An Algorithm for Bayes Classification of Interval Information

Abstract

KOWALSKI Piotr Andrzej[1] and KULCZYCKI Piotr[2]

[1] *Cracow University of Technology*
ul. Warszawska 24, PL-31-155 Cracow
Poland
E-mail: pkowal@pk.edu.pl

[2]*System Research Institute, Polish Academy of Sciences*
ul. Newelska 6, PL-01-447 Warsaw
Poland
E-mail: kulczycki@ibspan.waw.pl

Lately, interest in interval analysis has grown notably in many practical applications [1]. Fundamental to this concept is the assumption for the only information about theinvestigated quantity $x \in \mathbf{R}$ being the fact that it fulfils the condition $\underline{x} \leq x \leq \overline{x}$, and consequently can be treated as the interval $[\underline{x}, \overline{x}]$. Dynamic development has also taken place in information technology [4] in the domain of data analysis and exploration - here, the classification and clustering tasks belong to basic procedures [6]. Classification consists in assigning the element under consideration to one among a number of previously defined classes. These are most often represented by samples constituting sets of elements representative for particular classes. In many practical problems - including those in which information containing uncertainties is treated - this representation causes elements generating samples to be defined uniformly.

This paper presents a procedure for classifying uncertain information, given in the form of the interval vector $[\,[\underline{x}_1, \overline{x}_1], [\underline{x}_2, \overline{x}_2], ..., [\underline{x}_n, \overline{x}_n]\,]^T$, where $\underline{x}_k \leq \overline{x}_k$ for $k = 1, 2, ..., n$, while the samples of the particular classes are defined uniformly, i.e. when $\underline{x}_k = \overline{x}_k$. This concept of classification is based on the Bayes approach, ensuring a minimum of potential losses arising from misclassifications. For a such-formulated problem the methodology of statistical kernel estimators has been used [2, 5], which freed the above procedure from arbitrary assumptions concerning the shapes of samples their identification constitutes an integral part of the presented algorithm. A procedure has also been worked out for reducing samples by those elements having negligible or even negative influence on the correctness of classification. Its concept has been based on a sensitivity method, inspired by artificial neural networks theory, while the goal is to increase the number of proper classifications as well as, primarily, calculation speed. The veracity of the presented method has also been tested for the case where the samples of particular classes were obtained as a result of clustering [3].

References

[1] Jaulin L., Kieffer M., Didrit O., Walter E., *Applied Interval Analysis.*, Springer, (2001).

[2] Kulczycki P., *Estymatory jadrowe w analizie systemowej.*, WNT, (2005).

[3] Kulczycki P., Charytanowicz M., *A Complete Gradient Clustering Algorithm by Kernel Estimators.*, in press.

[4] Kulczycki P., Hryniewicz O., Kacprzyk J. (ed.), *Techniki infomacyjne w badaniach systemowych.*, WNT, (2007).

[5] Wand M.P., Jones M.C., *Kernel smoothing.*, Chapman and Hall, (1995).

[6] Webb A., *Statistical Pattern Recognition.*, Wiley, (2003).