UDC: 51
Original scientific paper

CLASSIFICATION OF MULTISPECTRAL DATA

Eva Ocelíková, Ján Krištof
Faculty of Electrical Engineering and Informatics, Technical University of Košice,
Slovak Republic
e-mail: ocelike@tuke.sk/kristof@lib.tuke.sk

This paper deals with the classification of objects into a limited number of classes. Objects are characterised by n-features, e.g. n-dimensional vector are used to describe them. The paper focuses on the Bayes classifier based on the probability principle, with a fixed number of features during the classification process. Bayes classifier, that is which uses the criterion of the minimum error, was applied to the set of the multispectral data. They represented real images of the Earth's surface obtained from remote Earth sensing. This paper describes the experiences and results obtained during the classification of extensive sets of this multispectral data and an analysis of the influence of dispersions and the mean values of the features of the classification results.

Keywords: classification, Bayes classifier, features, multispectral data, decision rule.

1. INTRODUCTION

The classification process may be applied within different areas of research and practice, e.g. farms, military, medicine, remote Earth sensing, etc. The classical classification techniques use the statistical approach, which typically assumes the normal multidimensional distribution of probability in the experimental data set. Data classification may be supervised or unsupervised [1] [2].

The supervised classification method requires the presence of a training data set typically defined by the expert namely the teacher. Each class of objects is characterised by the basic statistical parameters (mean values vector, co-variance matrix), which are computed from the training set. These parameters guide the discrimination process. The Bayesian classifiers are your typical representatives (i.e. Bayes classifier, Fisher, Wald sequential)[4].

The unsupervised classification is also known as the classification without the teacher. In most cases this classification uses the methods of cluster analysis.

The device that performs the function of classification is called a ***classifier***. The classifier is a system that contains several inputs that are transported with signals carrying information about the objects. The system generates information about the competence of the objects in a particular class within the output. Particular inputs represent features that describe the object.

In accordance with to the number of features, it is possible to divide classifiers into the following classes:

- The classifiers with s ***variable number of features***, the so-called sequential classifiers. These classifiers start the process of classification with an evaluation of one feature and if the process is decidable, the classification ends. In the inverse case the next feature is added and the process repeats itself while the process of classification is undecided.
- The classifiers with a ***fixed number of features***. These are the characteristics that have an exact number of features. This number is fixed during the whole classification process.

According to the other aspects it is possible to divide the classification systems into simple and complex systems, parametric and nonparametric systems, etc.

This paper mentions problems (such as similarity and a large dispersion of particular classes), which can arise in the classification by the Bayes classifier.

In the following example we will look at the Bayes classifier.

2. THE BAYES CLASSIFIER

As we mentioned earlier the Bayes classifier is the classifier with a fixed number of features. The classified patterns $\mathbf{x}$ are described by the n-features $\mathbf{x} = [x_1, x_2, ..., x_n]^T$ and by a pattern space, which contains R disjunctive subsets Q_r, $r = 1,...,R$. The Bayes classifier uses the criterion of the minimum error

$$J = \sum_{r=1}^{R} \int_{Q_r} \sum_{s=1}^{R} \lambda(d_B(\mathbf{x}) \mid \omega_s) p(\mathbf{x} \mid \omega_s) P(\omega_s) d\mathbf{x}, \tag{1}$$

where $\lambda(d_B(\mathbf{x}) \mid \omega_s)$ is the loss function and $d_B(\mathbf{x})$ is the optimal Bayes decision rule, and this minimises the average loss (1). An evaluation of the criterion in this form is difficult therefore it is transformed into a simpler form

$$L_{\mathbf{x}}(\omega_r) = \sum_{s=1}^{R} \lambda(\omega_r \mid \omega_s) p(\mathbf{x} \mid \omega_s) P(\omega_s). \tag{2}$$

Equation (2) represents the loss that would occur after the classification of pattern $\mathbf{x}$ into the class ω_r, if the pattern did not belong to this class. The optimal decision rule is $L_{\mathbf{x}}(d_B(\mathbf{x})) = \min L_{\mathbf{x}}(\omega_r)$. It is possible to realize this in the principle discriminate functions [7]. The form of the discriminate function depends on the choice of the loss functions.

If the ***symmetric loss function*** is selected within the form

$$\begin{aligned} \lambda(\omega_r \mid \omega_s) &= 0 \qquad \text{if} \qquad r = s \\ \lambda(\omega_r \mid \omega_s) &= 1 \qquad \text{if} \qquad r \neq s \end{aligned} \tag{3}$$

then the discriminate function is $g_r(\mathbf{x}) = p(\mathbf{x} \mid \omega_s) P(\omega_s)$. This loss function states that we will lose one unit for only one misrecognition and nothing is lost from the correct classification.

If the ***diagonal loss function*** is selected within the form

$$\begin{aligned} \lambda(\omega_r \mid \omega_s) &= -h_i \qquad \text{if} \qquad r = s \\ \lambda(\omega_r \mid \omega_s) &= 0 \qquad \text{if} \qquad r \neq s \end{aligned} \tag{4}$$

where $h_i > 0$, then the discriminate function is $g_r(\mathbf{x}) = -h_s p(\mathbf{x} \mid \omega_s) P(\omega_s)$. In this case the loss function states a negative loss for a correct decision and no loss for an incorrect decision. The choice of constants h_i is intuitive.

If the patterns $\mathbf{x}$ from miscellaneous classes are like a statistical set, they can be approximated by normal distribution. The conditional probability density function has the form

$$p(\mathbf{x} \mid \omega) = \frac{1}{2\pi^{R/2}\sqrt{|W|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T W^{-1}(\mathbf{x} - \mu) \right), \qquad (5)$$

where $|W|$ is the determinant of co-variance matrix. The co-variance matrix W is

$$W = \frac{1}{K}\sum_{i=1}^{K}\left(\left(\mathbf{x}^{i} - \boldsymbol{\mu}\right)^T \left(\mathbf{x}^{i} - \boldsymbol{\mu}\right)\right), \qquad (6)$$

where K the is number of patterns in the class and $\boldsymbol{\mu}$ is the mean vector

$$\boldsymbol{\mu} = \frac{1}{K}\sum_{i=1}^{K} \mathbf{x}^{i}. \qquad (7)$$

If $g_s(\mathbf{x}) > g_r(\mathbf{x})$ is true then the pattern $\mathbf{x}$ is nearest to the class ω_s. If there are noises in the input set then the classical Bayes classifier assigns this pattern (noise) to the nearest cluster. The augmented Bayes classifier allows us to solve this problem. This augmented classifier tests Mahalanobis distance D_M by chi-quadrate distribution $\chi_{\alpha}^{2}(n)$ with the number of features n and with the significance level α.

The Mahalanobis distance becomes for class ω_r, $r=1, 2, \ldots, R$

$$D_{Mr}^{2} = (\mathbf{x} - \boldsymbol{\mu}_r)^T W^{-1}(\mathbf{x} - \boldsymbol{\mu}_r). \qquad (8)$$

At the base of the test the result of classification is:

1. If $D_{Mr}^{2} \leq \chi_{\alpha}^{2}(n)$ then $\mathbf{x}$ belongs to class ω_r.
2. If $D_{Mr}^{2} > \chi_{\alpha}^{2}(n)$ then $\mathbf{x}$ belongs to class T_0, where T_0 is the *"other"* class.

3. THE CLASSIFICATION OF IMAGE DATA OF EARTH SURFACE

The testing of the augmented Bayes classifier was applied to a training set of image data obtained from remote Earth sensing. The data source contained 800 patterns. The patterns had multispectral character and seven dimensional vectors represented them. The patterns were divided into 8 classes:

1. stubble 2. grass 3. asphalt 4. concrete
5. roof 6. water 7. shadow 8. tree.

The contingent tables (Table 1 and Table 2) show an evaluation of the classification results that were obtained.

The symmetric loss function was used for the realization of the tests. The tests were made with the significance level $\alpha = 0{,}05$ and $\alpha = 0{,}01$. Contingent Table 1 contains the results of the classification with the significance level $\alpha = 0{,}05$. Less satisfactory results were obtained in the 4th and 5th classes. The similarity between mean values and the large dispersions in patterns are part of the reason for this (see Figure 1 and Figure 2). The best results were obtained in the 2nd and 6th were classification was 97% successful.

The results of classification were similar when the significance level was $\alpha = 0{,}01$, (see contingent Table 2). Correct classification increased with the choice of significance level $\alpha = 0{,}01$ by about 2% in the 5th class, and by about 3% in the 3rd and 4th classes. The best results were obtained in the 6th class with classification, being 100% successful. In this class a little dispersion of the patterns was observed.

Table 1. The contingent table – the results of Bayes classification with $\alpha = 0.05$

Original classes	Classes after the classification									N_1	N_1%	E_1%
	1	2	3	4	5	6	7	8	9			
1	95	5	0	0	0	0	0	0	0	100	95	5
2	0	97	2	0	0	0	0	0	1	99	97	3
3	0	0	95	0	2	0	0	0	3	97	97	3
4	0	0	5	85	6	0	0	0	4	96	88	12
5	0	1	0	13	80	0	3	0	3	97	82	18
6	0	0	0	0	0	97	0	0	3	97	100	0
7	0	0	0	0	3	0	90	5	2	98	91	9
8	0	0	0	0	0	0	8	92	0	100	92	8
N_2	95	103	102	98	91	97	101	97	16	784		
E_2%	0	6	7	14	13	0	11	6				

Table 2. The contingent table – the results of Bayes classification with $\alpha = 0.01$

Original classes	Classes after the classification									N_1	N_1%	E_1%
	1	2	3	4	5	6	7	8	9			
1	95	5	0	0	0	0	0	0	0	100	95	5
2	0	98	2	0	0	0	0	0	0	100	98	2
3	0	0	98	0	2	0	0	0	0	100	98	2
4	0	0	5	88	6	0	0	0	1	99	88	12
5	0	1	0	13	82	0	3	0	1	99	82	18
6	0	0	0	0	0	100	0	0	0	100	100	0
7	0	0	0	0	3	0	91	6	0	100	91	9
8	0	0	0	0	0	0	8	92	0	100	92	8
N_2	95	104	105	101	93	100	102	98	2	798		
E_2%	0	6	7	13	12	0	11	7				

An explanation of symbols (Table 1 and Table 2)

- 9 - the "other" class
- N_1 - the number of patterns in the original classes
- N_1% - the number of correctly classified patterns percentage wise
- E_1% - an error of the first kind percentage; wise this it means error which arise after the he assignment of patterns from a certain original class to other classes created after classification
- N_2 - he number of patterns classified into appropriate classes on the map after classification
- E_2% - error of the second kind percentage; wise this means error which arise after the assignment of patterns from all of the original classes to one certain class created after classification.

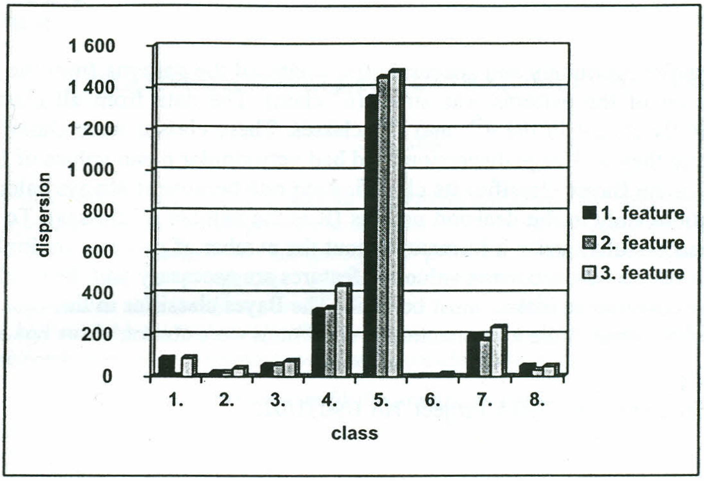

Figure 1. Dispersions of features in the classes

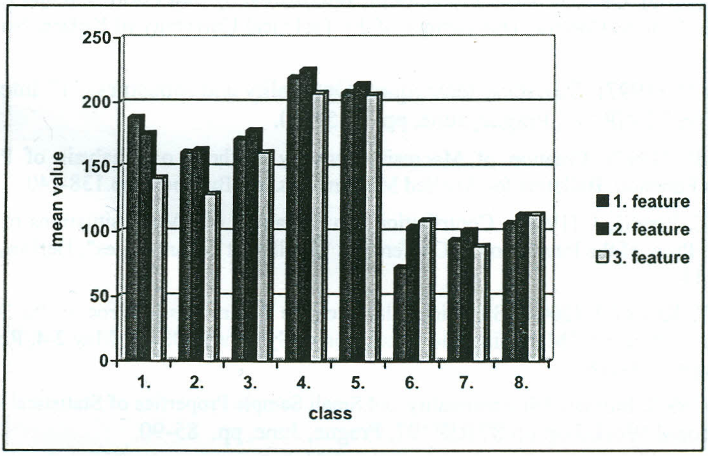

Figure 2. Mean values of features in the classes

4. CONCLUSION

The Bayes classifier absolutely and successfully recognized the patterns from the class in which the dispersion of the patterns was small (6^{th} class). The data from all classes was classified successfully except in the 4^{th} and 5^{th} classes. These classes were classified less successfully because they had large dispersions and had very similar mean values of features. The advantage with the Bayes classifier its classification rate because it always calculates R discriminate equations only in the decision process (R is the number of classes). To use the Bayes classifier successfully, prior information about the number of patterns, multiplicity of patterns in the various classes, the mean values of features are necessary and the condition of the semi definition co-variance matrix must be valid. The Bayes classifier in this case did not classify the data when some of the aforementioned conditions were absent.

Acknowledgement:
This work is supported by the VEGA Project No 1/9031/02.

REFERENCES

[1] Devijaver, P. E.-Kittler, J. (1982). *Pattern recognition: a statistical approach.* Prentice-Hall, Inc., London.

[2] Fukunaga, K. (1972). *Introduction to statistical pattern recognition.* Academic Press, New York,

[3] Hrubina, K. (2000). On some basic Problems of Information System Formation and Realization for Complex Systems Control. Transactions of the Technical University of Košice, No. 1, pp. 1-10.

[4] Kharin Yu. S. (1997). Statistical techniques: Optimality and robustness, 1st International Workshop on STRIP '97, Prague, June, pp. 135-140.

[5] Ocelíková, E. (1995). Creation of Macrosituations by Methods of Analysis of Probability Distribution Function. Bulletins for Applied Mathematics. TU Budapest, pp.138-140

[6] Ocelíková, E.-Paralič, J. (1996). Contribution to Creation of Macrosituations of Complex Systems. In: Proc. of the International Conference: "Intelligent Technologies", Herl'any, Volume 1, pp.103-111.

[7] Ocelíková, E.-Krištof, J. (2000). Statistical Classification Techniques In: Proc. of the 34th Spring International Conference "Modelling and Simulation of Systems MOSIS", May 2-4, Rožnov pod Radhoštem, pp. 111-116

[8] Raudys, S. (1997). Intrinsic Dimensionality and Small Sample Properties of Statistical Classifier. 1st International Workshop on STRIP '97, Prague, June, pp. 85-90.

Received: 08 September 2001
Accepted: 07 November 2001

Eva Oceliková
Ján Krištof

KLASIFIKACIJA MULTISPEKTRALNIH PODATAKA

Sažetak

U ovom radu klasificiraju se objekti koji su karakterizirani pomoću n svojstava u ograničen broj klasa. Objekti su reprezentirani pomoću n-dimenzionalnih vektora. Posebno težište stavlja se na Bayesov klasifikator temeljen na vjerojatnosnom principu. Bayesov klasifikator koristi kriterij minimalne pogreške, a primjenjuje se na multispektralne podatke. Multispektralni podaci repezentiraju realne slike Zemljine površine dobivene iz udaljenih izvora. Ovaj rad opisuje iskustva i rezultate koji su dobiveni tijekom klasifikacije obimnih skupova multispektralnih podataka i analize utjecja disperzije i srednjih vrijednosti svojstava na rezultate klasifikacije.

Ključne riječi: klasifikacija, Bayesov klasifikator, multispektralni podaci, pravilo odlučivanja.