multistage pattern recognition,
fuzzy loss function, Bayes rule

Robert BURDUK*

CASE OF FUZZY LOSS FUNCTION IN MULTISTAGE RECOGNITION ALGORITHM

The work deals with a recognition problem using a probabilistic-fuzzy model and multistage decision logic. A case where a loss function is described using fuzzy numbers has been considered. The globally optimal Bayes strategy has been calculated for this case with stage-dependent and dependent on the node of the decision tree fuzzy loss function. The obtained result is illustrated by a calculation example in which some methods for ranking fuzzy numbers were used.

1. INTRODUCTION

In many practical problems of recognition we are often to deal with uncertain and imprecise information. The first situation results from ambiguous, in general, relationships between a class, to which belongs an object being recognised, and features that describe it. However, the second situation involves, occurring, linguistic variables that have word values in the recognition problem. As an example medical diagnostics in which feature-symptom (i.e. temperature) not determines the disease, furthermore it is accessible only in a form of word description (i.e. slight temperature, elevated temperature, etc.) serves. The lack of precision of information may also result from objective reasons, from the impossibility of precise determination of values that certain variables receive.

This paper deals with such a recognition problem, in which – assuming a probabilistic model with a full information – we will assume that values of loss function are fuzzy numbers. We will also consider the so-called multistage recognition problem [6]. It consists – speaking in short – in reaching a final classification as a result of decisions and measurements of features of the object being recognised. Consecutive decisions in sequence point to less and less numerous sets of classes and simultaneously describes features, which should be measured in order to make next decision. It is convenient to describe multistage decision logic of classifier by means of decision tree, in which terminal nodes are connected with individual classes where internal nodes denote corresponding groups of classes and recognition consists in path transition from a root to an end node [7].

In the further part, after the introduction of necessary symbols, we will calculate the optimal set of recognition algorithms for individual internal nodes, minimising global quality indicator. As a criterion of optimality we will assume the mean value of the fuzzy loss function (risk), values of

* Chair of Systems and Computer Networks, Faculty of Electronics, Wroclaw University of Technology. Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland, e-mail: burduk@zssk.pwr.wroc.pl

which depends on the node and stage of the decision tree, on which an error has occurred. The presented algorithm will be illustrated by a calculation example in which some methods for ranking fuzzy numbers were used.

2. THE PROBLEM STATEMENT

Let us consider a pattern recognition problem, in which the number of classes is equal to M. Let us assume that classes were organised in (N+1) horizontal decision tree. Let us number all nodes of the constructed decision-tree with consecutive numbers of 0, 1, 2,..., reserving 0 for the root-node and let us assign numbers of classes from the M={1,2,...,M} set to terminal nodes so that each one of them is labelled with the number of the class which is connected with that node. This allows the introduction of the following notation:

M(n) – the set of numbers of nodes, which distance from the root is n, n=0,1,2,...,N.
In particular M(0)={0}, M(N)=M,

$\overline{\mathsf{M}} = \bigcup_{n=0}^{N-1} \mathsf{M}(n)$ – the set of interior node numbers (non terminal),

$\mathsf{M}_i \subseteq \mathsf{M}(N)$ – the set of class numbers attainable from the i-th node ($i \in \overline{\mathsf{M}}$),

M^i – the set of numbers of immediate descendant nodes i ($i \in \overline{\mathsf{M}}$),

m_i – number of immediate predecessor of the i–th node (i≠0),

$d(i)$ - distance nodes i from the root.

We will continue to adopt the probabilistic model of the recognition problem, i.e. we will assume that the class number of the pattern being recognised $j_N \in \boldsymbol{M}(N)$ and its observed features x are realisations of a couple of random variables j_N and X. Complete probabilistic information denotes the knowledge of a priori probabilities of classes:

$$p(j_N) = P(J_{\boldsymbol{N}} = j_N), \quad j_N \in \mathsf{M}(N) \tag{1}$$

and class-conditional probability density functions:

$$f_{j_N}(x) = f(x/j_N), \quad x \in X, \quad j_N \in \mathsf{M}(N). \tag{2}$$

Let

$$x_i \in \mathsf{X}_i \subseteq \mathsf{R}^{d_i}, \; d_i \le d, \; i \in \mathsf{M} \tag{3}$$

denote vector of features used at the i-th node, which have been selected from the vector x.

Our target now is to calculate the so-called multistage recognition strategy $\pi_N = \{\Psi_i\}_{i \in \overline{\mathsf{M}}}$., that is the set of recognition algorithms in the form:

$$\Psi_i : \mathsf{X}_i \rightarrow \mathsf{M}^i, \; i \in \overline{\mathsf{M}} \tag{4}$$

so that it minimises the mean risk, that is the mean value of the fuzzy loss function [4, 5]:

$$\widetilde{R}^{*}(\pi_N) = \min_{\Psi_{i_n},\ldots,\Psi_{i_{N-1}}} \widetilde{R}(\pi_N) = \min_{\Psi_{i_n},\ldots,\Psi_{i_{N-1}}} E[\widetilde{L}(\boldsymbol{I_N},\boldsymbol{J_N})] \tag{5}$$

The π_N^* strategy we will be called the globally optimal N-stage recognition strategy. In the next chapter we will calculate globally optimal strategy for the loss function dependent on the node of the decision tree and stage-dependent and on crisp values of the features.

3. THE RECOGNITION ALGORITHM

3.1. DEPENDENT ON THE NODE OF THE DECISION TREE FUZZY LOSS FUNCTION

Let $L(i_N, j_N)$ denotes the loss incurred when objects of the class j_N is assigned to the class i_N ($i_N, j_N \in \mathcal{M}(N)$). Let us assume now

$$\widetilde{L}(i_N, j_N) = \widetilde{L}_w \tag{6}$$

where w is the first common predecessor of the nodes i_N and j_N. The loss function so defined means that the loss depends on the node of the decision tree at which misclassification is made. Applying procedure similar to [1], we obtain searched globally optimal strategy with decision algorithms as follows:

$$\Psi_{i_n}^{*}(x_{i_n}) = i_{n+1}\text{, } while \tag{7}$$

$$(\widetilde{L}_{i_n} - \widetilde{L}_{i_{n+1}})\,p(i_{n+1})\,f_{i_{n+1}}(x_{i_n}) +$$

$$+ \sum_{j_{n+2}\in \mathcal{M}^{i_{n+1}}} [(\widetilde{L}_{i_{n+1}} - \widetilde{L}_{j_{n+2}})q^{*}(j_{n+2}/i_{n+1}, j_{n+2})\,p(j_{n+2})\,f_{j_{n+2}}(x_{i_n}) +$$

$$+ \ldots + \sum_{j_N \in \mathcal{M}^{j_{N-1}}} [\widetilde{L}_{j_{N-1}} q^{*}(j_N / i_{n+1}, j_N)\,p(j_N)\,f_{j_N}(x_{i_n})]\ldots] \quad =$$

$$= \max_{k\in \mathcal{M}^{i_n}} \Bigg\{ (\widetilde{L}_{i_n} - \widetilde{L}_k)\,p(k)\,f_k(x_{i_n}) +$$

$$+ \sum_{j_{n+2}\in \mathcal{M}^{k}} [(\widetilde{L}_k - \widetilde{L}_{j_{n+2}})q^{*}(j_{n+2}/k, j_{n+2})\,p(j_{n+2})\,f_{j_{n+2}}(x_{i_n}) +$$

$$+ \ldots + \sum_{j_N \in \mathcal{M}^{j_{N-1}}} [\widetilde{L}_{j_{N-1}} q^{*}(j_N / k, j_N)\,p(j_N)\,f_{j_N}(x_{i_n})]\ldots] \Bigg\},$$

for $i_n \in \mathrm{M}(n)$, n=0,1,2,...,N–1, where $q^*(j_N / i_{n+1}, j_N)$ denotes probability of accurate classification of the object of the class j_N in further stages using π_N^* strategy rules on condition that on the n-th stage the i_{n+1} decision has been made.

3.2. STAGE-DEPENDENT FUZZY LOSS FUNCTION

Let us assume now

$$\widetilde{L}(i_N, j_N) = \widetilde{L}^s_{\mathrm{d}(w)}, \tag{8}$$

where w is the first common predecessor of the nodes i_N and j_N ($i_N, j_N \in \mathrm{M}(N)$). So defined fuzzy loss function means that the loss depends on the stage at which misclassification is made. Stage-dependent and dependent on the node of the decision tree fuzzy loss function are presented in Fig. 1.

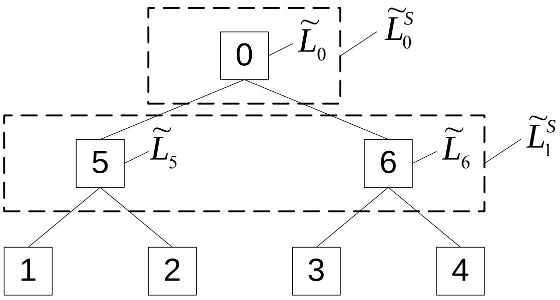

Fig.1. Stage-dependent and dependent on the node of the decision tree fuzzy loss function

Applying procedure similar to [1], we obtain searched globally optimal strategy with decision algorithms as follows:

$$\Psi^*_{i_n}(x_{i_n}) = i_{n+1}, \textit{ while} \tag{9}$$

$$(\widetilde{L}^s_{\mathrm{d}(i_n)} - \widetilde{L}^s_{\mathrm{d}(i_{n+1})})\, p(i_{n+1})\, f_{i_{n+1}}(x_{i_n}) +$$

$$+ \sum_{j_{n+2} \in \mathrm{M}^{i_{n+1}}} [(\widetilde{L}^s_{\mathrm{d}(i_{n+1})} - \widetilde{L}^s_{\mathrm{d}(j_{n+2})}) q^*(j_{n+2} / i_{n+1}, j_{n+2})\, p(j_{n+2})\, f_{j_{n+2}}(x_{i_n}) +$$

$$+ \ldots + \widetilde{L}^s_{\mathrm{d}(j_{N-1})} \sum_{j_N \in \mathrm{M}^{j_{N-1}}} [q^*(j_N / i_{n+1}, j_N)\, p(j_N)\, f_{j_N}(x_{i_n})]\ldots] =$$

$$= \max_{k \in \mathrm{M}^{i_n}} \Big\{ (\widetilde{L}^s_{\mathrm{d}(i_n)} - \widetilde{L}^s_{\mathrm{d}(k)})\, p(k)\, f_k(x_{i_n}) +$$

$$+ \sum_{j_{n+2} \in \mathrm{M}^{k}} [(\widetilde{L}^{s}_{\mathrm{d}(k)} - \widetilde{L}^{s}_{\mathrm{d}(j_{n+2})}) q^{*}(j_{n+2}/k, j_{n+2}) p(j_{n+2}) f_{j_{n+2}}(x_{i_n}) +$$

$$\left. + \ldots + \widetilde{L}^{s}_{\mathrm{d}(j_{N-1})} \sum_{j_N \in \mathrm{M}^{j_{N-1}}} [q^{*}(j_N/k, j_N) p(j_N) f_{j_N}(x_{i_n})] \ldots] \right\} .$$

4. ILLUSTRATIVE EXAMPLE

Let us consider the two-stage binary classifier of Fig. 1. Four classes have identical a priori probabilities which are equal 0.25. Class-conditional probability density functions of features $\boldsymbol{X_5}$ i $\boldsymbol{X_6}$ are presented in Fig. 2.

The feature $\boldsymbol{X}_0$ is normally distributed in each class with the following class-conditional probability density functions:

$$f_1(x_0) = f_2(x_0) = N(1,1), \quad f_3(x_0) = f_4(x_0) = N(3,1)$$

The fuzzy loss function dependent on the node of the decision tree are the following:

<u>case 1</u> $\widetilde{L}_0 = (1{,}5,\ 2,\ 2{,}5,\ 3)_{Tr}$, $\widetilde{L}_5 = (0,\ 0,\ 0{,}5,\ 1)_{Tr}$, $\widetilde{L}_6 = (0{,}5,\ 1,\ 1{,}5,\ 2)_{Tr}$,

<u>case 2</u> $\widetilde{L}^S_0 = (1{,}5,\ 2,\ 2{,}5,\ 3)_{Tr}$, $\widetilde{L}^S_1 = (0,\ 0,\ 0{,}5,\ 1)_{Tr}$,

<u>case 3</u> $\widetilde{L}^S_0 = (1{,}5,\ 2,\ 2{,}5,\ 3)_{Tr}$, $\widetilde{L}^S_1 = (0{,}5,\ 1,\ 1{,}5,\ 2)_{Tr}$.

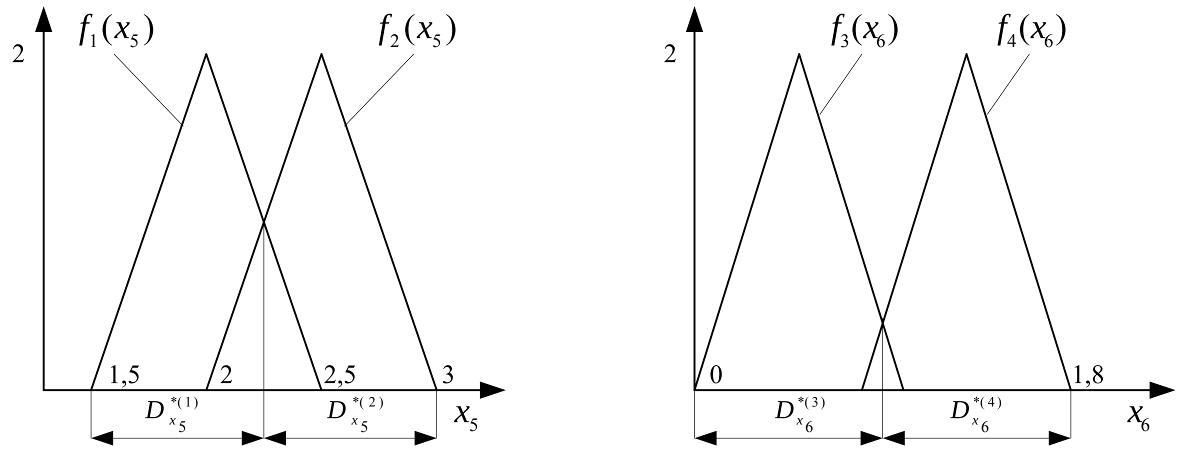

Fig.2. Illustrate of example – class conditional probability density function and decision regions at the second stage

Due to the peculiar distribution of $\boldsymbol{X_5}$ and $\boldsymbol{X_6}$, the decision rules Ψ^{*}_{5} and Ψ^{*}_{6}, at the second stage of classification, are obvious. Their decision regions are shown in Fig. 2. Let us now determine the rule Ψ^{*}_{0} at the first stage of classification. From (7) we have in case 1:

$$\Psi_0^*(x_0)=\begin{cases}5, & \text{if } (\widetilde{L}_0-\widetilde{L}_5)\,p(5)\,f_5(x_0)+\\ & +\widetilde{L}_5(q^*(1/5,1)\,p(1)\,f_1(x_0)+q^*(2/5,2)\,p(2)\,f_2(x_0))>\\ & >(\widetilde{L}_0-\widetilde{L}_6)\,p(6)\,f_6(x_0)+\\ & +\widetilde{L}_6(q^*(3/6,3)\,p(3)\,f_3(x_0)+q^*(4/6,4)\,p(4)\,f_6(x_0)),\\ 6 & \text{otherwise,}\end{cases}$$

Putting now the data of the example, and using Yager [8], Chen [3] and Campos-Gonzales [2] method for comparison fuzzy risk, we finally obtain results presented in Tab. 1.

Tab.1. Results for some method for ranking fuzzy numbers

case	Method				
	Yager	Chen	Campos-Gonzalez		
			λ=0	λ=0,5	λ=1
1	1,994	1,996	2,153	1,995	1,956
2	1,990	1,992	1,999	1,991	1,988
3	1,969	1,969	1,943	1,969	1,974

5. CONCLUSION

The multistage recognition strategy obtained in the paper, which minimises the global fuzzy mean risk for fuzzy loss function poses a generalisation of the result presented in [6]. It may pose a good starting point for consideration of other cases of multistage recognition with a fuzzy factor and in particular recognition with a learning set.

BIBLIOGRAPHY

[1] BURDUK R., Komputerowe algorytmy rozpoznawania wieloetapowego z modelem probabilistyczno-rozmytym i ich zastosowanie w medycynie, Raport Katedry Systemów i Sieci Komputerowych, Wrocław, 2003 (in Polish).

[2] CAMPOS L.M., GONZÁLEZ A., A Subjective Approach for Ranking Fuzzy Numbers, Fuzzy Sets and Systems 29, 145-153, 1989.

[3] CHEN S.H., Ranking Fuzzy Numbers with Maximizing Set and Minimizing Set, Fuzzy Sets and Systems 17, 113-129, 1985.

[4] GIL M.A., LÓPEZ-DÍAZ M., Fundamentals and Bayesian Analyses of Decision Problems with Fuzzy-Valued Utilities, International Journal of Approximate Reasoning, 203-224, 1996.

[5] GIL M.A., LÓPEZ-DÍAZ M., A Model for Bayesian Decision Problems Involving Fuzzy-Valued Consequences, Proc. 6th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge Based Systems, Granada, Spain, 1996.

[6] KURZYŃSKI M., Rozpoznawanie Obiektów - Metody Statystyczne, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 1997 (in Polish).

[7] KURZYŃSKI M., Algorytmy rozpoznawania wieloetapowego oraz ich zastosowania medyczne i techniczne, Prace naukowe Inst. Sterowania I Techniki Systemów Pol. Wr. Nr 5, Seria Monografie Nr 3, Wrocław, 1987 (in Polish).

[8] YAGER R., A Procedure for Ordering Fuzzy Subsets of the Unit Interval, Information Sciences 24, 143-161, 1981.