Methods of Computing Epsilon Thresholds in the Estimates' Calculation's Algorithms

Mirzayan Kamilov[1], Mirzaakbar Hudayberdiev[2], Alisher Khamroev[3]
Centre for the Development of Software and Hardware-Program Complexes, Tashkent, Uzbekistan
[1]*kamilov_m@mail.ru*, [2]*mirzaakbarhh@gmail.com*, [3]*alisherhsh@mail.ru*

***Abstract*— Methods for calculating epsilon thresholds' values in the family of estimates' calculation's algorithms are discussed. A new approach for choosing the method, which improves quality of recognition results, is proposed.**

***Keywords*— pattern recognition; estimates' calculation's algorithms; training sample**

I. Introduction

The estimates' calculation's algorithm is one of the widely used pattern recognition algorithms. One of the important problems of the estimate's calculation's algorithms (ECA) is to calculate the optimal values of epsilon thresholds [1,2].

The epsilon threshold values express distinctions between corresponding features of two objects. If given object is described by n features, then for each feature of the object its threshold value is defined and we'll have following epsilon thresholds: $\varepsilon_1, \varepsilon_2, ..., \varepsilon_i, ..., \varepsilon_n$.

It is known that solving the classification problem is calculating the values of the unknown parameters of recognition algorithm by analyzing a training sample.

The problem of choosing the optimal method among the epsilon thresholds' calculation methods is considered in this article.

II. Formulation of the Problem

Pattern recognition is generally categorized according to the type of learning procedure used to generate the output value: supervised learning and unsupervised learning systems. Supervised learning assumes that a set of training data (the training set) has been provided, it consists of a set of instances that have been properly labeled by hand with the correct output. A learning procedure then generates a model that attempts to meet two sometimes conflicting objectives: Perform as well as possible on the training data, and generalize as well as possible to new data. The following objects and corresponding features are given: $S_j = (x_1(S_j), x_2(S_j), ..., x_n(S_j))$, $j = \overline{1,m}$, $x_i(S_j) = \alpha_i \in X_i$, $i = \overline{1,n}$. The training set's objects are divided into the classes $K_1, K_2, ..., K_l$ so that: $K_u \cap K_q = \varnothing$, $u,q=1,...,l$. And methods for calculating epsilon thresholds' are given as u_v ($v = \overline{1,t}$).

The proximity function between two objects in ECA is given by:

$$B(S_j, S_q) = \begin{cases} 1, & \sum_{i=1}^{n} \sigma(\alpha_{ij}, \alpha_{iq}) < E \\ 0, & \end{cases} \tag{1}$$

where

$$\sigma(\alpha_{ij}, \alpha_{iq}) = \begin{cases} 1, & |\alpha_{ij} - \alpha_{iq}| \le \varepsilon_i, \\ 0, & |\alpha_{ij} - \alpha_{iq}| > \varepsilon_i. \end{cases}$$

$$1 \le E < n,\ E \in N;\ j = \overline{1,m}.$$

Calculation of optimal value of epsilon threshold is required to find correct proximity function $B(S_j, S_q)$ in ECA. Let's consider methods for solving the problem of epsilon thresholds' calculation.

III. The Method of Solution

We consider several methods for calculating the values of epsilon thresholds, they are divided into the following categories:

1. Normally, the epsilon thresholds' values are defined as the distances between maximum and minimum values of i-th features of K_g class's objects:

$$\bar{\varepsilon}_i^g = \frac{1}{2}(\max_{i,j} x_{ij}^g - \min_{i,j} x_{ij}^g) \tag{2}$$

2. The calculation of epsilon thresholds' values by the mathematical expectation (Mx_i^g) of i-th features of K_g class's objects:

$$Mx_i^g = \frac{1}{m_g - m_{g-1}} \sum_{j=m_{g-1}+1}^{m_g} x_{ij}^g$$

$$\tilde{\varepsilon}_i^g = \begin{cases} Mx_i^g - \min x_{ij}^g, & |Mx_i^g - \max x_{ij}^g| \le |\max x_{ij}^g - Mx_i^g|, \\ \max x_{ij}^g - Mx_i^g, & \end{cases} \tag{3}$$

where $j = \overline{1,m}$.

3. The calculation of epsilon thresholds by min-max values of features, calculated by classes, and mathematical expectation of the objects' *i*-th features:

$$\varepsilon_i^g = \frac{1}{2} | \bar{\varepsilon}_i^g - \tilde{\varepsilon}_i^g | . \qquad (4)$$

4. The calculation of epsilon thresholds' values on the base of ECA [2]:

$$\varepsilon_i = \frac{\sum_{j=1}^{m-1} \left[(m-j)\alpha_{ij} - j\alpha_{ij+1} \right]}{C_m^2} \qquad (5)$$

IV. Experimental Results

The values of quality function φ_v are calculated by the formulae (2)-(5) (u_v, $v = \overline{1,t}$) using $\Gamma_u(S_j)$ voices on the training sample (TS). This process is applied to Iris, Wine, DogWolf training samples and results of experiments are given in the following table 1.

TABLE I.

TS / u_v	Iris	Wine	DogWolf
MatExpect	94	93,8	92,9
Avr	94,6	90,5	83,3
MinMax	92	80,9	81
method of ECA	86,7	93,3	92,9

We introduce $Q(u_v)$ function for analyzing the results of table 1:

$$Q(u_v) = \max_{v=\overline{1,t}} (\varphi_v), \qquad (6)$$

V. Conclusions

The results presented above help us to select suboptimal method for calculating epsilon threshold values. Values obtained are used as initial values of epsilon threshold parameters. The iteration methods are used to find optimal values of epsilon threshold parameters. The values of $Q(u_v)$ function are very useful to improve recognition quality and the algorithm can adapte its parameters by itself.

VI. Acknowledgment

This work was supported in part by the Grants F4-FA-F004 and A5-FA-F019 of the Committee of Science and Technology of the Republic of Uzbekistan.

References

[1] Zhuravlev Yu. I., Selected Scientific Works Magistr, Moscow, 1998, 417 p. (in russian),

[2] Kamilov M.M., Aliev E.M., Kim A.N. About calculation of the epsilon threshold for recognition objects by the voiting algorithms. "Problems Cybernetics", No 43, Tashkent, IC with CC AS UzSSR, 1971. – Pp. 72-79. (in russian).

[3] Hudayberdiev M. Kh., Akhatov A.R., Hamroev A.Sh. On a Model of Forming the Optimal Parameters of the Recognition Algorithms. International journal of KIMICS, vol.9, no.5. – Korea, Oktober 2011. – Pp. 607-609.