 | 0.004601526 |
| Ensemble | 72.917% | WJ48    | 2.16802E-09 |

**Table 9.** Comparative analysis of efficiency for task №2 (number of components – 100).

| T-test | 92.386% | 71.992% | 92.454% |
|--------|---------|---------|----------|
| ANN    | 89.027% | SVM     | 2.51327E-05 | 7.9544E-10 | 2.13644E-05 |
| WJ48   | 92.386% | ANN     | 3.34761E-10 | 0.356727509 |
| Ensemble | 71.992% | WJ48    | 2.07942E-10 |

According to the results which were presented in the tables 7-9, we can conclude that the ensemble has the best mean precision value with any number of the components, while differences of the results obtained by the ensemble and alternative methods are statistically significant.

In paper [21], recognition results of the methods Eigenfaces, LDA and Gabor SVM for database AT&T are presented (with various train and test percentages of database are presented).
Table 10. Comparative analysis of efficiency for task №2 for some algorithms.

| Algorithm       | Accuracy  |
|-----------------|-----------|
| SVM             | 89.027%   |
| ANN             | 92.386%   |
| WJ48            | 71.992%   |
| Ensemble        | 92.454%   |
| Eigenfaces      | 82.7%     |
| LDA             | 92.6%     |
| Gabor SVM       | 90.2%     |

According to the results which were presented in the table 10, we can conclude that the ensemble has the best mean precision value. According to the results presented above, one may conclude that collective decision making allows to increase the accuracy of the solution.

Acknowledgements
This research is supported by the Russian Foundation for Basic Research, within project № 16-01-00767.

References
[1] Freeman J A, Skapura D M 1991 Neural Networks Algorithms Applications and Programming Techniques
[2] Ossowski S 2002 Neural networks for information processing 344 p
[3] Takagi T, Sugeno M 1985 Fuzzy identification of systems and its applications to modeling and control IEEE Trans. on Systems, Man, and Cybernetics
[4] Quinlan J R 1993 C4.5: Programs for Machine learning
[5] Hansen L K 1990 Neural network ensembles IEEE Trans. Pattern Analysis and Machine Intelligence 12 (10) pp 993-1001
[6] Semenkin E S, Shabalov A A, Efimov S N 2011 Automated design of intelligent information technology ensembles by means of genetic programming techinique Bulletin of the Siberian State Aerospace University named after academician M.F. Reshetnev 3 pp 77-81
[7] Bukhtoyarov V V, Semenkin E S 2013 Evolutionary three-stage approach for designing of neural networks ensembles for classification problems Lecture notes in computer science 1 pp 467-477
[8] Bukhtoyarov V V, Semenkina O E 2010 Comprehensive evolutionary approach for neural network ensemble automatic design 2010 IEEE World Congress on Computational Intelligence - 2010 IEEE Congress on Evolutionary Computation
[9] Zhou Z-H, Jiang Y 2004 NeC4.5: Neural Ensemble Based C4.5 IEEE Trans. Knowl. Data Eng. 16(6) pp 770-773
[10] Special Interest Group on Discourse and Dialogue - URL: http://www.sigdial.org/.
[11] Pantic M, Rothkrantz L J M 2003 Toward an Affect-Sensitive Multimodal Human-Computer Interaction Proceedings of the IEEE, Spec. Issue on Human-Computer Multimodal Interface 91(9) pp 1370-1390
[12] Zhao W, Chellappa R, Rosenfeld A, Phillips P J 2003 Face Recognition: A Literature Survey ACM Computing Surveys pp 399-458
[13] Shervin E. 2012 Face Recognition using Eigenfaces or Fisherfaces Mastering OpenCV with Practical Computer Vision Projects
[14] Yang M H, Kriegman D J, Ahuja N 2002 Detecting faces in images *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 pp 34-58
[15] Brester C, Semenkin E, Sidorov M, Minker W 2014 Self-adaptive multi-objective genetic algorithms for feature selection *Proceedings of International Conference on Engineering and Applied Sciences Optimization (OPT-i’14)* pp 1838–1846
[16] Brester Ch, Sidorov M, Semenkin E 2014 Acoustic Emotion Recognition: Two Ways of Features Selection Based on Self-Adaptive Multi-Objective Genetic Algorithm *Proceedings of the International Conference on Informatics in Control, Automation and Robotics (ICINCO)* pp 851–855
[17] AT&T Facedatabase - URL: http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
[18] Collection of Facial Images: Faces95 - URL: http://cswww.essex.ac.uk/mv/allfaces/faces95.html
[19] RapidMiner Studio URL: https://rapidminer.com/
[20] Bansal A K, Chawla P 2013 Performance evaluation of face recognition using PCA and N-PCA *International Journal of Computer Applications* 76(8) pp 14–20.
[21] Daoud E 2015 Face Recognition Using Fuzzy Clustering and Kernel Least Square *Journal of Computer and Communications* 3 pp 1-7