Neural Classification for Interval Information

Piotr A. Kowalski[1,2] and Piotr Kulczycki[1,2]

[1] Cracow University of Technology,
Department of Automatic Control and Information Technology,
ul. Warszawska 24, PL-31-155 Cracow, Poland
{pkowal,kulczycki}@pk.edu.pl

[2] Systems Research Institute, Polish Academy of Sciences,
ul. Newelska 6, PL-01-447 Warsaw, Poland
{pakowal,kulczycki}@ibspan.waw.pl

Abstract. The subject of the presented research is to determine the complete neural procedure for classifying inaccurate information, as given in the form of an interval vector. For such a formulated task, a basic functionality Probabilistic Neural Network was extended upon the interval type of information. As a consequence, a new type of neural network has been proposed. The presented methodology was positively verified using random and benchmark data sets. In addition, a comparative analysis of existing algorithms with similar conditions was made.

Keywords: neural networks, probabilistic neural networks, data analysis, classification, interval data, imprecise information.

1 Introduction

Recently, in many applications, there has been a growth of interest in interval analysis. The basis of this concept is the assumption that the only possessed information about the tested quantity x, is the fact that it fulfils the relationship $\underline{x} \leq x \leq \overline{x}$, and, consequently, it may be identified with the interval:

$$[\underline{x}, \overline{x}]. \tag{1}$$

Interval analysis is a separate area of mathematics which has its own formal apparatus based on the axiom theory [15]. Formerly, its primary use was to provide the required accuracy within numerical calculations, as these are often affected by the control error resulting from rounding [1]. However, as a result of its continuous development, this area is becoming frequently used in engineering, econometrics, and other related fields [5]. Its main advantage is the fact that, by its nature, it is modelling the uncertainty of an examined quantity by using the simplest possible formula. In many applications, interval analysis has found to be completely sufficient, and it requires low computation effort (which allows its employment in very complex tasks). Moreover, this methodology is easy to identify and interpret, while also having a convenient formalism based on a mathematical apparatus.

G. Agre et al. (Eds.): AIMSA 2014, LNAI 8722, pp. 206–213, 2014.

The goal of the research is to reveal the complete neural procedure for classifying inaccurate information (1) as applied in cases of multi-dimensional data that are expressed in the form of the interval vector:

$$[[\underline{x}_1, \overline{x}_1], [\underline{x}_2, \overline{x}_2], \ldots [\underline{x}_n, \overline{x}_n]]^T , \tag{2}$$

where $\underline{x_k} \leqslant \overline{x_k}$ for $k = 1, 2, ..., n$, when the patterns of individual classes are determined on the basis of unambiguously defined sets of items, that is

$$\underline{x_k} = \overline{x_k} \text{ for } k = 1, 2, ..., n. \tag{3}$$

The concept of classification is based on employing the Probabilistic Neural Network approach by way of using Bayes theorem, when provided with a minimum of potential losses resulting from misclassification. For such a task, a formulated statistical kernel estimator methodology is used. This procedure is not dependent on arbitrary assumptions about character patterns. Their identification will be an integral part of the presented algorithm.

2 Kernel Density Estimator

The Statistical Kernel Density Estimators (KDE) belong to the set of nonparametric methods. They allow the designation and illustration of the characteristics of random variable distribution, without possessing the information on the membership of a particular class.

Consider a n-dimensional random variable whose distribution has density function f. Its kernel estimator $\hat{f}$ is determined on the basis of the m-element random sample:

$$x_1, x_2, \ldots, x_m \tag{4}$$

and is defined by the formula:

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^{m} K\left(\frac{x - x_i}{h}\right). \tag{5}$$

The positive coefficient h is called 'smoothing parameter'. A measurable function K, symmetric with respect to zero at this point, having weak local maximum and satisfying the condition $K(x)\colon \mathbb{R} \to [0, \infty)$, is referred as a 'kernel'. The form of the kernel K practically does not affect the statistical quality of the estimation. In this work, we will use the one-dimensional Cauchy kernel

$$K(x) = 2/\pi(x^2 + 1)^2. \tag{6}$$

In the case of multivariate situation, this will be generalized using the concept of a kernel product.

More detailed information about the practical issues of employing KDE methods, as well as usage examples, can be found in cited references [11] and [19].

3 Neural Network for Interval Imprecise Information

The Probabilistic Neural Network (PNN), which is very often considered as being a neural realization of a set of KDE, is a special type of a Radial Neural Network. It is used mainly for regression [16], prediction [18], classification [14] [3] and identification [2] tasks, but also for non-linear time series analysis.

In this part of this paper, the generalization of PNN as used in processing interval information, will be introduced. This neural structure is based on Specht's Probabilistic Network [17], but it has a several new elements which enable us to classify interval information. Figure 1 reveals the topological scheme of a generalized probabilistic neural network. This structure, in this paper, is treated as a network implementation of the interval information classifier.

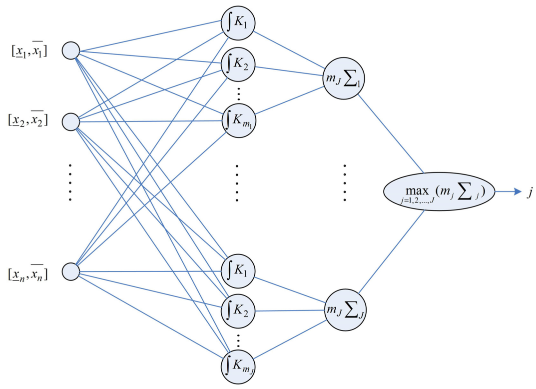

Fig. 1. The structure of a PNN extended for processing imprecise information

In this created network, there are four layers. The first is the input layer, with size equal m, wherein the inputs correspond to the dimensions of the interval element (3) under classification. The next layer is a subset of neurons representing the successive patterns of classes. Each of these consists of an appropriate number of neurons whose function is to bring about the operation of integration (9). The third layer provides a summation of neuronal signals within a pattern class, as well as a multiplication of the result value by group cardinality (8). The final single neuron, located in the output layer, determines the highest values obtained from the pattern layer and fixes the final result of this classification task.

In a situation wherein information is given by the interval $[\underline{x}, \overline{x}]$, the tested element, based on Bayes Theorem, belongs to the class for which the value:

$$\frac{m_1}{\overline{x} - \underline{x}} \int_{\underline{x}}^{\overline{x}} \hat{f}_1(x)\, dx,\ \frac{m_2}{\overline{x} - \underline{x}} \int_{\underline{x}}^{\overline{x}} \hat{f}_2(x)\, dx, \ldots,\ \frac{m_J}{\overline{x} - \underline{x}} \int_{\underline{x}}^{\overline{x}} \hat{f}_J(x)\, dx \tag{7}$$

is the largest. This is a natural extension of Bayes' theorem into interval information type. In the above formula, the positive constants $1/(\overline{x} - \underline{x})$, can be omitted as these are negligible for the optimization problem under consideration. Therefore finally, it can be presented in the following form:

$$m_1 \int_{\underline{x}}^{\overline{x}} \hat{f}_1(x)\,dx,\ m_2 \int_{\underline{x}}^{\overline{x}} \hat{f}_2(x)\,dx, \ldots,\ m_J \int_{\underline{x}}^{\overline{x}} \hat{f}_J(x)\,dx \tag{8}$$

Moreover, for every $\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_J$ one can note:

$$\int_{\underline{x}}^{\overline{x}} \hat{f}(x)\,dx = \hat{F}(\overline{x}) - \hat{F}(\underline{x}), \tag{9}$$

where $\hat{F}$ means the primitive of the function $\hat{f}$. For the Cauchy Kernel 6 used here, the following analytical formula can be obtained:

$$\hat{F}(x) = \frac{1}{m} \sum_{i=1}^{m} \left[\frac{(x^2 - 2xx_i + x_i^2 + h^2)\arctan(\frac{x-x_i}{h}) + h(x - x_i)}{x^2 - 2xx_i + x_i^2 + h^2} + \frac{\pi}{2} \right] \tag{10}$$

(note that the constant $\pi/2$ could be again omitted for equal cardinality of pattern sets). In the multidimensional case, when information is represented by the interval vector (3), this can be easily extended by using the concept of a product kernel [9,13] .

4 Numerical Verification

The correctness of the presented method was verified through was conducted by the way of numerical simulation. Due to the specific conditioning of the presented method, this type of data was not found in public repositories. The following are the results for data obtained using the random number generator with normal distribution. This was done using a given vector of expected value and a covariance matrix. This, in turn, was derived from the implemented multivariate normal distribution generator based on the concept of Box-Muller [4].

The quality assessment methods presented here were obtained by generating a set of random numbers of the assumed distribution, and provide an analysis of the correctness of the results of the classification procedures used for data that was either of an interval type, or (for comparison) of an unambiguous type. In order to ensure the reproducibility of the results, for each of the pseudo-random sets, the seed value that defines it was strictly determined.

After obtaining the sequences of pseudo-random patterns representing the different classes, test data consisting of classified items was generated. These included those of an interval type, and occasionally those that were of an unambiguous type, for comparative purposes. Each class corresponded to a set of a size of 1,000 items.

4.1 Basic Synthetic Data

This section will firstly put forward the basic form of the research conducted for the classification method developed herein that is build upon the information interval and upon uniform given patterns. In the case of a one-dimensional ($n = 1$)

pattern the first class was obtained by using a pseudo-random number generator with a normal distribution $N(0, 1)$ and the other as $N(2, 1)$. The results of this example are presented in Table 1.

The classified elements were obtained through generation by one of the aforementioned generators with normal distribution of the first pseudo-random number, as well as the second taken from a generator with uniform distribution. This defines the location of the first as within an interval of an arbitrarily assumed length. Moreover, this represents information of interval type when there are no circumstances for the considered imprecision, although its size is known. Such an interpretation seems to be the most appropriate for the majority of practical interval analysis applications. The tables below show the results with the following size of patterns: 10, 20, 50, 100, 200, 500 and 1000. In the mentioned tables, each cell contains the results obtained from 100 tests, giving an average classification error that is defined on the basis of these 100 random samples.

Table 1. Average classification error for the basic concept of Neural Classification

interval length no. of elements	0.00	0.1	0.25	0.5	1.00	2.00	5.00
10	0.1713	0.1720	0.1720	0.1723	0.1729	0.1761	0.1944
20	0.1655	0.1669	0.1669	0.1672	0.1680	0.1713	0.1888
50	0.1602	0.1605	0.1606	0.1609	0.1617	0.1652	0.1848
100	0.1596	0.1601	0.1602	0.1604	0.1615	0.1650	0.1827
200	0.1596	0.1602	0.1604	0.1609	0.1618	0.1650	0.1840
500	0.1591	0.1595	0.1596	0.1602	0.1613	0.1647	0.1844
1000	0.1579	0.1584	0.1588	0.1591	0.1603	0.1637	0.1833

4.2 Iris Benchmark Data

In the following studies, a real data set was employed. This is derived from a well-known repository located at the Center for Machine Learning and Intelligent Systems at the University of California, Irvine. The pattern set and the reference sample testing are not distinguished. For this reason, in the study, data elements were randomly divided into subsets of elements which include both reference patterns and test samples. The results shown in Table 2 are the average of 1000 tests made into random divisions. The intervals were generated in the same manner as in previous studies.

The results that are displayed, underline the many positive features of employing the classification methods mentioned in this paper. The first is small in practice, and, while sensitivity is often mentioned as being the curse of dimensionality, yet, herein, the classification of a four-dimensional feature vector has been satisfactorily performed on the basis of patterns containing about 25 items. Additional confirmation of the effectiveness of the method proposed within this paper was obtained by comparison with the results presented in [7] for the unambiguous data. In the aforementioned article, a classification error of no less

Table 2. The results of the numerical verification for the *Iris data*

interval length	0.00×0.00	0.10×0.10	0.25×0.25	0.50×0.50	1.00×1.00	2.00×2.00
mean error	0.041	0.045	0.047	0.048	0.049	0.066

than 4.5 % was obtained. A similar result was obtained in this study for the unambiguous data (cf. second column of the Table 2). Despite reducing the accuracy of the classified information by processing it into inaccurate information, the results had not deteriorated to the length of the interval of 1.0 - which is worth especially underlining.

4.3 Comparison with Similar Algorithms Used for Classification

The purpose of the following research is to compare the quality of classification of inaccurate information with other works available in the literature which are suitable for adoption.

The first one is based on a method very broadly used today due to its certain advantages, that of support vectors machines; while the other is employed for comparing the number of elements of each pattern contained in the investigated interval.

The results were obtained using the technique of supporting vectors, according to the algorithm presented in the work [20]. As a result of this procedure, three types of decisions are generated: assignment of an interval element to the first or to the second class, or the lack thereof. The study considered the amount of misclassification, lack of decision, and, in addition, the total error which is the sum of bad decisions and those of the elements for which there is no decision. Information found inside the latter was classified by drawing lots in relations proportional to the number of patterns. Upon comparing the results in the base case, it is clear that the results obtained using the method of support vectors are worse by 5% to up to 50%.

The second, relatively simple method of classifying interval type information is the procedure for patterns counting. This consists of reckoning how many elements of the learning sample are contained in the interval which is under consideration. In each case study of this algorithm, the obtained results were distinguished to be within four situations: the amount of misclassification is equal to the cardinality of elements drawn from both patterns belonging to the tested element of the interval; that the interval elements do not contain any element that is referenced; that further total error is the sum of wrong decisions; and that the consequential errors draw a ratio of 0.5 and 0.5 for those cases where the number of elements of both patterns were the same.

The effects obtained from the use of the concept of counting, revealed themselves to be absolutely worse than those obtained using the method developed in this article. However, current methods for classifying interval data are not limited to those presented this subsection. There is also a very interesting algorithm with similar conditions described in [6]. A comparison of the proposed neural algorithm with the cited method will be the subject of further research.

5 Conclusions

In conclusion, the results presented in the previous section, through numeric verification, confirm the correctness of the developed herein neural classification method of the interval type for dealing with contained inaccurate information. The results were compared with those obtained when the element was classified as being uniquely defined, as well as with those gained through utilizing other algorithms commonly employed for classifying interval data. In all the studies, enlarging the cardinality patterns resulted in a decrease of the average value of the error classification. This, in practice, allows the gradual improvement of the quality of the classification as new data is acquired. Furthermore, with the increasing length of the interval, classification errors were seen to increase to a certain limit that is justified by the data structure.

These conclusions are worth emphasizing from the application point of view. This is because they indicate that it is possible to increase the quality of classification by way of enlarging the available information through placing this in the form of numerous patterns, and by accurately classifying the interval element. In practical matters, therefore, it becomes necessary to establish a compromise between the amount of available data and the quality of the results. In situations in which there are very large representations of classes, the neural networks size rapidly increases. For this reason, we recommend using a method of reducing the sample size. With respect to employing a generalized PNN on interval information, particularly advantageous results are gained through enlisting the reduction method described in [10].

What is more, if there are no previously distinguished classes before the learning process is undertaken, the data set should be divided into smaller groups by the way of utilising a clustering method. If the number of classes is known, the application of the simple *k-means* method is recommended. Otherwise, the algorithm that is required should determine the optimal number of groups during the process of clustering. An example of a procedure satisfying the above task is an algorithm based on the Kernel Estimators Methodology [8,12].

The issue of information classification on the basis of interval data can be illustratively interpreted when unambiguous examples of the patterns contain specific, precisely measured data, while the compartments represent limitations within the plans or estimates, or when it is difficult to perform the measurements. In particular, this neural method can be used for generating a classification where a set of unambiguous data is treated as being specific information from the past (for example, temperature or exchange rates), while the classification element represents the inaccuracies forecast as being naturally limiting.

References

1. Alefeld, G., Hercberger, J.: Introduction to Interval Computations. Academic Press, New York (1986)
2. Araghi, L.F., Khaloozade, H., Arvan, M.R.: Ship identification using probabilistic neural networks (PNN). In: Proceedings of the International MultiConference of Engineers and Computer Scientists, vol. 2, pp. 18–20 (2009)
3. Bascil, M.S., Oztekin, H.: A Study on Hepatitis Disease Diagnosis Using Probabilistic Neural Network. J. Med. Syst. 36, 1603–1606 (2012)
4. Brandt, S.: Data Analysis. Springer, Heidelberg (1999)
5. Jaulin, L., Kieffer, M., Didrit, O., Walter, E.: Applied Interval Analysis. Springer, Berlin (2001)
6. Kaytoue, M., Kuznetsov, S.O., Napoli, A., Duplessis, S.: Mining gene expression data with pattern structures in formal concept analysis. Information Science 181(10), 1989–2001 (2011)
7. Kotsiantis, S.B., Pintelas, P.E.: Logitboost of Simple Bayesian Classifier. Informatica 29, 53–59 (2005)
8. Kowalski, P.A., Lukasik, S., Charytanowicz, M., Kulczycki, P.: Data-Driven Fuzzy Modelling and Control with Kernel Density Based Clustering Technique. Polish Journal of Environmental Studies 17, 83–87 (2008)
9. Kowalski, P.A.: Bayesian Classification of Imprecise Interval-Type Information (in Polish). SRI, Polish Academy of Sciences, Ph.D. Thesis (2009)
10. Kowalski, P.A., Kulczycki, P.: Data Sample Reduction for Classification of Interval Information Using Neural Network Sensitivity Analysis. In: Dicheva, D., Dochev, D. (eds.) AIMSA 2010. LNCS (LNAI), vol. 6304, pp. 271–272. Springer, Heidelberg (2010)
11. Kulczycki, P.: Statistical Inference for Fault Detection: A Complete Algorithm Based on Kernel Estimators. Kybernetika 38(2), 141–168 (2002)
12. Kulczycki, P., Charytanowicz, M., Kowalski, P.A., Lukasik, S.: The Complete Gradient Clustering Algorithm: properties in practical applications. Journal of Applied Statistics 39(6), 1211–1224 (2012)
13. Kulczycki, P., Kowalski, P.A.: Bayes classification of imprecise information of interval type. Control and Cybernetics 40, 101–123 (2011)
14. Kusy, M., Kluska, J.: Probabilistic Neural Network Structure Reduction for Medical Data Classification. In: Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) ICAISC 2013, Part I. LNCS, vol. 7894, pp. 118–129. Springer, Heidelberg (2013)
15. Moore, R.E.: Interval Analysis. Prentice-Hall, Englewood Cliffs (1966)
16. Rutkowski, L.: Computational Intelligence: Methods and Techniques. Springer, Berlin (2008)
17. Specht, D.F.: Probabilistic Neural Networks. Neural Networks 3, 109–118 (1990)
18. Tran, T., Nguyen, T., Tsai, P., Kong, X.: BSPNN: boosted subspace probabilistic neural network for email security. Artif. Intell. Rev. 35, 369–382 (2011)
19. Wand, M.P., Jones, M.C.: Kernel Smoothing. Chapman and Hall, London (1995)
20. Zhao, Y., He, Q., Chen, Q.: An Interval Set Classification Based on Support Vector Machines. In: 2nd International Conference on Networking and Services, Silicon Valley, pp. 81–86 (2005)