

# The Reduction of Attribute Space Dimensionality in the Fuzzy Classification

Sophia,  
Sosinskaya

Irkutsk National Research  
Technical University  
Irkutsk, Russia  
e-mail: sosinskaya@mail.ru

Daniel,  
Nelle

Potsdam University  
Potsdam, Germany  
e-mail: dnelle@unipotsdam.den

Anastasia,  
Tarasova

Irkutsk National Research  
Technical University  
Irkutsk, Russia  
e-mail: sosinskaya@mail.ru

## ABSTRACT

The article deals with methods of fuzzy classification of observations and means of dimension reducing of their characteristics.

## Keywords

Computer science, informatics, fuzzy classification, dimensionality reduction.

## 1. INTRODUCTION

The task of the classification is rating the observations described with a set of attributes as one of the known classes. The method of fuzzy classification understandable to the specialist of a certain application domain is discussed. The classification method is a fuzzy rule base, where every rule is the expression of the cause-effect pattern relating the observation as one of the known classes expressed with linguistic variables. The methods of attribute space dimensionality reducing and their influence on the accuracy of recognition and the speed of problem solving by giving an example from the archive of UCI Machine Learning Repository [1] used for the empiric algorithm analysis of MATLAB machine learning are considered.

## 2. FUZZY CLASSIFIER

The task of the classification is object rating specified with an informative attribute vector  $X=(x_1, x_2, \dots, x_n)$ , to one of the beforehand defined classes  $\{d_1, d_2, \dots, d_m\}$  i.e. carrying out the representation of the following view:

The classification on the base of the fuzzy deduction [2] takes place in the base of fuzzy rules as follows:

$$\bigcup_{p=1}^{k_j} \bigcap_{i=1}^n x_i = a_{i,j_p} w_{j_p} \rightarrow y = d_j, j=1, \dots, m.$$

The grades of membership of the classification of the object having informative attributes specified with the vector  $X^*=(x_1^*, x_2^*, \dots, x_n^*)$ , to the classes  $d_j$  from the rulebase are defined by the formulae:

$$\mu_{d_j}(X^*) = \bigvee_{p=1}^{k_j} w_{j_p} \bigwedge_{i=1}^n \mu_{j_p}(X^*) \quad j=1, \dots, m, \quad \text{where } \bigvee(\bigwedge) - \text{operations OR (AND).}$$

More often for the operation OR the maximum finding is used, for the operation AND the minimum finding is used.

As a solution the class with the maximum grade of membership is chosen:

$$Y^* = \underset{(d_1, d_2, \dots, d_m)}{\operatorname{argmax}} (\mu_{d_1}(X^*), \mu_{d_2}(X^*), \dots, \mu_{d_m}(X^*))$$

In MATLAB system there are 3 functions for the fuzzy conclusion system construction according to a given learning sample containing a matrix of the observation input attributes by some classed and corresponding values of the

class numbers. The distinction of these functions is the way the fuzzy inference system (FIS) is constructed.

GENFIS1 function generates the FIS of the Sugeno type, the number of input quantity terms and their functions of the membership are chosen by the user. The amount of knowledge base rules is defined as a product of term power of input variables.

GENFIS2 function also generates the FIS of the Sugeno type, preliminarily using the algorithm of subtractive clustering for the observation matrix division into the classes according to their proximity in the attribute space. Automatically the Gaussian function of the membership is chosen. The number of rules of the knowledge base coincides with the number of classes defining the algorithm of the subtractive classification.

GENFIS3 function generates the FIS of the type (Mamdani or Sugeno), chosen by the user, preliminarily utilizing the algorithm of the fuzzy clustering for the division of the observation matrix into classes. The Gaussian function of the membership is also chosen. The amount of the rules of knowledge base concurs with the number of classes defining the algorithm of fuzzy classification.

The FIS, generated by each function, does not reflect the patterns represented by the data between the input and output. It is an initial system for the teaching by means of the ANFIS function, as a result, the patterns given in the data will be taken into account in the parameters.

Having a great number of input variables and rules, accordingly, considerable time for the FIS generation can be spent as well as its training.

Therefore, the ways of attribute space dimensionality reducing without loss of self-descriptiveness are considered.

## 3. THE METHODS OF ATTRIBUTE SPACE DIMENSIONALITY REDUCING

In many cases the determinant for or against the deletion of the attribute is its relevance. In practice, the relevance of the attributes is defined, all attributes, the relevance of which is lower than the preset threshold, are excluded.

The analytical model constructed on the base of the reduced attribute ensemble should become easier for processing, realization and understanding than the model constructed on the initial ensemble.

### 3.1. The method of principal components

Let the matrix of variables  $X$  equal the dimensionality of  $(I \times J)$ , where  $I$  is the number of lines, and  $J$  – the number of independent variables (columns), which are a lot ( $J \gg 1$ ), as a rule. The essence of the method of principal components is considerable lowering of data dimensionality. The initial matrix  $X$  is replaced with two other matrices  $T$  and  $P$ , the

dimensionality of which, A, is less than the number of columns J of the initial matrix X.

The second dimensionality is when the number of lines I is remained. If the dimensionality A is correctly chosen, the matrix T carries as much information as it was in the matrix X.

In the method of principal components new formal variables  $t_a$  ( $a=1, \dots, A$ ) are used being a linear combination of the initial variables of  $x_j$  ( $j=1, \dots, J$ ).

$$t_a = p_{a1}x_1 + \dots + p_{aJ}x_J$$

With the help of these new variables the matrix X is the product of two matrices T and P.

$$X = TP^t + E = \sum_{a=1}^A t_a p_a^t + E$$

The method of principal components is closely connected with another singular value expansion. In the last case the initial matrix X is expanded into the product of the three matrices  $X=USV^t$ .

Here U is a matrix formed with orthonormalized eigen vectors  $u_r$ , the matrix of  $XX^t$ , corresponding values  $l_r$ ;  $XX^t u_r = l_r u_r$ ;

V is a matrix formed with orthonormalized eigen vectors  $v_r$ , the matrices of  $X^t X$ ;  $X^t X v_r = l_r v_r$ ;

S is a positively defined diagonal matrix, the elements of which are  $\sigma_1 \geq \dots \geq \sigma_R \geq 0$  equaling the square roots of proper values  $l_r$

$$\sigma_y = \sqrt{l_y}$$

After the construction of the principle components (usually they locate descending) the matrix of observations in the coordinates of two first principle components is formed. As a rule, they are enough for the model self-descriptiveness saving.

### 3.2. Factor analysis

In the model of the factor analysis the dimensioned variables depend on less number of unobserved (latent) factors.

It is supposed that each variable depends on the linear combination of the factors, the coefficients of this linear combination are called loading.

The model of a simple factor analysis can be presented as follows:

$$X = \mu + \lambda f + e,$$

where X is a vector of observations of a multidimensional random value,

$\lambda$  is a matrix of loading of simple factors,

$\mu$  is a vector of average attribute values of a multidimensional random value,  $X = \{X_1, X_2, \dots, X_d\}$

f is a vector of mutually independent standardized factors,

e is a vector of independent specific factors.

The vector of the values of a multidimensional random value equals  $X = \{X_1, X_2, \dots, X_d\}$ , where  $X_i$  is the i-th attribute of a multidimensional random value.

The dimension of the matrix  $\lambda$  equals  $d \times m$ , where d is the dimensionality of a multidimensional random value, m is the number of simple factors. The matrix element  $\lambda_{ij}$  is called the loading of the j-th factor to the i-th variable.

The number of elements of the vectors  $\mu$ , f and e equals d.

The number of simple factors f should be less than the number of attributes of a multidimensional random value X. This makes it possible to reduce the task dimension from d to m. In a simple factor analysis it is assumed that simple factors are mutually independent and their variance equals a unit, the specific factors  $e_j$  are independent from any  $f_i$ , where  $i=1 \dots m, j=1 \dots d$ . The most possible number of simple factors d is defined with the inequality  $(d+m) \leq (d-m)^2$  (1)

The factor analysis in MATLAB makes it possible to calculate the matrix of the loading.

As the results show, the loading, calculated with the factors, which were not taken with rotating, have a complex structure. The aim of the factor rotation consists in finding the parameters, when each attribute has only several heavy loads. This can often make simpler the interpretation of what the factors are.

If one considers each line of the loading matrix as the coordinates of the point in the m-dimensional space, then each factor conforms to the axis.

The factor rotation is equivalent to the rotation of the axes and the calculation of new loads in a new coordinate frame. Some methods keep orthogonal axis, whereas others are oblique-angled, changing the angle between them, what is more next-generation.

Finally, it is possible to determine the values of a multidimensional random value F according to the calculated loading matrix. The lines of the matrix F fit the observations, the columns conform the factors of a multidimensional random value.

## 4. MODEL EXPERIMENTS

The experiments were carried out with 3 samplings from the archive of UCI Machine Learning Repository.

1. **Survival datum** (306 copies, 2 classes, 3 attributes)

The sampling contains the data about the survival of the patients, who have been operated against the cancer of mammary gland. The input attributes are:

1. The age of the patient during operation
2. The year of the operation
3. The number of positive axillary nodes found

The classifying indication is:

The status of survival (1 – a patient lived 5 years or more after the operation; 2 – a patient died within a 5 year period).

2. **Birth rate** (100 copies, 2 classes, 9 attributes)

100 volunteers gave their sperm, which was analyzed according to the criteria of World Health Organization in 2010. The concentration of spermatozoa is connected with social and demographic datum, environment factors, level of health and habits.

The input attributes are:

1. Season
2. Age during the test
3. Infectious diseases (chickenpox, measles, epidemic parotiditis, poliomyelitis)
4. Serious injury
5. Operative intervention
6. Recently high temperature
7. The frequency of alcohol consumption
8. Smoking habit
9. The number of hours spent in the sitting position

The classifying indication is:

The normal or altered diagnostics

3. **Glass types** (192 copies, 4 classes, 9 attributes), defined in the terms of oxide content.

The input attributes are:

1. index of refraction
2. sodium content
3. magnesium content
4. aluminum content
5. silicon content
6. potassium content
7. calcium content
8. barium content
9. iron, Fe content

The classifying indication is: glass type

Table 1 GENFIS1 Use

N	FULL		PCA		FACT		
	PT/PC	TCT/TR	PT/PC	TCT/TR	PT/PC	TCT/TR	CF
1	100/60	0.29/0.001	80.83/80.83	0.13/0.002	100/100	0.006/0.001	1
2	100/66	18.25/0.38	88/88	0.34/0.1	96/90	0.29/0.01	5
3	Out of memory		65/46	0.26/0.04	97/90	0.28/0.09	5

Table 2 GENFIS2 Use

N	FULL		PCA		FACT		
	PT/PC	TCT/TR	PT/PC	TCT/TR	PT/PC	TCT/TR	CF
1	100/61	1.25/0.15	85.83/85.83	0.17/0.0005	100/100	0.027/0.001	1
2	100/80	6.2/0.02	88/88	0.31/0.04	96/94	0.28/0.06	5
3	100/78	0.43/0.04	65/46	0.26/0.04	97/86	0.34/0.08	5

Table 3 GENFIS3 Use

N	FULL		PCA		FACT		
	PT/PC	TCT/TR	PT/PC	TCT/TR	PT/PC	TCT/TR	CF
1	100/59.17	0.42/0.00118	80.83/80.83	0.11/0.0011	100/100	0.065/0.002	1
2	94/78	0.55/0.04	88/88	0.32/0.08	100/68	0.46/0.01	5
3	97/75	0.68/0.07	80/80	0.12/0.1	100/74	0.24/0.02	5

All samplings were divided into two parts: training and monitoring with approximately equal number of observations in each class. On the basis of the training samplings the systems of the fuzzy output were created with the help of genfis1, genfis2, genfis3 functions and adapted with anfis function. After that the values of the output variables for the copies of the training and monitoring samplings were calculated. The results of the calculations were compared with actual values of the classifying attribute.

At the next step the space of the input attributes was reduced according to the method of principal components and with the help of the factor analysis.

The matrices of the attributes were generated in the coordinates of two maximum principle components and in the coordinates of the factors, the number of which was chosen according to the inequality (1).

All work with the systems of the fuzzy output was repeated for new attribute spaces.

During all stages the time of FIS generation, their training and recognition was computed.

The results of the experiments are shown in tables 1-3.

The signs of the table columns are as follows:

N – sampling number

FULL – a full set of attributes

PCA – the method of principle components

FACT – factor analysis

PT – the percentage of detected objects in the training sampling

PC – the percentage of detected objects in the checking sampling

TCT – the time of FIS creating and training

TR – the time of recognition in FIS

CF – the number of factors

The analysis of the experiment results shows that the use of GENFIS1, forming the rules by a full combination search of the whole terms of the rules, leads to approximately great losses of time and sometimes overflow of the main memory. That is why this method should be rejected or one should start to form the hierarchy of the fuzzy output systems, when FIS of each level uses a part of interrelated input attributes and the output of FIS of the previous level as one of the inputs.

The application of the method of principal components leads to reducing of the time of calculations, as the number of attributes are reduced to 2 and to the lowering of the percent of correctly recognized objects.

Finally, the method of the factor analysis shortens the time of calculations in comparison with the use of the full set of attributes and gives sufficiently high precision of recognition. Therefore, this method can be recommended for the observation classification in other subject areas.

## REFERENCES

- [1] <http://mlr.cs.umass.edu/ml/datasets.html>
- [2] Штовба С.Д. Проектирование нечетких систем средствами MATLAB -М.: Горячая линия – Telecom, 2007. -228 с.