INTERNATIONAL CONFERENCE ON APPLIED ELECTRICAL ENGINEERING AND INFORMATICS 2008

HELD UNDER AUSPICES

OF THE FACULTY OF ELECTRICAL ENGINEERING AND INFORMATICS

OF THE TECHNICAL UNIVERSITY OF KOŠICE

AEI ´2008

SEPTEMBER 8 - 11

GREECE, ATHENS 2008

EDITORS: LIBERIOS VOKOROKOS, ANTON ČIŽMÁR

ISBN: 978-80-553-0066-5 © 2008

DIMENSION DECREASING OF FEATURE SPACE

Eva OCELÍKOVÁ*, Dana KLIMEŠOVÁ**, Iveta ZOLOTOVÁ*
*Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics,
Technical University of Košice, Letná 9, 042 00 Košice, tel. 055/602 2570, e-mail: eva.ocelikova@tuke.sk, iveta.zolotova@tuke.sk,
** Dept. of Information Engineering , Faculty of Economics and Management
Czech University of Life Sciences, Prague, Kamýcká 129, 165 21 Prague 6 – Suchdol, e-mail: klimesova@pef.czu.cz

ABSTRACT

A big number of monitored features in examined objects often complicates a technical realization of decision-making and extends the time necessary for providing a decision. It is possible to decrease dimensionality of the tasks and along with not to decrease a quality of decision-making. The main subject of this contribution relates to some of the possible approaches. Basis of these methods stays in finding a linear transformation of original m-dimensional space of features into a new n-dimensional feature space where $n \le m$. *New features are arisen by suitable linear combination of original features and they are descending sorted according to their variance. The contribution describes experience and results obtained with the decreasing of feature space and the classification of sets of the objects which where represented by real pictures of the Earth surface.*

Keywords: *decision, feature space, dimension reduction, Karhunen - Loeve transformation, principal component method*

1. INTRODUCTION

The quality of decision is always related to the quality and quantity of information available to the subject taking decisions about given objects. For that reason there is tendency to monitor the maximum number of features on the followed objects. But the number of monitored features may not be arbitrarily increased because with the increased number of features the technical realization of their processing becomes more complicated and the required sorting time very long.

From that point of view it is convenient to select so-called informative features, i.e. the features considerably influencing the classification. In such a way the dimension of a problem to be solved can be significantly reduced, better clearness achieved and the calculations simplified whereas the tolerated error size of the right decision stays satisfied.

2. DIMENSION REDUCTION

Reduction of the solved problem can be achieved in two different ways - either by selecting the most informative features from the original feature set or by creation of new features from the original ones. The new features are created following some, e.g. as a linear or non-linear combination of the original features. Dimension reduction methods can be therefore divided into two groups [4].

The main idea of the first group of methods is that from the basic set of available features are chosen those which are most informative, i.e. most advantageous from the next processing point of view, e.g. classification point of view. Many of these methods are based on the information value of features following some criteria and their sorting with respect to these values with decreasing informationallity. When the features are chosen from the beginning of such sorted features, the information lost related to the dimension reduction will be minimal with respect to the chosen criterion. In such a way the whole problem lies in the selection of the suitable criterion for the evaluation of the usefulness feature information. It is necessary that the criterion used for selecting features will be an accordance with the criterion used by the following processing methods.

The sophisticated case occurs when the group of k-individually best features is not for the selected criterion the best group of k-features according to that criterion. It is obvious that to find the optimal subset of features means to examine all possible combinations. Understandably this approach is in most practical problems inapplicable. Therefore the main attention is focused on suboptimal solutions.

This group of methods includes the ones which conjugate features into groups while each of them characterizes one particular side of the object. The method of feature sorting and the method of correlation groups belong here. They serve for finding such groups of features where the correlation coupling, i.e. the sum of absolute values of correlation coefficients between features from the same group (innergroup correlation) is high whereas the correlation coupling between features from different groups (intergroup correlation) is significantly lower.

The core of the second group of methods is in finding a linear or non-linear transformation of *m*-dimensional space of original features into the *n*-dimensional space of informative features $(n \le m)$. Creation of a new system of features have to follow certain requirements. Required is the highest possible information with respect to the right sorting of the feature set, the feature may not be correlate, etc. With linear transformations the new set of features is created as a linear combination of input measurements, with non-linear transformations the original coordinates are transformed curvilinear.

3. FEATURE SPACE TRANSFORMATION

The methods of feature sorting based on transformation convert the original problem of sorting to the problem of finding the proper transformation A

ISBN 978-80-553-0066-5 © 2008

$$A \; : \; W^m \rightarrow \; W^n \qquad (1)$$

With which the original m-dimensional feature space W^m is transformed into the new n-dimensional space W^n, where $n \leq m$. Required is the transformation A should only minimally affect the original amount of information. Probability of classification error while using n new features should not be bigger than with original m features. The only problem is how to measure the information. The classical theory of information is not applicable for this purpose. Such problem is solved in two ways [5]:

1. Transformation A is chosen the way that the representations from W^n would be the best approximation of the original representations from W^m pursuant to square deviation.
2. Transformation A is chosen the way that the representations from W^n would minimize the estimation of probability error in case of using the criterion of minimum error.

The methods based on Karhunen - Loeve transformation belong into the first group. The method where the feature selection is based on ratio of dispersions between classes and inside of classes belongs into the second group. In that case the probability of incorrect decision will be negative proportional to that ratio. When the dispersions of representations inside of classes will be smaller and the distance between the classes will be bigger the probability of incorrect decision will be smaller.

4. FEATURE REDUCTION BY KARHUNEN - LOEVE TRANSFORMATION

Let us assume the total number of investigated objects is K, each of them is characterized by m-dimensional column vector

$$x_k \;\in\; W^m, \qquad k = 1, 2, ..., K. \qquad (2)$$

Let us pick n orthonormal vectors from the space while $n \leq m$. Each representation x_k will be approximated by the linear combination of these vectors, i.e. by the vector

$$\widehat{x}_k \;=\; \sum_{i=1}^{n} c_{ki} e_i \qquad (3)$$

while keeping the square deviation x_k from $\widehat{x}_k$

$$\varepsilon_k^2 = \left\| x_k - \widehat{x}_k \right\|^2 \qquad (4)$$

to be minimal.

It is provable [4] that the previous expression is minimal for the orthonormal vector system

$$e_i \;=\; v_i \;, \qquad i = 1, 2, ...n \qquad (5)$$

where v_i are characteristic vectors corresponding to increasingly ordered characteristic numbers of the autocorrelative matrix of original features

$$\lambda_1 \;\geq \lambda_2 \;\geq \lambda_3 \;\geq \; ... \;\geq \lambda_m \;\geq 0. \qquad (6)$$

Calculation of the selective autocorrelative matrix is based on the following expression

$$\chi(x) \;=\; \frac{1}{K} \sum_{k=1}^{K} x_k x_k^T \qquad (7)$$

where K is the number of objects.

The principal component method for the orthonormal system e_i, $i = 1, 2, ...n$ is using characteristic vectors of a dispersion matrix figured out from

$$\chi^0(x) \;=\; \frac{1}{K} \sum_{k=1}^{K} (x_k - \mu)(x_k - \mu)^T \qquad (8)$$

where μ is the vector of mean values calculated from the all objects.

Authors Chien and Fu suggested for the purpose of ordering and feature selection to use characteristic vectors of the selective dispersion matrix as the orthogonal system

$$\chi^1(x) \;=\; \frac{1}{K} \sum_{r=1}^{R} \sum_{k=1}^{K_r} (x_k - \mu_r)(x_k - \mu_r)^T \qquad (9)$$

where R is the number of classes of representations and K_r is the number of representations belonging into r-th class. The vector μ_r is the vector of mean values r-th class

$$\mu_r \;=\; \frac{1}{K} \sum_{k=1}^{K_r} x_k \;, \qquad r = 1, 2, ..., R. \qquad (10)$$

According to the above-mentioned expressions this method completely ignores the influence of mean values μ_r what may be disadvantage in cases when the mean values of classes are markedly different and contain a big part of information usable for classification.

5. REALIZATION OF THE DIMENSION REDUCTION METHODS

The described techniques of dimension decreasing of the feature space have been applied on a sets of actual objects that were represented with the real pictures of Earth surface from the remote Earth sensing. The pictures were sensed in seven spectral zones, i.e. each object was described by seven features. Each set of representations was divided into four samples (water, forest, agricultural area, city). All pictures from one sample represented an equal type of the object.

Original seven features have been transformed into new features following Karhunen - Loeve transformation and the method of dispersion ratio. By taking into consideration that the characteristic numbers of the

ISBN 978-80-553-0066-5 © 2008

dispersion matrix represent dispersions of new features [6] the variability contribution of *j*-th new feature to the total dispersion value can be calculated by following the expression

$$H_j^* = \frac{\lambda_j}{\lambda_1 + \lambda_2 + \ldots + \lambda_m} \tag{11}$$

and the cumulative contribution of the first p features to the total dispersion value according to the expression

$$H(p) = \frac{\lambda_1 + \lambda_2 + \ldots + \lambda_p}{\lambda_1 + \lambda_2 + \ldots + \lambda_m} \tag{12}$$

where $p \le m.$ It is obvious that if $p = m,$ then $H(p) = 1.$ The contribution $H(p)$ is also called the relative part of dispersion.

The following tables display the results of reduction dimension by principal component method and Chien and Fu method in set mix2.

Table 1 Results of principal component method

Feature	1	2	3	4	5	6	7
Contrib. H_j^*	0,723	0,2312	0,0145	0,0113	0,0109	0,0052	0,0037
Cum.contrib. $H(p)$	0,723	0,9543	0,9688	0,9801	0,9911	0,9963	1

Table 2 Results of Chien and Fu method

Feature	1	2	3	4	5	6	7
Contrib. H_j^*	0,7379	0,2266	0,0127	0,0077	0,0067	0,0041	0,0039
Cum.contrib. $H(p)$	0,7379	0,9646	0,9774	0,9852	0,9919	0,9961	1

Cumulative contributions $H(p)$ of the new features obtained with all mentioned methods are represented on Fig. 1.

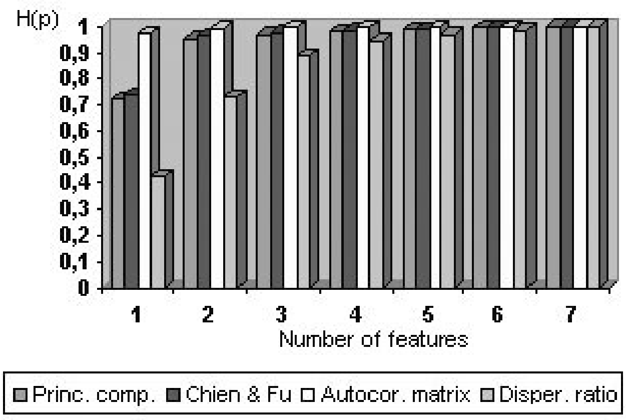

Fig. 1 Cumulative contributions H(p) of new transformed features

6. CONCLUSIONS

According to the obtained results it is possible to reduce the original seven measured features of Earth pictures into two features while satisfying 95% of information with using almost all described methods. An exemption is the method of dispersion ratio where four features are necessary to achieve 95% information of all original features. Quite related is the method of principal components and Chien and Fu method where contribution of the first and second feature are crucial. The contribution of the others are markedly lower and even. Important method is the method of calculation based on the autocorrelative matrix where the required information is reached already during the first transformed feature. It is caused with the high value of the first characteristic number. The reason of it is that in this method the values of representations are not centered so the elements of the autocorrelative matrix are high.

Classifications of Earth pictures into classes have been done using CLASS cluster method with respect to the original as well as reduced dimension. We obtained very satisfied results by clustering of pictures based on all seven original features as well as based on the only one principal component.

ACKNOWLEDGMENTS

This work is supported by the VEGA project No 1/0386/08 and the project information and knowledge support of strategic control - MSM 6046070904.

REFERENCES

[1] AJVAZJAN, S .- BEZAJEVOVA, Z..: Metody vícerozměrné analýzy. Praha, SNTL, 1981

[2] GRIM, J.: Multivariate Statistical Pattern Recognition with Non-reduced Dimensionality. Kybernetika, Vol.22, No.2, pp.142-157, 1986.

[3] HARTIGAN, J. A.: Clustering Algoritms. J. Wiley, New York-London-Sydney-Toronto, 1975

[4] HUAN, L. - HIROSHI, M.: Feature Extraction, Construction and Selection. Kluwer Academic Publishers, Boston, 1998, 440 pp.

[5] JANARDAN, J. Y. R. – LI, Q. - PARK, H.: Feature Reduction via Generalized Uncorrelated linear Discriminant Analysis. *IEEE Trans.on Knowledgw and Data Engineering*, Vol. 18, No.10, 2006, pp.1312-1322

[6] KITTLER, J.: Feature Selection and Extraction. In: Handbook of Pattern Recognition and Image Processing, Young, T.Y. and Fu, K.S. Edition, Academic Press, Inc., 1986, pp. 60-81.

[7] KLIMEŠOVÁ, D. - OCELÍKOVÁ, E**.:** Knowledge Uncertainty and Composed Classifier. *International Journal of Mathematical Modelss and Methods in Applied Sciences*, Issue 2, Volume 1, 2007, pp. 101-105.

[8] OCELÍKOVÁ, E**.**: Dimensionality reduction of Feature Space. *Acta Electrotechnica et Informatica*, No. 1, Vol.1, 2001, pp. 61-65.

[9] OCELÍKOVÁ, E. : Multikriteriálne rozhodovanie. Vyd. elfa, 2002, 87 pp.

[10] PUDIL, P. - NOVOVIČOVÁ, J. - KITTLER, J.: Floating Search Methods in Feature Selection. Pattern Recognition Letters, Vol.15,No. 11. 1994, pp. 1119-1125

[11] OCELÍKOVÁ, E. - MADARÁSZ, L.: Contribution to Creation of Macrosituations of Complex System. *Journal of Advanced Computional Intelligence*, FUJI Technology Press, Ltd. Tokyo, Japan 2002, Vol. 6, No.2, pp. 79-83.

[12] OCELÍKOVÁ, E. - KOČIŠ, M. - KRIŠTOF, J.: Comparing of Dimension Reduction methods in Research of Earth Surface Real Pictures. Bulletins for Applied Mathematics, TU Budapest, 1997, pp. 139-144.

BIOGRAPHIES

Eva Ocelíková defended her Ph.D. thesis, which dealt with multicriterial classification of situations in the complex system, in 1985 at the Slovak Technical University of Bratislava. She is working at the Department of Cybernetics and Artificial Intelligence of the Faculty of Electrical Engineering and Informatics at Technical University in Košice as professor. Her research work includes problems of decision processes, especially the problems of multicriterial classification, composed classifiers designing and high dimensionality reduction of feature space of multidimensional data.

Dana Klimešová is with the Czech University of Life Sciences Prague, Faculty of Economics and Management, Kamýcká 129, 165 21 Prague, Czech Republic, (e-mail: klimesova@pef.czu.cz) and with the Institute of Information Theory and Automation ASCR, Pod vodárenskou věží 4, 182 00 Prague 8, as associate professor. Education: Master Degree in Numerical Mathematics, Faculty of Mathematics and Physics, Charles University in Prague, 1973. Ph.D. Degree in Computer Science, Thesis: Texture Processing in Remote Sensing, Czech Academy of Sciences, Prague, 1987. Research interests: Spatial modelling, geographical information management, context mapping, GIS, image processing. Education activities: Geographical information systems and image processing, remote sensing, system engineering.

Iveta Zolotová was born on 26.7.1959. In 1983 he graduated (MSc) with distinction at the Department of Technical Cybernetics of the Faculty of Electrical Engineering at Technical University in Košice. She defended her CSc. in the field of hierarchical representation of digital image in 1987. Since 1987 she is working as a teacher at the Department of Cybernetics and Artificial Intelligence at the Faculty of Electrical Engineering and Informatics at Technical University in Košice. Her scientific research is focusing on supervisory control, data acquisition, human machine interface and web labs. In addition, she also investigates questions related with the on digital image processing

ISBN 978-80-553-0066-5 © 2008