Application of Clustering Methods for Recognition of Technical Objects

Jan Matuszewski

Abstract - **In this paper the clustering methods with different similarity measures were used to classifying the intercepted signals generated by electromagnetic sources to one of the previously identified *L* classes in *N*-dimensional feature space. The quality of these methods performing technical objects recognition was examined by using the computer simulated date.**

Keywords – **Clustering methods, similarity measures, classification.**

INTRODUCTION

On the modern electromagnetic environment a great deal of information collected by the receivers is processed in real time and computer must be used to analyze, feature extraction and recognize the intercepted unknown signals.

The Electronic Support Measures System (ESM) measures the basic parameters of intercepted radar signals. These basic (typical) parameters are as follows: radio frequency (RF), time of arrival (ToA), pulse width (PW), angle of arrival (AoA), pulse repetition interval (PRI) and period of antenna rotation.

Since each electromagnetic source (emitter, radar) has limited signal parameter ranges (e.g. transmission within a limited frequency band) and often identifiable characteristics, it is assumed that emitter signals with similar characteristics originate with the same object [2], [3].

The classifier compares the measured signal's characteristics (signatures) with a library of stored emitter types, which may have a high degree of inherent uncertainty arising from the methods of data gathering and processing.

SIMILARITY MEASURES

Let X be our set of the feature vectors representing the measured signal parameters from unknown emitters, that is,

$$X = \{\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_M\} \tag{1}$$

The vectors $\underline{x}_i$ are viewed as points in the N-dimensional space. Our objective is to divide X into L sets (clusters). The clusters are described as "continuous regions of this space containing a relatively high density of points, separated from other high density regions by regions of relatively low density of points." The vectors contained in the cluster i are "more similar" to each other and "less similar" to the feature vectors of the other clusters j.

In the approach of objects grouping known as nearest neighbor (NN) rule the different similarity measures in N-dimensional space are used [4]. The most common similarity measures for real-valued vectors used in practice are:

a) *Euclidean*

$$d_E(\underline{x}_i, \underline{x}_j) = \left[\sum_{k=1}^{N} (x_{ik} - x_{jk})^2\right]^{1/2} \tag{2}$$

b) *Seuclidean*

$$d_S(\underline{x}_i, \underline{x}_j) = \sum_{k=1}^{N} \left|x_{ik} - x_{jk}\right|^2 \tag{3}$$

c) *Cityblock*

$$d_C(\underline{x}_i, \underline{x}_j) = \sum_{k=1}^{N} \left|x_{ik} - x_{jk}\right| \tag{4}$$

d*) Minkowski* of rank s=1, 2,... (In calculations assumed: s=3).

$$d_M(\underline{x}_i, \underline{x}_j) = \left[\sum_{k=1}^{N} | x_{ik} - x_{jk} |^2\right]^{1/s} \tag{5}$$

e) *Mahalanobis*

$$d_H = (\underline{x}_i - \underline{u}_j)^t \, \Sigma^{-1} \, (\underline{x}_i - \underline{u}_j) \tag{6}$$

where: $\underline{u}_j = [u_{j1}, u_{j2}, ..., u_{jN}]$ - the mean vector of class j;

Σ^{-1} - the reverse of covariance matrix calculated on the base of set of measured signal parameters.

A feature vector $\underline{x}_k^i \in X$ is assigned to the class of its nearest neighbour. Provided that the number of training samples is large enough, this simple rule exhibits good performance. In minimum-distance method the transformation C is applied to the different similarity measures in the measured features space. In transformation C the similarity measures of unknown object $\underline{z} \in X$ to the class i, $i \in \langle 1, L \rangle$ are calculated, where L - number of classes.

These calculated values determine the similarity measures of object **z** to the respective of class i. The NN-method depends on the choice this class i, $i \in \langle 1, L \rangle$ to which belongs the nearest object $\underline{x}_k^i \in X$ (in accordance with the acceptance of the distance measured) to recognized object $\underline{z}$.

This rule can be written as:

$$C^i(\underline{z}) = \frac{1}{d(\underline{z}, \underline{x}_k^i) + \varepsilon}, \quad i = 1,2,...,L \tag{7}$$

where: $\underline{z}$ – vector representing unknown object, $\underline{x}_k^i$ – object k belonging to class i, (k=1, 2, …, p), p – number of all objects, L – number of identified class, d – chosen similarity measure, ε – positive constant assuring the condition that

$$C^i(\underline{z}) < \infty \tag{8}$$

One of disadvantages of NN-method is that if belonging of one objects $\underline{x}_k$ is false determined than the whole of its neighborhood will be false classified.

Jan Matuszewski, Military University of Technology
2 Gen. Sylwestra Kaliskiego St., 00-908 Warsaw, POLAND
E-mail: jan.matuszewski@wel.wat.edu.pl

THE METHODS OF PATTERNS GROUPING

In clustering the following methods of patterns grouping are:

a) *Single*

A distance (similarity measure) between clusters r and s is determined as the distance between two the nearest neighbours belonging to these clusters, that is

$$d(r,s) = \min(\mathrm{dist}(x_{ri}, x_{sj})),\ i \in <i, n_r>,\ j \in <1, n_s> \quad (9)$$

b) *Complete*

A distance between clusters r and s is determined as the distance between two the farthest neighbours belonging to these clusters, that is

$$d(r,s) = \max(\mathrm{dist}(x_{ri}, x_{sj})),\ i \in <i, n_r>,\ j \in <1, n_s> \quad (10)$$

c) *Average*

In this method the distance is calculated as a difference between two centres of weights

$$d(r,s) = \left\|\tilde{x}_r - \tilde{x}_s\right\|_2 \quad (11)$$

where:

$$\|\ \|_2 \text{ - Euclidean metric, } \tilde{x}_r = \frac{1}{n_r}\sum_{i=1}^{n_r} x_{ri},\ \ \tilde{x}_s = \frac{1}{n_s}\sum_{j=1}^{n_j} x_{sj} \quad (12)$$

d) *Median*

The distance is calculated in the similar way to as in Eq. 12

$$d(r,s) = \left\|\tilde{x}_r + \tilde{x}_s\right\|_2 \quad (13)$$

If cluster r was created from connection of clusters p and q then

$$\tilde{x}_r = \frac{1}{2}(\tilde{x}_p + \tilde{x}_q) \quad (14)$$

e) *Centroid*

$$d(r,s) = \frac{1}{n_r n_s}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{dist}(x_{ri}, x_{sj}) \quad (15)$$

f) *Ward*

$$d^2(r,s) = n_r n_s \frac{\left\|\tilde{x}_r - \tilde{x}_s\right\|_2^2}{(n_r + n_s)} \quad (16)$$

These described above methods of patterns grouping are illustrated on the Fig. 1.

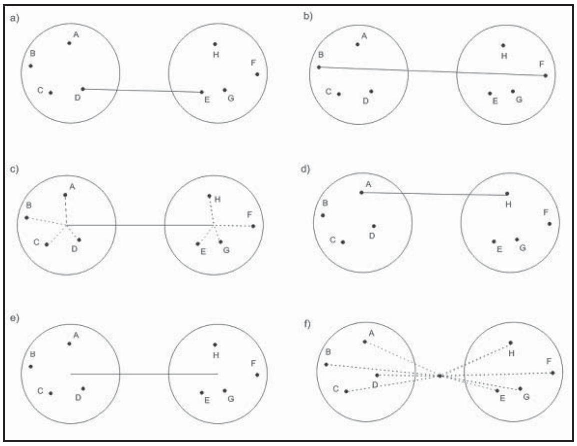

Fig. 1 The rules of objects grouping: a) Single, b) Complete, c) Average, d) Median, e) Centroid, f) Ward

EXPERIMENT

The result of simulated data for 4 classes containing 100 patterns in each class is illustrated on the Fig. 2 and probability of true classification in the Table 1, [1].

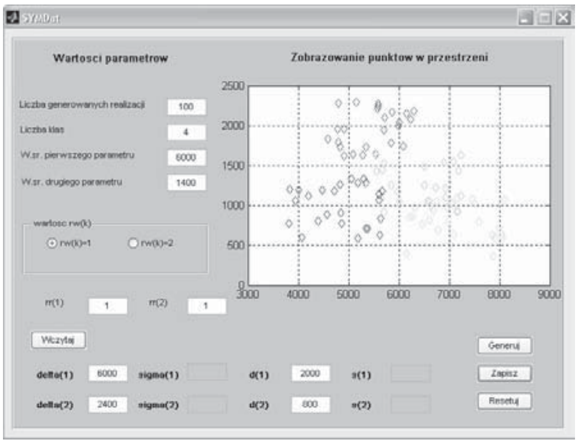

Fig. 2 Simulated results for 4 classes and 2 parameters

TABLE 1. THE PROBABILITY OF TRUE CLASSIFICATION

	Euclidean	Seuclidean	Cityblock	Mahalanobis	Minkowski
Single	0,452	0,496	0,566	0,562	0,419
Complete	0,703	0,670	0,660	0,617	0,700
Average	0,714	0,678	0,715	0,679	0,707
Centroid	0,713	0,679	-	0,671	-
Median	0,651	0,634	-	0,631	-
Ward	0,695	0,666	-	0,645	-

CONCLUSION[1]

After making a few experiments it is difficult to confirm which version of clustering method is better. To choose the best method it should be performed much more calculations with the different distance measures so that to find the smallest error probability of objects recognition. In an electromagnetic environment the values of emitter signal parameters often change. Further testing of these methods should show which of them is better and more convenient to emitter recognition.

REFERENCES

[1] U. Kołodziejska, "Badanie jakości algorytmu rozpoznawania obiektów przy użyciu metod nieparametrycznych," *Praca dyplomowa*, WAT, Warszawa, 2009.

[2] J. Matuszewski, "Identification of radar signals using discriminant vectors," *Proc. XVII International Conference on Microwaves, Radar and Wireless Communications MIKON-2008*, May 19-21, 2008, Wrocław, pp. 433-436.

[3] J. Matuszewski, A. Pieniężny, S. Konatowski, "Emitter identification using neural networks and clustering algorithms," *International Conference on Telecommunication Systems and Technologies* (*ICTST'2008)*, October 22-24, 2008, Charkow, Ukraina, pp. II-301-II-304.

[4] S. Theodoridis, K. Kountroumbas, "Pattern recognition," Academic Press, London, 1998.

[1] This paper is financed from sources for science in the years 2007-2010 in a frame of ordered scientific project PBZ-MNiSW-DBO-04/I/2007.