New Method for Generation Type-2 Fuzzy Partition for FDT

Łukasz Bartczuk[1], Piotr Dziwiński[1], and Janusz T. Starczewski[1,2,⋆]

[1] Department of Computer Engineering, Czestochowa University of Technology, Al. Armii Krajowej 36, 42-200 Czestochowa, Poland
{lukasz.bartczuk,piotr.dziwinski,janusz.starczewski}@kik.pcz.pl
[2] Academy of Management (SWSPiZ), Institute of Information Technology, ul. Sienkiewicza 9, 90-113 Łódź, Poland

Abstract. One of the most important tasks during application of fuzzy decision tree algorithms is to generate a fuzzy partition. In this paper, we introduce a new method to perform this task. The proposed method is a two stage process. The second stage is based on the classical Fuzzy C-means (FCM) clustering.

1 Introduction

Decision trees are common methods in data mining and machine learning. They are appreciated because of theirs clarity and interpretability as well as for the attribute selection mechanism. The most popular algorithms to create such structures are CART [1] and ID3 [2] or its successor C4.5 [3].

Fuzzy decision trees (FDTs) of type-1 [4] or of type-2 [5] are generalized versions of classical decision trees and combine their advantages with possibilities of modelling an uncertain or imprecise knowledge. These algorithms require that each continuous attribute have to be previously split into a number of fuzzy intervals.

In this paper, a new method to generate type-2 fuzzy partition is presented. The method takes into account an inner uncertainty that is contained in a modeled process. The paper is organized as follows: the next section briefly describes the type-2 fuzzy decision trees, section 3. depicts a proposed method to obtain type-2 fuzzy sets, section 4 illustrates experimental results, and the last section draws final conclusions.

2 Interval Type-2 Fuzzy Decision Trees

The interval type-2 fuzzy decision trees [5,6] are created by a recursive procedure from data set E and attribute set A. Every pattern contained in the data set is a

⋆ This work was partly supported by Polish Ministry of Science and Higher Education (Habilitation Project N N516 372234 2008–2011, Polish-Singapore Research Project 2008–2010).

L. Rutkowski et al. (Eds.): ICAISC 2010, Part I, LNAI 6113, pp. 275–280, 2010.

vector $\mathbf{x}_t = \left[x_1(t), \ldots, x_{|A|}(t), y(t)\right]$ consisting both input values and a value of decision attribute. Every attribute $A^k \in A$, $k = 1, \ldots, |A|$, contains $|A^k|$ different values a_m^k and describes some characteristic feature of objects from the data set. All values included in the attribute set are characterized by some interval type-2 fuzzy sets.

Let us remind that the interval fuzzy set of type-2 [7,8,9] defined in $X \subset \mathbb{R}$, is an ordered collection of pairs $\{x, \mu_{\tilde{A}}(x)\}$ where $x \in X$, and $\mu_{\tilde{A}}(x)$ is a fuzzy grade of membership defined on $[0,1]$ interval as $\mu_{\tilde{A}}(x) = \int_{u \in [0,1]} 1/u$, where u is called a primary membership grade. To simplify the notation, the interval fuzzy membership grade can be represented as minimum and maximum points of its support: $\mu_{\tilde{A}}(x) = \left[\underline{\mu_{\tilde{A}}(x)}, \overline{\mu_{\tilde{A}}(x)}\right]$.

In each step of recursion for the decision tree creating procedure, the data set is split, according to a chosen attribute A^k, into S fuzzy subsets, where $2 \leq S \leq |A^k|$. The choice of the attribute is performed by an algorithm based on the assumption that resulting data subsets have a higher degree of homogeneity than a complete data set. The detailed description of this algorithm can be found in paper [5].

2.1 Inference Process

Every decision tree can be considered as a set of decision rules, whose number is equal to the number of leaves in the tree. Obviously in case of the interval type-2 fuzzy decision tree, the type-2 fuzzy decision rule have to be under consideration, and can be presented in the following form:

$$R^o : \mathsf{IF} \bigcap_{\tilde{a}_m^k \in \mathcal{P}^o} \left(x^k \ \mathsf{IS} \ \tilde{a}_m^k\right) \ \mathsf{THEN} \ y \ \mathsf{IS} \ \tilde{a}_o^D; \quad o = 1, \ldots, |O|, \tag{1}$$

where $|O|$ denotes a number of leaves (rules), $\tilde{a}_o^D$ — a value of the decision attribute assigned to leaf o, $\mathcal{P}^o$ — a path from a root node to leaf o. A premise part of this rule is constituted by values of attributes, which can be found on path $\mathcal{P}^o$, for which this rule has been generated. A conclusion part contains the value of decision attribute $\tilde{a}_o^D$ that have been assigned to this leaf by the procedure creating a tree structure.

At the beginning of the process of classification previously unseen patterns, an activation grade for each fuzzy decision rule has to be computed as the t-norm combination of all type-2 fuzzy sets from the premise part of rule (1):

$$\mu_{A^o}(\mathbf{x}) = \mathop{\tilde{\mathsf{T}}}_{\tilde{a}_m^k \in \mathcal{P}^o} \mu_{\tilde{a}_m^k}(x_k) = \left[\mathop{\mathsf{T}}_{\tilde{a}_m^k \in \mathcal{P}^o} \underline{\mu_{\tilde{a}_m^k}}(x_k), \mathop{\mathsf{T}}_{\tilde{a}_m^k \in \mathcal{P}^o} \overline{\mu_{\tilde{a}_m^k}}(x_k)\right], \tag{2}$$

where: $\mu_{A^o}(\mathbf{x})$ is the activation grade of the rule o.

If we also assume that all values of the decision attribute are singletons, the activation grade and the conclusion grade are equal, so that:

$$\underline{\tau} = \underline{\mu_{A^o}}(\mathbf{x}) * \underline{\mu_{\tilde{a}_o^D}}(y) = \underline{\mu_{A^o}}(\mathbf{x}), \tag{3}$$

$$\overline{\tau} = \overline{\mu_{A^o}}(\mathbf{x}) * \overline{\mu_{\tilde{a}_o^D}}(y) = \overline{\mu_{A^o}}(\mathbf{x}). \tag{4}$$

When we determine activated rules, we can take a decision, into which value of the decision attribute the pattern have to be assigned. In the simplest case, we can find the rule with the highest activation grade. However, a single pattern can activate more than one rule, possibly with the same consequent. Thus better results may be obtained if we use a total activation grade, that is defined for each value of the decision attribute as follows:

$$\mu_{\tilde{a}_m^D}(\mathbf{x}) = \underset{o:\mu_{\tilde{a}_o^D}=\mu_{\tilde{a}_m^D}}{\widetilde{S}} \mu_{A^o}(\mathbf{x}) = \left[\underset{o:\mu_{\tilde{a}_o^D}=\mu_{\tilde{a}_m^D}}{S} \underline{\mu_{A^o}}(\mathbf{x}), \underset{o:\mu_{\tilde{a}_o^D}=\mu_{\tilde{a}_m^D}}{S} \overline{\mu_{A^o}}(\mathbf{x}) \right]. \quad (5)$$

In both considered cases, we have to compare interval values. This can be performed by one of the following method: MAX, AVG, MIN-MAX, and comparison of fuzzy intervals (CFI).

The MAX method searches for an interval value with the highest upper bound. The AVG method is very similar but it uses averages of intervals. The MIN-MAX method identifies intervals with the highest upper and lower bounds, separately. If both indicate the same interval, the corresponding value is chosen. If two different intervals are found, then the result is undetermined. The last method is based on an algorithm comparing fuzzy numbers, proposed in [10]. This algorithm determines a degree from $[0, 1]$ of satisfying an inequality between two fuzzy quantities A_1 and A_2. The value 1 for this degree indicates a full certainty that A_1 is greater than A_2, 0 indicates $A_1 < A_2$, and 0.5 is reserved for equal fuzzy quantities. In the simulations, we assume the correct classification if the fuzzy interval firing degree of a rule satisfies the inequality with all remaining rules with the inequality degree greater than 0.6. Otherwise, the system cannot make any certain decision.

3 Generation of Type-2 Fuzzy Partition

In this section, we would like to propose a new method for generating type-2 fuzzy partition. This is a two-phase method, in which the first phase is responsible for determining the parameters of type-1 fuzzy membership functions, and the second phase creates type-2 fuzzy membership functions.

3.1 Phase 1: Determining Parameters of Type-1 Fuzzy Membership Functions

In this phase, the parameters of type-1 fuzzy membership function are determined. For the purposes of this paper we use the asymmetric Gaussian membership function, which is defined by following formula:

$$a\_gauss\,(x; m, \sigma_l, \sigma_r) = \begin{cases} \exp\left[-\left(\frac{x-m}{\sigma_l}\right)^2\right] & \text{for } x < m, \\ \exp\left[-\left(\frac{x-m}{\sigma_r}\right)^2\right] & \text{otherwise,} \end{cases} \quad (6)$$

where m, σ_l, σ_r are respectively: the center of the function, its left and right spreads.

We assume that this phase is performed using the FCM (Fuzzy C-Means) clustering algorithm [11]. Let the number of cluster centers obtained as a result of the FCM algorithm be denoted by C. Each cluster center is defined as a vector $\mathbf{v}_c = \left[m_1^c, \ldots, m_{|A|}^c\right]$. The value m_c^k is also the center of the c^{th} membership function defined for the k^{th} attribute.

The left and right spreads of the asymmetrical Gaussian function can be computed independently for ${E_l}_n^k = \left\{x^k(t) : x_n^k(t) \le m_n^k\right\}$ and ${E_r}_n^k = \left\{x^k(t) : m_n^k < x_n^k(t)\right\}$ with the following formula:

$$\sigma{\delta}_n^k = \sqrt{\frac{\sum_{t=1}^{|E{\delta}_n^k|} u_{kt}\left[x^k(t) - m_n^k\right]^2}{\sum_{t=1}^{|E{\delta}_n^k|} u_{kt}}}, \tag{7}$$

where: $\delta = \{l, r\}$ signifies the left or right spread, u_{kt} is a fuzzy membership value computed by the FCM method.

3.2 Phase 2: Creating Type-2 Membership Function

The actual degree of membership assigned by the FCM algorithm and computed on the base of the membership functions specified in the previous phase will vary for real-world data. If we assume that training data were collected without a measurement error, this difference between membership grades is caused by an inner uncertainty of a modeled process. Therefore, we can expand upper and lower membership functions over training data, such that memberships of all data points are covered by the footprint of uncertainty of a type-2 fuzzy membership function, i.e. the area between the upper and the lower membership function. Consequently, the upper membership function can be a normal asymmetrical Gaussian function,

$$\mu_{A_n^k}(x_n) = a\_gauss(x_n, m_n^k, \overline{\sigma l}_n^k, \overline{\sigma r}_n^k)$$

being a superior limit of the function family drawn through points $(x_n(t), u_{kt})$, i.e.,

$$\overline{\sigma l}_n^k = \max_t \sigma_{lt} : \left\{a\_gauss(x_n(t), m_n^k, \sigma_{lt}, \sigma_r) \mid x_n(t) \le m_n^k\right\} \tag{8}$$

$$\overline{\sigma l}_n^k = \max_t \frac{m_n^k - x_n(t)}{\sqrt{-\log u_{kt}}} \tag{9}$$

and

$$\overline{\sigma r}_n^k = \max_t \sigma_{rt} : \left\{a\_gauss(x_n(t), m_n^k, \sigma_l, \sigma_{rt}) \mid m_n^k < x_n(t)\right\} \tag{10}$$

$$\overline{\sigma r}_n^k = \max_t \frac{x_n(t) - m_n^k}{\sqrt{-\log u_{kt}}} \tag{11}$$

The lower membership function can be a scaled (by h) asymmetrical Gaussian function,

$$\underline{\mu}_{A_n^k}(x_n) = \underline{h}_n^k a\_gauss(x_n(t); m_n^k, {\sigma_l}_n^k, {\sigma_r}_n^k) \tag{12}$$

being an inferior limits of the function family drawn through points $(x_n(t), u_{kt})$, i.e.,

$$\underline{h}_n^k = \min_t h_t : \left\{ h_t a\_gauss(x_n(t); m_n^k, {\sigma_l}_n^k, {\sigma_r}_n^k) \right\} \tag{13}$$

$$\underline{h}_n^k = \min_t \frac{u_{kt}}{a\_gauss(x_n(t); m_n^k, {\sigma_l}_n^k, {\sigma_r}_n^k)} \tag{14}$$

4 Experiments

To illustrate the capability of the proposed method, the Wisconsin Breast Cancer Dataset from UCI Machine Learning Repository [12] has been used. This dataset contains description of 699 cases from two classes. Benign class is assigned to 458 patterns and Malignant class to 241 patterns. Each pattern is characterized by 10 numerical attributes. The complete dataset consists 16 instances with missing values which have been removed.

The whole dataset has been split into two subsets used for training and testing. The training set has consisted of 547 and testing set — 136 patterns chosen randomly. The training set has been used to generate type-2 fuzzy partition as well as to build the decision tree. The testing set has been applied to determine the classification rates of the obtained decision tree.

The experiments have been also performed for a type-2 fuzzy partition method described in paper [6]. The results are summarized in Tables 1-2.

As we can see the interval fuzzy decision trees of type-2 constructed with fuzzy sets generated by the algorithm based on FCM have lower incorrect classification rates for the MIN-MAX and CFI inference methods than rates for the method based on uncertain fuzzy clustering. Unfortunately, these structures are also characterized by lower rates of correct classification. However we believe that

Table 1. Results obtained for the proposed method

	MAX	AVG	MIN-MAX	CFI
correct classification	0.93	0.96	0.84	0.84
incorrect classification	0.07	0.04	0.01	0.01
undecided	-	-	0.15	0.15

Table 2. Results obtained for the method described in paper [6]

	MAX	AVG	MIN-MAX	CFI
correct classification	0.96	0.96	0.96	0.96
incorrect classification	0.04	0.04	0.04	0.04
undecided	-	-	0	0

in case of medical diagnosis problems, it is better that system cannot give the answer than it gives the wrong answer.

5 Conclusion

In this paper, the new method for generation of type-2 fuzzy partition have been proposed. The method has been thought as a two stage process. In the first stage parameters for each asymmetrical Gaussian membership function have been determined. The second stage has generated type-2 fuzzy membership functions. The source of these interval type-2 membership functions has been regarded as an inner uncertainty contained in a modeled process. The obtained fuzzy partitions have been used to construct type-2 interval fuzzy decision trees.

References

1. Breiman, L., Friedman, J., Olshen, R., Stone, C.: Classification and regression trees. Wadsworth, Belmont (1984)
2. Quinlan, J.: Induction of decision trees. Machine Learning 1, 81–106 (1986)
3. Quinlan, J.: C4.5 Programs for Machine Learning. Morgan Kaufmann, San Francisco (1993)
4. Janikow, C.: Fuzzy decision trees: Issues and methods. IEEE Trans Systems, Man, Cybernetics - Part B: Cybernetics 28, 1–14 (1998)
5. Bartczuk, L., Rutkowska, D.: Fuzzy decision trees of type-2. In: Some Aspects of Computer Science. EXIT Academic Publishing House, Warsaw (2007) (in Polish)
6. Bartczuk, L., Rutkowska, D.: Type-2 fuzzy decision trees. In: Rutkowski, L., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) ICAISC 2008. LNCS (LNAI), vol. 5097, pp. 197–206. Springer, Heidelberg (2008)
7. Castillo, O., Melin, P.: Intelligent systems with interval type-2 fuzzy logic. International Journal of Innovative Computing, Information and Control 4(4), 771–783 (2008)
8. Liang, Q., Mendel, J.: Interval type-2 fuzzy logic systems: Theory and design. IEEE Transactions on Fuzzy Systems 8, 535–550 (2000)
9. Mendel, J., John, R.: Interval type-2 fuzzy logic systems made simple. IEEE Transactions on Fuzzy Systems 10, 622–639 (2002)
10. Dorohonceanu, B.: Comparing fuzzy numbers. Dr. Dobb's Journal 343, 38–45 (2002)
11. Bezdek, C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York (1981)
12. UCI, M.L.R.: `http://mlearn.ics.uci.edu/mlrepository.html`