Classification by Ordered Fuzzy Decision Tree

Jan Rabcan, Makpal Zhartybayeva

***Abstract*—** Nowadays, the classification represents a significant part of the data mining. The object of the classification is assigned to the new data sample the output property (class label) based on previous, learned experience. In this paper the approach of ordered fuzzy decision tree is considered. The fuzzy logic can reduce the uncertainness of initial data and it is closer to natural way of human thinking. Chosen classification model is evaluated by estimation of error and accuracy of the resulting classification.

***Keywords*—**Classification, Fuzzy, Decision tree, Data mining.

I. Introduction

Data mining is the process of analysis of data from the various perspective and summarization the results on useful information [1]. The goal of data mining is to discover salutary knowledge stored in huge databases and repositories [1]. Data mining includes many techniques like prediction, classification, clustering, association rules, estimation and affinity grouping [2]. One of the mentioned data mining tasks is classification. The aim of the classification is to assign the class label for new instance. The popular way of classification is decision tree technique. Nowadays, there are many methods for induction of decision tree. One of the first decision tree technique has been published by J. R. Quinlan in [3]. The main idea of the ID3 algorithm is to choose the associate attribute to each node with minimal entropy or maximal information gain [3]. J. R. Quinlan modified ID3 algorithm in [4]. The modified version is called C4.5. Also C4.5 algorithm deals with information entropy. The splitting criterion is the normalized information gain [4]. Many real words problems are uncertainties and noisy. In this case, the crisp classification can be difficult to perform. The usage of fuzzy sets can be useful to describe real-world problems with higher accuracy [5] and more naturally to the way of human thinking. For this reason, the fuzzy decision trees are considered in this paper. At present time, many algorithms for induction of fuzzy decision tree have been proposed. One popular method has been described by Yuan and Shaw in [6]. The induction is based on the reduction of classification ambiguity with fuzzy evidence. Another way of FDT induction is based on fuzzy rules and published by Xianchang Wang in [7]. In contrast with "traditional" decision trees in which only a single attribute is taken into account at an each node, the node of the proposed decision trees in [7] involves a fuzzy rule which take into account multiple attributes. The next approach has been presented in [8]. This algorithm for Ordered Decision Tree (OFDT) is proposed in [8] and used for needs of this paper. OFDT algorithm takes only one attribute to each level of the decision tree. This feature can be considered as an advantage, because it allows constructing OFDT as a parallel process [5]. The criterion to choose attribute associated with given level is cumulative information estimations of fuzzy sets [9].

The usage of considered classification method has been evaluated on well know public dataset *Pima Indians Diabetes Database*. The dataset contains medical records of female patients and the goal is to estimate whatever a patient has signs of diabetes or not. Evaluation of OFDT is done by estimation of error and accuracy of the resulting classification.

II. Fuzzy logic

Fuzzy logic is one way to represent multi-valued logic. In classical Boolean logic, variables can reach only "crisp" values 1 or 0. Fuzzy logic describes variables using membership function

J. Rabcan, University of Zilina, Faculty of Management Science and Informatics, Zilina, Slovakia (e-mail: jan.rabcan@fri.uniza.sk)
M. Zhartybayeva, Faculty of Information Technology, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan.

where variables are between 0 and 1 which represent degrees of membership [10]. Let set U is Universe of discourse. Universe of discourse contains all investigated samples. The fuzzy set A in set U has following definition. The set A is characterized by membership function as follow [11]:

1. $\mu_A(u) = 0$ if and only if x is not the member of set A
2. $\mu_A(u) = (0,1)$ if and only if x is not the full member of set A
3. $\mu_A(u) = 1$ if and only if x is the full member of set A

The function $\mu_A(u)$ is the membership function. This function for all elements in U assigns the value between 0 and 1. Fuzzy set is defined as an ordered set of pairs [11]:

$$A = \{(u, \mu_A(u)), u \in A\} \tag{1}$$

The following example illustrates usage of fuzzy logic on speed variable. If speed of car traveling on the highway is 100 km/h, it is hard to tell, if car is moving fast or not. In classical "crisp" logic the result is *low speed = 0, medium speed =1, high speed = 0*. In fuzzy logic the result is *low speed = 0, medium speed =0.6, high speed = 0.3*.

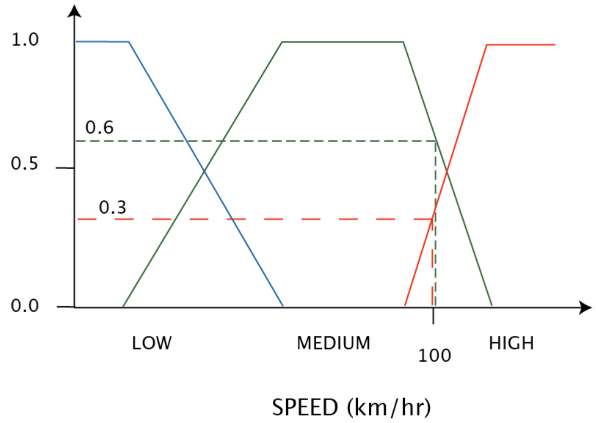

Fig. 1 Fuzzy set of the car movement speed

III. Data

The repository for data storage is represented in tabular form (Table 1). The table consists of n+1columns that correlate with n input attributes and 1 output attribute. The i-th column, for $i = 1,\ldots, n+1$, is divided into q_i sub-columns. The j_i-th sub-column, for $j_i = 1,\ldots, q_i$, corresponds to the j-th value of the attribute represented by the i-th column. Every row of the table represents one sample of initial (collected) data.

The proposed classification system deals with fuzzy attributes. Fuzzy attribute A_i is a linguistic attribute, and its possible values are also stored in the repository. Each value $A_{i_q} \in \langle 0,1 \rangle$ and $\sum_{q=1}^{q_i} A_q = 1$. These possibilities correspond to a membership function of fuzzy data [12]. These requirements on representation of initial data are caused by the method of FDT induction.

TABLE 1
TRAINING AND TESTING DATASETS

Input attributes, Ai								Output attribute, B		
	A1		A2			An				
A11	A12	A13	A21	A22	....	An1	An2	B1	B2	B3
0.1	0.5	0.4	0.6	0.4	...	0.5	0.5	0.0	0.0	1.0
0.2	0.1	0.7	0.1	0.9	...	0.8	0.2	0.3	0.7	0.0
	...		...		...	...			...	
0.3	0.3	0.4	0.0	1.0	...	0.4	0.6	0.1	0.4	0.5

Repository (data collection) for the FDT induction.
Results of measurement, expert evaluation, or monitoring

The used dataset also contains numerical attributes. OFDT works only with fuzzy attributes. Therefore, all numerical attributes must be fuzzied. Yuan and Shaw have suggested a simple method of how to generate a set of membership functions for transforming numeric data to fuzzy data, which is described in [1]. Let attribute A be numeric attribute. The algorithm in first step divides values x_{i} of A into Q intervals $\{Q_1, \dots, Q_q, \dots, Q_Q\}$ by using arbitrary clustering algorithm. Every interval Q_q contains the center C_q. Then the membership functions are created based on the following rules. For the first linguistics term $x_{i,1}$, the next membership function is used [5]:

$$\mu_{x_{i,1}}(x) = \begin{cases} 1 & x \leq C_1 \\ \frac{C_2 - x}{C_2 - C_1} & C_1 < x < C_2 \\ 0 & x \geq C_2 \end{cases} \tag{2}$$

Every linguistics term$x_{i,q}$, for $q = 2,3, \dots, N-1$, has a membership function of the following form:

$$\mu_{x_{i,q}}(x) - \begin{cases} 0 & x \leq C_{j-1} \\ \frac{x - C_{j-1}}{C_j - C_{j-1}} & C_{j-1} < x \leq C_j \\ \frac{C_{j+1} - x}{C_{j+1} - C_j} & C_j < x \leq C_{j+1} \\ 0 & x \geq C_{j+1} \end{cases} \tag{3}$$

The last term $x_{i,Q}$ has the membership function of the form of:

$$\mu_{x_{i,Q}}(x) = \begin{cases} 0 & x \leq C_{k-1} \\ \frac{x - C_{k-1}}{C_k - C_{k-1}} & C_{k-1} < x \leq C_k \\ 1 & x \geq C_k \end{cases} \tag{4}$$

The following example illustrates the fuzzification method, where a numerical attribute represent the age of female. The investigated attribute is transformed into four groups: young, early adulthood, middle-aged and old age (Fig. 2). The first set young and the last set old have a form which can be described by four corners and the form is trapezoidal [1], while the sets between the first and the last set has triangular form. Therefore, they are described by three corners. For example, the set young is described by (0, 0, 16, 32), while the set early adulthood can be expressed using (16, 32, 48). Fuzzy set of investigated attribute (age) is shown on Fig. 2.

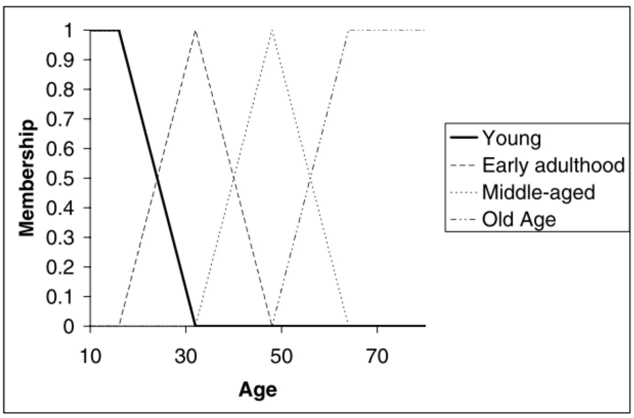

Fig. 2 Membership functions obtained after transformation

IV. ORDERED FUZZY DECISION TREE

An Ordered FDT (OFDT) is a specific type of decision trees, which has exactly one attribute in every level of the tree. The algorithm for OFDT induction has two main aspects. The first is choosing expandable attributes for each node, and the second is establishing the leaf nodes. Expandable attribute is chosen by cumulative mutual information [13]. Usage of this information characteristic allows parallel tree induction. The cumulative mutual information is estimated as follows:

$$\Delta\mathbf{I}\left(B; A_{i_1}, \dots, A_{i_{q-1}}, A_{i_q}\right) = \mathbf{I}\left(B; A_{i_1}, \dots, A_{i_{q-1}}, A_{i_q}\right) - \mathbf{I}\left(B; A_{i_1}, \dots, A_{i_{q-1}}\right) \quad (5)$$

The attribute with the greatest information is chosen to associate with all nodes of the given level. The criterion for choosing expandable attributes of the OFDT can take into account the cost to measure value of attribute A. Then the criterion to choose attribute is following:

$$q = \operatorname{argmax}\left(\frac{\Delta\mathbf{I}\left(B; A_{i_1}, \dots, A_{i_{q-1}}, A_{i_q}\right)}{Cost\left(A_{i_q}\right)}\right) \quad (6)$$

The OFDT algorithm has to establish leaf node during the induction phase. The presented algorithm uses a defined threshold values α and β. The threshold β represents the confidence level. The threshold α reflects the frequency of occurrences in the given node. Every internal node of the tree is declared as a leaf if at least one of the following conditions is satisfied:

$$f\left(U_{i_q j_q}\right) = \frac{M\left(A_{i_1 j_1} \times \dots \times A_{i_q j_q}\right)}{N} \leq \alpha \quad (7)$$

$$2^{-\mathbf{I}\left(B | A_{i_1 j_1}, \dots, A_{i_q j_q}\right)} \geq \beta \quad (8)$$

where $U_{i_q j_q} = \left\{A_{i_1, j_1}, \dots, A_{i_1, j_1}\right\}$ means values of input attributes and $M\left(A_{i,j}\right) = \sum_{k=1}^{N} \mu_{Aj}(x_k)$. $\mathbf{I}\left(A_{i_1 j_1} | A_{i_2 j_2}\right) = \log_2 M\left(A_{i_1 j_1}\right) - \log_2 M\left(A_{i_2 j_2} \times A_{i_1 j_1}\right)$. Steps of the OFDT algorithm are described in Table 2.

TABLE 3
ALGORITHM FOR OFDT INDUCTION

0	• qis the current level. Put q = 0. • Insert all input attributes into the set of unused attributes ua.
1	• Select an attribute from ua by $\operatorname{argmax}\left(\frac{\Delta \mathrm{I}\left(\mathrm{B}; \mathrm{A}_{\mathrm{i}_1}, \dots, \mathrm{A}_{\mathrm{i}_{q-1}}, \mathrm{A}_{\mathrm{i}_q}\right)}{\mathrm{Cost}\left(\mathrm{A}_{\mathrm{i}_q}\right)}\right)$
2	• Associate the chosen attribute with each node in level q and remove it from ua.
3	• Check every node in level q, if node is a leaf.
4	• If all nodes in level qare leaf or uais empty go to 5. • Else put q = q + 1 and recursively go to 1.
5	• Exit.

The threshold values affect on tree depth and size. Size of the tree presents number of nodes. The parameter β is the threshold value of the confidence level. If some node has bigger or equal confidence level as β, this node has to become a leaf. Increasing the value of beta causes an increase of the size of the tree. The parameter α represents the threshold of the minimal frequency of a given branch. If some node has lower or equal frequency as α, this node has to become the leaf. The parameter α also affects the size of the tree. In this case, bigger α causes the smaller tree size. The threshold values must be set to satisfied value to perform accurate classification [3], [9]. One possible way, how to estimate threshold values is run the algorithm more times with different threshold values and then the best combination is chosen. The process is described by diagram on Fig. 2.

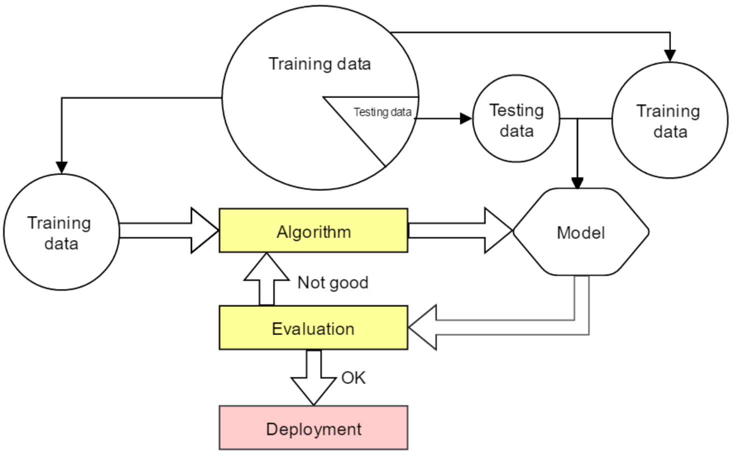

Fig. 2. Diagram of threshold values estimation

V. Analysis of fuzzy decision tree

The trees are analyzed by error and accuracy of the classification. During the error estimation, the tree is built from the whole training set and all instances in the training set are classified. The result of error estimation is the ratio between classification mistakes and the number of instances in the dataset, which represents the percentage of incorrectly classified instances. In case of accuracy estimation, the training dataset is divided into two dataset. The bigger dataset contains 80 % of the divided training dataset and this dataset is used to build decision tree. The second dataset contains 20 % of the divided dataset and samples from this dataset are used for classification. The result of accuracy of classification is percentage of correctly classified instances.

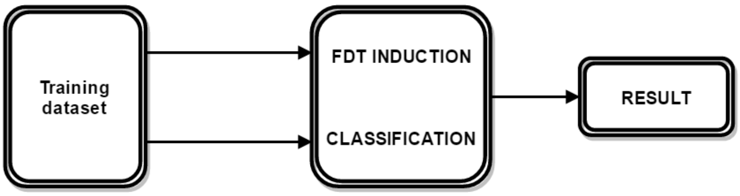

Fig. 3 Diagram of classification error estimation

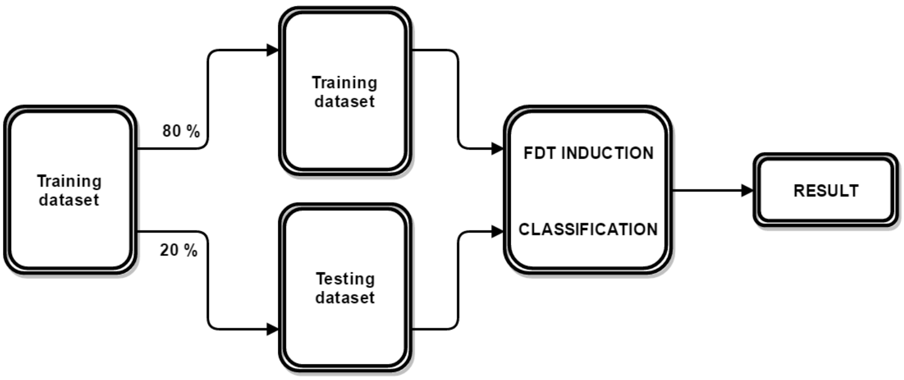

Fig. 4 Diagram of classification accuracy estimation

VI. USAGE OF ORDERED DECISION TREE

The usage of OFDT is demonstrated on the public dataset *Pima Indians Diabetes Database.* The dataset is available on [14]. Dataset contains medical records of diabetic patients. In particular, all patients here are females at least 21 years old of Pima Indian heritage. The population lives near Phoenix, Arizona, USA. The goal is investigate whether the patient shows signs of diabetes. The dataset has been gathered by National Institute of Diabetes and Digestive and Kidney Disease. The dataset contains 786 number of instance. All attributes are numerical. The count of input attributes is 8. The dataset contains also one numerical output attribute. The Table 2 contains attribute description and brief statistical analysis of attributes.

TABLE 4
ATTRIBUTE DESCRIPTION AND BRIEF STATISTICAL ANALYSIS OF ATTRIBUTES

Attribute	Mean	Standard Deviation
Number of times pregnant	3.8	3.4
Plasma glucose concentration a 2 hours in an oral glucose tolerance test	120.9	32.0
Diastolic blood pressure	69.1	19.4
Triceps skin fold thickness	20.5	16.0
2-Hour serum insulin	79.8	115.2
Body mass index	32.0	7.9
Diabetes pedigree function	0.5	0.3
Age (years)	33.2	11.8

The illustration of full OFDT is not suitable for this paper because the resulting tree consists of 39 nodes and its depth is 8. Hence, the threshold values have been chosen to induct a tree with suitable size and depth for visualization. Illustrated OFDT has following threshold values $\beta = 0.75$ and $\alpha = 0.25$. The classification error of this tree is 33.268 %. The accuracy of this example is 35.651 %. The tree is painted on Fig. 6. The optimal threshold values for classification error estimation are $\alpha = 0$ and $\beta = 1$. These values have been established by repeated runs of algorithm and the best combination has been chosen. In this case the tree is much robust, but classification error is only 0.036 %. The accuracy of classification is 0.947 %.

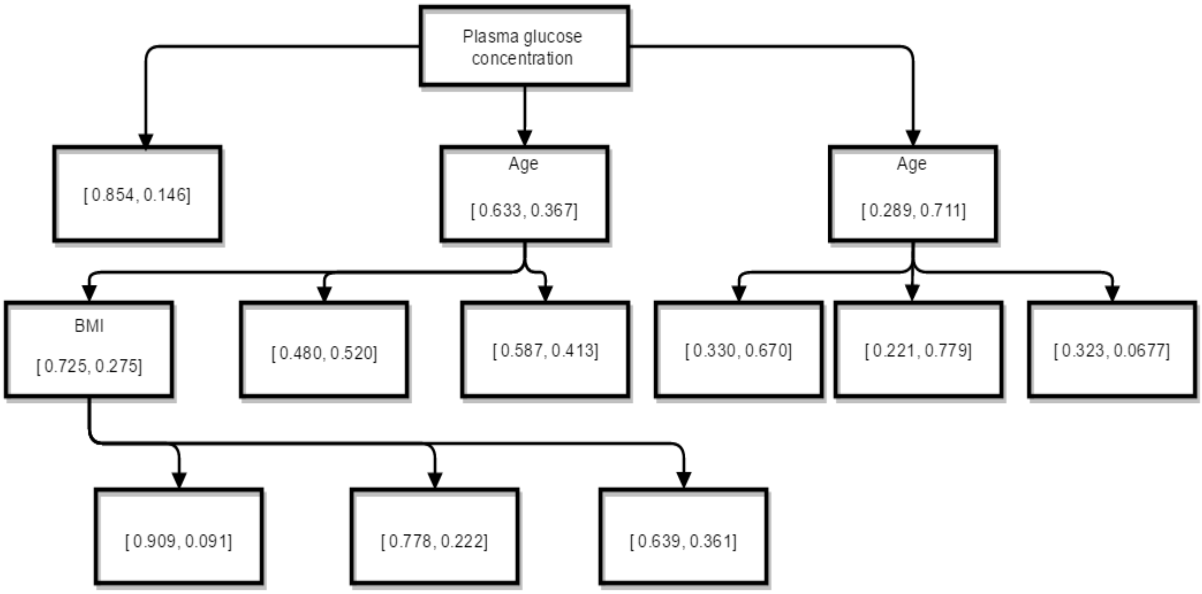

Fig. 5 Picture of small OFDT for Pima Indians Diabetes Database

VII. CONCLUSION

In this paper the classification method using approach of fuzzy decision trees is considered. In the paper the fuzzyification algorithm is also described. The presented algorithm OFDT is based on cumulative information estimations of initial data. The considered method is demonstrated on the public dataset: *Pima Indians Diabetes Database.* Algorithm is able to classify new instance with satisfied accuracy is 94.7 %. The classification error is 0.036 %.

REFERENCES

[1] J. I. Maletic and A. Marcus, *Data Mining and Knowledge Discovery Handbook*. 2005.

[2] V. Levashenko, E. Zaitseva, M. Kvassay, and Deserno, “Reliability Estimation of Healthcare Systems using Fuzzy Decision Trees,” *Ann. Comput. Sci. Inf. Syst.*, vol. 8, pp. 331–340, 2016.

[3] J. R. Quinlan, “Induction of Decision Trees,” *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.

[4] J.R. Quinlan, “Programs for machine learning. Morgan Kaufmann,” p. C4. 5, 1993.

[5] J. Rabcan, “Ordered Fuzzy Decision Trees Induction based on Cumulative Information Estimates and Its Application,” *ICETA*, p. 6, 2016.

[6] Y. Yuan and M. J. Shaw, “Induction of fuzzy decision trees,” *Fuzzy Sets Syst.*, vol. 69, no. 2, pp. 125–139, 1995.

[7] X. Wang, X. Liu, W. Pedrycz, and L. Zhang, “Fuzzy rule based decision trees,” *Pattern Recognit.*, vol. 48, no. 1, pp. 50–59, 2015.

[8] V. Levashenko, E. Zaitseva, and S. Puuronen, “Fuzzy Classifier Based on Fuzzy Decision Tree,” *EUROCON 2007 - Int. Conf. Comput. as a Tool*, pp. 823–827, 2007.

[9] V. Levashenko and E. Zaitseva, “Fuzzy Decision Trees in Medical Decision Making Support System,” *Computer Science and Information Systems*. ” in 2012 Federated Conference on Computer Science and Information Systems, pp. 213–219, 2012.

[10] L. Biacino and G. Gerla, “Fuzzy logic, continuity and effectiveness,” *Arch. Math. Log.*, vol. 41, no. 7, pp. 643–667, 2002.

[11] Y. Yao, “A comparative study of fuzzy sets and rough sets,” *Inf. Sci. (Ny).*, vol. 109, no. 1–4, pp. 227–242, 1998.

[12] H. Tanaka, L. T. Fan, F. S. Lai, and K. Toguchi, “Fault-tree analysis by fuzzy probabilities,” *Ieee*, vol. R-32, pp. 453–457, 1983.

[13] V. G. Levashenko and E. N. Zaitseva, “Usage of New Information Estimations for Induction of Fuzzy Decision Trees,” *Lect. Notes Comput. Sci.*, vol. 2412, pp. 493–499, 2002.

[14] “Pima Indians Diabetes Database.” [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes.

