SCIENTIFIC PAPERS
OF THE UNIVERSITY OF PARDUBICE
Series B
The Jan Perner Transport Faculty
11 (2005)

CLASSIFICATION BASED ON STABLE FUZZY DECISION TREE METHOD

Vitaly LEVASHENKO, Štefan KOVALÍK, Karol MATIAŠKO

Dept. of Informatics, Faculty of Management Science and Informatics, University of Žilina, Slovakia

1. Introduction

Data Mining, whose objective is to obtain useful knowledge from data stored in large repositories, is recognized as a basic necessity in many areas [7]. Since data represent a certain real world domain, patterns that hold in data show us interesting relations that can be used to improve our understanding of that domain. Classification is the basis step in the knowledge discovery process.

Current classification technologies, for example decision trees work well for pattern recognition and process control. A popular and efficient method for induction of decision trees is ID3 algorithm proposed J.R.Quinlan [10]. The main idea of this algorithm is selecting the attribute that takes the nominal values of the average mutual information and repeating this selection procedure. ID3 algorithms and its modification C4.5 make a crisp decision tree for classification. This tree consists of nodes for detecting attributes, edges for branching by values of symbols and leaves for deciding class names to be classified.

Unfortunately, while we are considering the problem of uncertainties and noise of real world, the data would be very difficult to be clearly classified. An experienced expert looking at an event in a real world environment can estimate quickly, and with a high degree of confidence the obtained information. The expert would define it as „highly possible”, „absolutely impossible” etc. On the other hands the exactly numeric data in many cases is against the nature of human beings (excluding statisticians, of course). People use their subjective feelings, background knowledge and short-time memory, rather than any probabilistic criteria, to distinguish different data [14].

The theory of fuzzy sets can certainly help us to construct decision trees, which more accuracy to reflect the real around environment [15]. Moreover, fuzzy sets are an optimal tool to model imprecise terms and relations as commonly employed by humans in communication and understanding. There are a lot of modern generalizing Fuzzy ID3 algorithms for fuzzy sets [1,8,9,11-14]. Algorithms [11,14] were based on deLuca & Termini entropies. Generalization Shannon's entropy was used in [1,12,13]. Analysis of well-known variants of Fuzzy ID3 algorithm can be found in [8,9].

In this paper, we are present our approach that can deal with fuzzy data. Also our approach should be able to analyze in which order input attributes detection should be performed in order to minimize the costs of classification and guarantee a desired predefined level of accuracy.

For these purposes, we use a technique to compute cumulative information estimations of fuzzy sets [6]. The application of such estimations allows inducing minimum cost decision trees based on new criterions of optimality. We had obtained new type of fuzzy decision trees (FDT): ordered FDT, in which to every node of one level associate with similar attribute (this allows to realize parallel search through different levels of tree) [3,5]. Now we are going to realize new type of FDT - stable ordered decision trees.

The paper is organized as follows. Section 2 contains brief information about representation fuzzy data. Section 3 short describes how cumulative information estimates are calculated, and then Section 4 shows how these estimations are used for FDT induction with a simple example. Section 5 illustrates usage FDT for classification new instance. And Section 6 demonstrates the basic results of our study on continuous valued benchmarks. Section 7 concludes this work

2. Fuzzy Set

Fuzzy logic is a popular approach to capture this vagueness of information [15]. The basic idea is to come from the "crisp" 1 and 0 values to a degree of truth or confidence in the interval [0,1].

Definition 1. A fuzzy set A with respect to a universe U is characterized by a *membership function* μ_A: U $\rightarrow$ [0,1], assign a A-membership degree, $\mu_A(u)$, to each element u in U. $\mu_A(u)$ gives us an estimation of the belonging of u to A.

For example, we divide the i-th attribute Ai that is real into *m=3* fuzzy partitions as it is depicted in Fig.1.

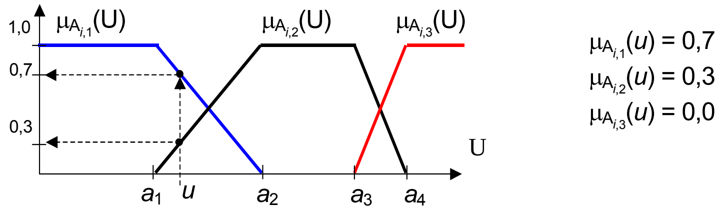

Figure 1. *Fuzzy membership functions of input attribute Ai*

Thus, the fuzzification of the initial data is performed by analyzing the corresponding values of a membership function. Here, each value of attribute can be seen as likelihood estimate. For these purposes, we use one of the algorithms to transform from numeric to triangular fuzzy data, which presented in [4].

A typical classification problem can be described as follows [14]. A universe of objects U={*u*} is described by N training examples and n input attributes **A**={A_1,...,A_n}. Each input attribute A_i (1≤i≤n) measures some important feature and is presented by a group of discrete *linguistic terms*. We assume that each group is a set of m_i (m_i ≥2) values of fuzzy subsets {$A_{i,1}$,…,$A_{i,j}$,…,$A_{i,mi}$}. The cost of the attribute A_i denoted as $Cost_i$ is an integrated measure that accounts financial and temporal resources. These resources are required to define the value of the A_i for a certain subject. We will suggest that each object u in the universe is classified by a set of classes {B_1,...,B_{mb}}. This set describes by class attribute B. Our goal is build model for new object classification. The model should be demand minimal resource cost.

Let us consider the following example.

Example 1. An object is presented with four input attributes: **A**={A_1,A_2,A_3,A_4} and one class attribute B. Each attribute has values: A_1={$A_{1,1}$, $A_{1,2}$, $A_{1,3}$}, A_2={$A_{2,1}$, $A_{2,2}$, $A_{2,3}$}, A_3={$A_{3,1}$, $A_{3,2}$}, A_4 = {$A_{4,1}$, $A_{4,2}$} and B = {B_1, B_2, B_3}.

Table 1. *A training set (adopted from [11,14])*

No	A_1			A_2			A_3		A_4		B		
	$A_{1,1}$	$A_{1,2}$	$A_{1,3}$	$A_{2,1}$	$A_{2,2}$	$A_{2,3}$	$A_{3,1}$	$A_{3,2}$	$A_{4,1}$	$A_{4,2}$	B_1	B_2	B_3
1.	0.9	0.1	0.0	1.0	0.0	0.0	0.8	0.2	0.4	0.6	0.0	0.8	0.2
2.	0.8	0.2	0.0	0.6	0.4	0.0	0.0	1.0	0.0	1.0	0.6	0.4	0.0
3.	0.0	0.7	0.3	0.8	0.2	0.0	0.1	0.9	0.2	0.8	0.3	0.6	0.1
4.	0.2	0.7	0.1	0.3	0.7	0.0	0.2	0.8	0.3	0.7	0.9	0.1	0.0
5.	0.0	0.1	0.9	0.7	0.3	0.0	0.5	0.5	0.5	0.5	0.0	0.0	1.0
6.	0.0	0.7	0.3	0.0	0.3	0.7	0.7	0.3	0.4	0.6	0.2	0.0	0.8
7.	0.0	0.3	0.7	0. 0	0.0	1.0	0.0	1.0	0.1	0.9	0.0	0.0	1.0
8.	0.0	1.0	0.0	0.0	0.2	0.8	0.2	0.8	0.0	1.0	0.7	0.0	0.3
9.	1.0	0.0	0.0	1.0	0.0	0.0	0.6	0.4	0.7	0.3	0.2	0.8	0.0
10.	0.9	0.1	0.0	0.0	0.3	0.7	0.0	1.0	0.9	0.1	0.0	0.3	0.7
11.	0.7	0.3	0.0	1.0	0.0	0.0	1.0	0.0	0.2	0.8	0.3	0.7	0.0
12.	0.2	0.6	0.2	0.0	1.0	0.0	0.3	0.7	0.3	0.7	0.7	0.2	0.1
13.	0.9	0.1	0.0	0.2	0.8	0.0	0.1	0.9	1.0	0.0	0.0	0.0	1.0
14.	0.0	0.9	0.1	0.0	0.9	0.1	0.1	0.9	0.7	0.3	0.0	0.0	1.0
15.	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.8	0.2	0.0	0.0	1.0
16.	1.0	0.0	0.0	0.5	0.5	0.0	0.0	1.0	0.0	1.0	0.5	0.5	0.0
Cost$_i$	2,5			1,7			2,0		1,8				

In this paper we use an approach for solve classification task based on stable FDT method.

3. Mutual Information Estimations

We have proposed *cumulative* information estimates, which describe it in paper [6]. Let us to make short introduction in these estimates. We have a sequence of q-1 input attributes $\boldsymbol{U}_{q\text{-}1} = \{A_{i_1},\ldots,A_{i_{q-1}}\}$ ($q \ge 2$) and class attribute B. These input attributes $\boldsymbol{U}_{q\text{-}1}$ have next values $U_{q\text{-}1} = \{A_{i_1,j_1},\ldots,A_{i_{q-1},j_{q-1}}\}$ accordingly.

Definition 2. The *cumulative joint* information into values B_j (j = 1,…, m_b) and $U_{q\text{-}1}$ is

$$\mathbf{I}(B_j, U_{q\text{-}1}) = -\log_2 M(B_j \times A_{i_1,j_1} \times \ldots \times A_{i_{q-1},j_{q-1}}) \text{ bits,} \qquad (1)$$

where M(A) is a *cardinality measure* of fuzzy set A: $M(A)=\Sigma_{u\in U}\, \mu_A(u)$ [14].

Definition 3. The *cumulative conditional* entropy between class attribute B and input attribute A_{i_q} (or its value A_{i_q,j_q}) with given $U_{q\text{-}1}$ is uncertainty into attribute B when attribute A_{i_q} (or its value A_{i_q,j_q}) and the sequence $U_{q\text{-}1}$ are known

$$\mathbf{H}(B|\, U_{q\text{-}1}, A_{i_q}) = \sum_{jq=1}^{m_{iq}} \mathbf{H}(B \,|\, U_q) = \sum_{jq=1}^{m_{iq}} \mathbf{H}(B \,|\, U_{q\text{-}1}, A_{i_q j_q}) \text{ bit,} \qquad (2)$$

$$\mathbf{H}(B|\, U_{q\text{-}1}, A_{i_q j_q}) = \sum_{j=1}^{m_b} M(B_j \times U_q) \times [\mathbf{I}(B_j, U_q) - \mathbf{I}(U_q)] \text{ bit,}$$

where $\mathbf{I}(B_j, U_q)$ *cumulative joint* information into values B_j and U_q is calculated by (1).

Definition 4. The *cumulative mutual* information into attribute A_{i_q} and the sequence of values $U_{q\text{-}1}=\{A_{i_1,j_1},\ldots,A_{i_{q-1},j_{q-1}}\}$ about class attribute B reflects the influence of attribute A_{i_q} on the attribute B when sequence $U_{q\text{-}1}$ of attributes is known

$$\mathbf{I}(B;\, U_{q\text{-}1}, A_{i_q}) = \mathbf{H}(B \,|\, U_{q\text{-}1}) - \mathbf{H}(B \,|\, U_{q\text{-}1}, A_{i_q}) \text{ bit,} \qquad (3)$$

where $\mathbf{H}(B|U_{q\text{-}1})$ and $\mathbf{H}(B|U_{q\text{-}1},A_{i_q})$ are cumulative conditional entropies (2).

Example 2. For training set from Table 1 and $U_1=\{A_{2,1}\}$ we obtain

$\mathbf{I}(B_1) = \log_2 N - \log_2 M(B_1) = \log_2 16 - \log_2 4{,}52 = 1{,}862$ bit;

$\mathbf{I}(B_1, A_{2,1}) = \log_2 N - \log_2 M(B_1 \times A_{2,1}) = \log_2 16 - \log_2 1{,}62 = 3{,}304$ bit;

$\mathbf{H}(B) = \sum_{j=1}^{3} M(B_j) \times \mathbf{I}(B_j) = 4{,}4\times1{,}862 + 4{,}4\times1{,}862 + 7{,}2\times1{,}152 = 24{,}684$ bit;

$\mathbf{H}(B|\, A_{2,1}) = \sum_{j=1}^{3} M(B_j \times A_{2,1}) \times \mathbf{I}(B_j, A_{2,1}) = 1{,}62\times1{,}913 + 2{,}925 + 2{,}797 = 8{,}820$ bit;

$\mathbf{H}(B|\, A_2)=\mathbf{H}(B|\, A_{2,1})+\mathbf{H}(B|\, A_{2,2})+\mathbf{H}(B|\, A_{2,3}) = 8{,}820+8{,}201+3{,}911=20{,}932$ bit;

$\mathbf{I}(B;\, A_2)= \mathbf{H}(B) - \mathbf{H}(B|\, A_2)= 24{,}684 - 20{,}932=3{,}752$ bit;

$\mathbf{H}(B|\, A_{2,1},A_{1,2})= \sum_{j=1}^{3} M(B_j \times A_{2,1} \times A_{1,2}) \times [\mathbf{I}(B_j,A_{2,1}, A_{1,2}) - \mathbf{I}(A_{2,1}, A_{1,2})] = 1{,}935+1{,}564+1{,}227 = 4{,}725$ bit;

$\mathbf{H}(B|\, A_{2,1}, A_1) = \mathbf{H}(B|\, A_{2,1},A_{1,1})+\mathbf{H}(B|\, A_{2,1},A_{1,2})+ \mathbf{H}(B|\, A_{2,1},A_{1,3}) = 4{,}725+1{,}927+1{,}000 = 7{,}652$ bit;

$\mathbf{I}(B; A_{2,1},A_1) = \mathbf{H}(B| A_{2,1}) - \mathbf{H}(B| A_{2,1},A_1) = 8{,}820 – 7{,}652 = 1{,}168$ bit.

These information estimations we have used for forming new criteria of FDT induction. In [5] we used it for building ordered FDT. Stable ordered FDT is a new type of FDT. We introduce it in next Section 4.

4. Stable Ordered FDT Induction

We had proposed new interpretation of Fuzzy ID3, which is based on cumulative information estimates [5,6].

Apart from the selection of expanded attributes, the determination of the leaf node is another important issue for FDT induction. The key points of a proposed algorithm for induction FDT are (a) a heuristic for selecting expanded attributes and (b) a rules for transform nodes into leaves. Expanded attributes are such attributes that according to values of attributes trees are expanded at the nodes considered.

4.1 Selection expanded attributes

The induction of an ordered FDT has less complexity, while it does not require information estimates calculations for each branch of a tree. When choosing an expanded attribute q it is enough to maximize the increment of information about the attribute with minimum of costs [5]:

The induction of an ordered FDT has less complexity, while it does not require information estimates calculations for each branch of a tree. When choosing an expanded attribute *q* it is enough to maximize the increment of information about the attribute with minimum of costs [5]:

$$q = \text{argmax} \;\; \Delta\mathbf{I}(B_q; \boldsymbol{U}_{q\text{-}1}, A_{i_q}) \,/\, Cost_{i_q} \quad . \qquad (4)$$

where $\boldsymbol{U}_{q\text{-}1} = \{A_{i_1}, ..., A_{i_{q-1}}\}$ and $\mathbf{I}(B_q; \boldsymbol{U}_{q\text{-}1}, A_{i_q})$ is calculated by (3).

Let's us to detail analyze of component $\mathbf{I}(B_q; \boldsymbol{U}_{q\text{-}1},A_{i_q})$ from rule (4) by (3).

$$\Delta\mathbf{I} (B; \boldsymbol{U}_{q\text{-}1}, A_{i_q}) = \mathbf{I}(B; \boldsymbol{U}_{q\text{-}1}, A_{i_q}) - \mathbf{I}(B; \boldsymbol{U}_{q\text{-}1}) = (\mathbf{H}(\boldsymbol{U}_{q\text{-}1},A_{i_q}) - \mathbf{H}(B, \boldsymbol{U}_{q\text{-}1},A_{i_q})) - (\mathbf{H}(\boldsymbol{U}_{q\text{-}1}) - \mathbf{I}(B, \boldsymbol{U}_{q\text{-}1}))$$
$$= (\mathbf{H}(\boldsymbol{U}_{q\text{-}1},A_{i_q}) - \mathbf{H}(\boldsymbol{U}_{q\text{-}1})) - (\mathbf{H}(B, \boldsymbol{U}_{q\text{-}1},A_{i_q}) - \mathbf{H}(B, \boldsymbol{U}_{q\text{-}1})) = \mathbf{H}(A_{i_q}| \boldsymbol{U}_{q\text{-}1}) - \mathbf{H}(A_{i_q}| B, \boldsymbol{U}_{q\text{-}1}). \quad (5)$$

In the rule (5):

- *conditional cumulative entropy* $\mathbf{H}(A_{i_q}| \boldsymbol{U}_{q\text{-}1})$ describes amount of *new information* from attribute A_{iq}, which we don't know in advance (when we know attributes values A_{i_1}, ..., $A_{i_{q-1}}$).
- *conditional cumulative entropy* $\mathbf{H}(A_{i_q}| B, \boldsymbol{U}_{q\text{-}1})$ describes amount of *un-useful information* from attribute A_{i_q}. This information separates situations, which belong to identical class (identical decision).

We had replaced one element $\mathbf{H}(A_{i_q}|\ \boldsymbol{U}_{q\text{-}1})$ from criterion (5) by new element $\mathbf{H}(A_{i_q})$. This element describes amount of *information received* from attribute A_{i_q}:

$\mathbf{H}(A_{i_q}) - \mathbf{H}(A_{i_q}|B,\ \boldsymbol{U}_{q\text{-}1}) = \mathbf{I}(A_{i_q};B,\ \boldsymbol{U}_{q\text{-}1})$. Therefore, our new criterion for stable FDT induction can be represent by rule (6):

$$q = argmax\ \mathbf{I}(A_{i_q}; B, \boldsymbol{U}_{q\text{-}1},) / Cost_{i_q} \qquad (6)$$

4.2 Transformation nodes into leaves

We have two kind of threshold β and α. This truth level threshold controls the growth of the tree. Nodes are usually regarded as leaves if:

- the frequency f of branch U_q is less than or equal to a given threshold value α :

$$\mathbf{I}(U_q) \geq -\log \alpha \times N \quad \text{or} \quad f = 2^{-\mathbf{I}(U_q)/N} \leq \alpha\ ; \qquad (7a)$$

- the relative frequency of one class is greater or equal than a given threshold value β:

$$min\ \mathbf{I}(B_j|\ U_q) \leq -\log \beta \quad \text{for } \forall\ j=1,\ldots m_b; \qquad (7b)$$

Lower β and greatest α (in (7a and 7b)) may lead to compact tree.
Example 3. Stable FDT consummated from data that presented in Table 1 can be seen in Fig.2.

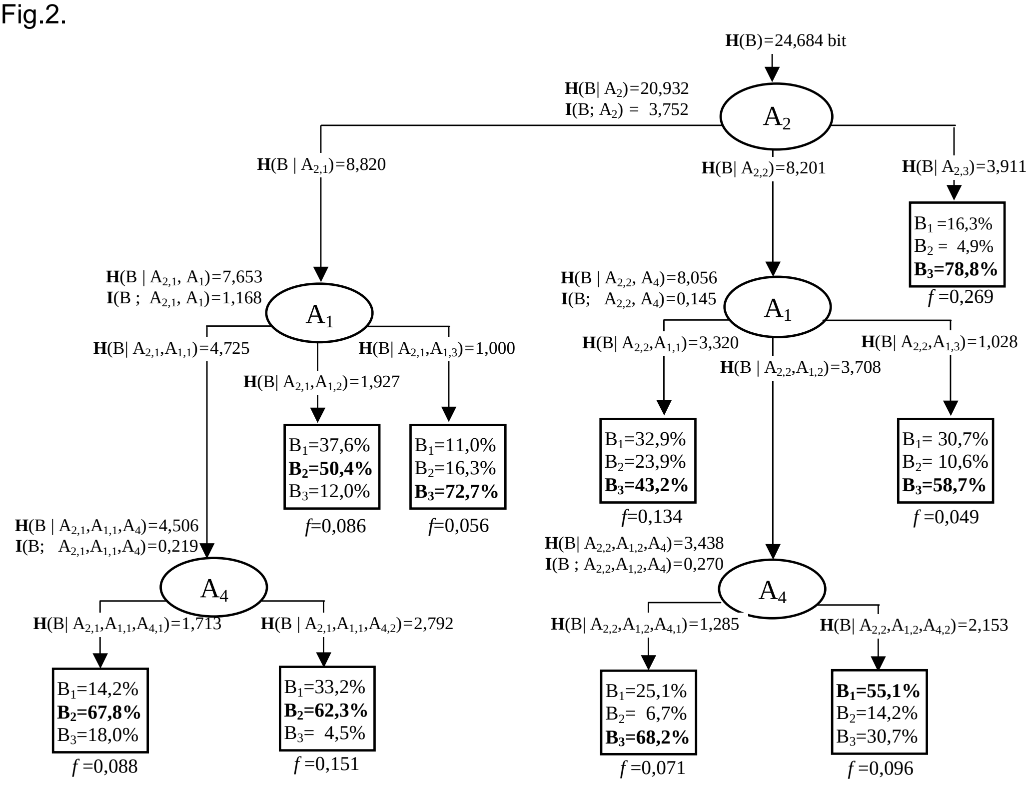

Figure 2. *Stable FDT (β=0,75 and α=0,16) realized by rule (6).*

Induction of a stable FDT allows reducing classification time due to the possibility of parallel checks of attribute values at several levels of the FDT.

Actually, when classifying an instance at level s, using the unordered FDT (s=2,…) we need to define Ais attribute's value and it is impossible to know in advance the value of which attribute we need to evaluate at the next level (s+1). On the contrary, when using an ordered decision tree the same attribute is associated to every level of the FDT. That is why we know the sequence of attributes to be evaluated for each branch in advance. Obviously, the construction of an ordered FDT implies additional costs, but its use can be beneficial in the situations when time factor is critical and there is a possibility to implement the check of several attributes simultaneously.

5. Usage FDT for classification

In this section, we explain the mechanism of making fuzzy rules from FDTs and their use for classification.

In the FDT, each non-leaf node is associated with an attribute from **A**. When A_i is associated with a non-leaf node, the node has m_i outgoing branches. The j-th branch of the node is associated with value $A_{i,j}$. The class attribute B has m_b possible values B_1,..., B_{j_b},..., B_{m_b}. Let the FDT have R leaves $\mathbf{L} = \{l_1, ..., l_r, ..., l_R\}$. It is also considered there is a value $\mathbf{F}^r = [F_1^r;..., F_{j_b}^r;..., F_{m_b}^r]$ for each r-th leaf l_r and each j_b-th class B_{j_b}. This value $F_{j_b}^r$ means the certainty degree of the class B_{j_b} attached to the leaf node l_r.

In fuzzy cases, a new instance e may be classified into different classes with different degrees. Then, each leaf $l_r \in \mathbf{L}$ corresponds to one r-th classification rule. The condition of the classification rule is a group of "*attribute* is *attribute's value*" which are connected with *and*-operator. These attributes are associated with the nodes in the path from the root to the leaf l_r. The attribute's values are the values associated with the respective outgoing branches of the nodes in the path. The conclusions of the r-th rule are values of class attribute B. Let's consider that in the path $P_r(e)=\{[A_{i_1,j_1}(e)]^r,…,[A_{i_s,j_s}(e)]^r,…,[A_{i_S,j_S}(e)]^r\}$ from the root to the r-th leaf. This path P_r consist of S nodes which associated with attributes A_{i_1},.., A_{i_s},…, A_{i_S} and respectively their S outgoing branches associated with the values A_{i_1,j_1},.., A_{i_s,j_s},… , A_{i_S,j_S}. Then the r-th rule has the following form:

IF (A_{i_1} is A_{i_1,j_1}) *and* … *and* (A_{i_s} is A_{i_s,j_s}) *and* … *and* (A_{i_S} is A_{i_S,j_S}) THEN B (with truthfulness $\mathbf{F}^r$).

Our approach uses several classification rules for classification of new instance *e.* That's why, there may be several paths whose all outgoing node's branches are associated with value $A_{i_s,j_s}(e)$ more than 0. Each path $P_r(e)$ brings about leaf node l_r and corresponds to one r-th classification rule. In this case each r-th classification rule should be included in the final classification with a certain weight W_r. The weight is for instance e and the r-th rule given by the rule $W_r(e)=\prod_{s=1}^{S}[A_{i_s,j_s}(e)]^r$, where $[A_{i_s,j_s}(e)]^r$ is the value of the attribute A_{i_s}

for new instance *e*. The weight W_r is equal 0 if there is a attribute's value A_{i_s,j_s} whose membership function's equals 0. A values of class attribute B for the new instance *e* are:

$\mu_B(e)=\sum_{r=1}^{R} \mathbf{F}^r \times W_r(e)$, where $\mathbf{F}^r$ is the truthfulness of the *r*-th rule.

Let us describe the process of transformation of the FDT into fuzzy rules and the their use for classification with the following example.

Example 4. The FDT has R=9 leaves (see in Fig.2.). The new instance *e* is described next conditions: $A_1=\{A_{1,1}; A_{1,2}; A_{1,3}\}=\{0.9; 0.1; 0.0\}$, $A_2=\{A_{2,1}; A_{2,2}; A_{2,3}\}=\{1.0; 0.0; 0.0\}$, $A_3=\{A_{3,1}; A_{3,2}\}=\{0.8; 0.2\}$ and $A_4=\{A_{4,1}; A_{4,2}\}=\{0.4; 0.6\}$. Our goal is to determine values of class attribute B for this new instance *e*.

Let's form 9 classification rules for each FDT leaves.

r=1: IF A_2 is $A_{2,1}$ *and* A_1 is $A_{1,1}$ *and* A_4 is $A_{4,1}$ THEN B truthfulness $\mathbf{F}^1$ = [0.142; 0.678; 0.180];
r=2: IF A_2 is $A_{2,1}$ *and* A_1 is $A_{1,1}$ *and* A_4 is $A_{4,2}$ THEN B truthfulness $\mathbf{F}^2$ = [0.332; 0.683; 0.045];
r=3: IF A_2 is $A_{2,1}$ *and* A_1 is $A_{1,2}$ THEN B truthfulness $\mathbf{F}^3$ = [0.376; 0.504; 0.120];
…
r=9: IF A_2 is $A_{2,3}$ THEN B truthfulness $\mathbf{F}^9$ = [0.163; 0.049; 0.788].

Calculate weight $W_r(e)$ (r = 1,…, 9). $W_1(e) = 0.9\times1.0\times0.4 = 0.36$. Similarly, $W_2(e) = 0.9\times1.0\times0.6 = 0.54$, $W_3(e) = 0.1\times1.0 = 0.1$. All the other $W_r(e)$ are equal 0.
Then, $\mu_{B_1}(e) = 0.142\times0.36 + 0.332\times0.54 + 0.376\times0.1 + 0.110\times0 + \ldots + 0.163\times0 = 0.268$. Similarly,
$\mu_{B_2}(e) = 0.678\times0.36 + 0.623\times0.54 + 0.504\times0.1 + 0.163\times0 + \ldots + 0.049\times0 = 0.631$ and
$\mu_{B_3}(e) = 0.180\times0.36 + 0.045\times0.54 + 0.120\times0.1 + 0.727\times0 + \ldots + 0.787\times0 = 0.101$.
The values of class attribute $B = \{B_1, B_2, B_3\} = \{0.268; 0.631; 0.101\}$ for the instance *e*. The maximum value has $\mu_{B_2}(e)$. And so, if classification only into one class is needed, instance *e* is classified into class B_2.

6. Experimental Results

The algorithm is coded in C++ and the experimental results are obtained on a Pentium III with 256Mb of memory. The main purpose of our experimental study is to compare our algorithm with another classification methods.

The experiments have been carried out on Machine Learning benchmarks (dataset) each of which has at least one continuous variable [2]. We had separated initial dataset into 2 parts. We used the first part (70% from initial dataset) for building classification models. The second part (another 30%) was used for verification previous models. This process of separation and verification was repeated 100 times to obtain the model's error rate for respective databases. The error rate is calculated as the ratio of the number of misclassification combinations to the total number of combinations.

The results of our experiments are in Tab.2. Columns [Total sets], [Number of attributes] and [Number of classes] describes dataset. The column labelled [Errors] gives the count of error classification. It is calculated as the ratio of the number of misclassification combinations to the total number of combinations. Label [Naïve] denotes Naïve Bayes Classifier [kNN] denotes k-Nearest Neighbour Classifier and [sFDT] denotes stable FDT Classifier. The numbers in brackets denote the order of the method for respective database; number 1 denotes the best case with minimal error rate. The last row contains average error rate for all databases.

Table *2.* *Results on the UCI machine learning benchmark set*

Dataset	Total sets	Number of attributes Input / Numeric	Number of classes	Errors		
				Naïve	kNN	sFDT
bupa	345	6 / 6	2	$0.4414^{(3)}$	$0.3832^{(1)}$	$0.4061^{(2)}$
cmc	1473	9 / 2	3	$0.5240^{(2)}$	$0.5816^{(3)}$	$0.5118^{(1)}$
glass	214	9 / 9	7	$0.5347^{(3)}$	$0.3152^{(1)}$	$0.3914^{(2)}$
haberman	306	3 / 3	2	$0.2595^{(1)}$	$0.3389^{(3)}$	$0.2722^{(2)}$
iris	150	4 / 4	3	$0.0449^{(2)}$	$0.0473^{(3)}$	$0.0322^{(1)}$
pima	768	8 / 8	2	0.2491	$0.2971^{(3)}$	$0.2597^{(2)}$
ecoli	336	7 / 5	8	$0.1675^{(2)}$	$0.2097^{(3)}$	$0.1603^{(1)}$
				$0.3173^{(3)}$	$0.3104^{(2)}$	$0.2905^{(1)}$

7. Conclusion

Induction of FDT is a useful technique to find patterns in data in the presence of imprecision, either because data are fuzzy in nature or because we must improve their semantics. We have proposed the technique to induction of new type of FDT: stable FDT. The use of our cumulative information estimations allows precisely estimating mutual influence of attributes. These evaluations are good tool for analysis of group of training examples.

We have shown application of our information estimations in "greedy" algorithm for stable FDT induction. We suppose, that these evaluations will be a basis of algorithm for induction fully optimal FDT. The approach outlined in this paper is a basis for our further investigations.

The use of such estimates allows inducing minimum cost of decision process based on different criteria of optimality. We introduced also the cost of diagnostics into classification algorithms.

Acknowledgements

This research was supported by the grants of Scientific Grant Agency of the Ministry of Education of Slovak Republic and the Slovak Academy of Science (grant ZU/05 VV_MVTS13). We would like to say thank Dr. Andrey Sedun, PhD and his colleagues from Belarussian State Economic University for constructive comments, which have been made on 3rd Int. Seminar “Decision Making Support Systems” (BSEU, Minsk, June 2006).

Lektoroval: prof. PhDr. RNDr. Zdeněk Půlpán, CSc.

Předloženo: 30.06.2006

References

1. Chiang I-Jen, Hsu J.Yung-jen, *Fuzzy Classification Trees for Data Analysis,* International Journal of Fuzzy Sets and Systems, 130 (2002) 87-99.
2. Hettich S., Blake C.L., Merz C.J.: *UCI Repository of Machine Learning Databases* [http://www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, CA, University of California, Dept. of Information and Computer Science (1998).
3. Kollár J.: *Paralelné programovanie*. Academic Press elfa, s.r.o., Košice, 1999, ISBN 80-88964-14-8, 96p.
4. Lee H.M., Chen C.M., Chen J.M., Jou Y.L.: *An Efficient Fuzzy Classifier with Feature Selection Based on Fuzzy Entropy*. Journal of IEEE Trans. on Systems, Man and Cybernetics, Part B:Cybernetics, vol.31, 3 (2001) 426-432.
5. Levashenko V., Martincová P.: *Fuzzy Decision Tree for Parallel Processing Support*. Journal of Information, Control and Management Systems vol.3, 1 (2005) 45-52.
6. Levashenko V., Zaitseva E.: *Usage of New Information Estimations for Induction of Fuzzy Decision Trees*. In: Proc of the 3rd IEEE Int. Conf. on Intelligent Data Engineering and Automated Learning (IEEE IDEAL’02), Kluwer Publisher (2002), 493-499.
7. Matiaško K.*: Database Systems*. EDIS Print, Zilina (2002) 303 p.
8. Mitra S., Konwar K.M., Pal S.K.: *Fuzzy Decision Tree, Linguistic Rules and Fuzzy Knowledge-Based Network: Generation and Evaluation*, Journal of IEEE Trans. on Syst., Man Cybernetics — Part C: Applications and Reviews, Vol. 32, 4 (2002) 328-339.
9. Olaru C.,Whenkel L.: *A Complete Fuzzy Decision Tree Technique*, International Journal of Fuzzy Sets and Systems 138 (2003) 221-254.
10. Quinlan J.R.: *Induction of Decision Trees*. International Journal of Machine Learning 1 (1986) 81-106.
11. Wang X., Chen B., Qian G., Ye F.: *On the Optimization of Fuzzy Decision Trees*. International Journal of Fuzzy Sets and Systems 112 (2000) 117-125.
12. Weber R.: *Fuzzy-ID3: a Class of Methods for Automatic Knowledge Acquisition*. In Proc. of the 2nd Int. Conf. on Fuzzy Logic and Neural Networks, Japan, (1992), 265-268.
13. Yeung D.S., Tsang E.C.C., Wang X.: *Fuzzy Rule Wining by Fuzzy Decision Tree Induction Based on Fuzzy Feature Subset*. In: Proc. of the IEEE Int. Conf. on Syst., Man Cybernet., Vol. 4 (2002), 599-604.
14. Yuan Y., Shaw M.J.*: Induction of Fuzzy Decision Trees*. International Journal of Fuzzy Sets and Systems 69 (1995) 125-139.
15. Zadeh L.: Fuzzy sets, Journal of Information and Control, 8 (1965) 407–428.

Resumé

KLASIFIKÁCIA ZALOŽENÁ NA METÓDE STABILNÝCH ROZHODOVACÍCH STROMOV

Vitaly LEVASHENKO, Štefan KOVALÍK, Karol MATIAŠKO

Tvorba rozhodovacích stromov je jedným z najvhodnejších prístupov, požívaných pre extrakciu znalostí z databáz. Značná časť týchto databáz predstavuje formálnu analýzu a modelovanie ľudských aktivít, ktoré majú fuzzy charakter. V reálnom svete sa vyskytujú úlohy, ktoré človek dokáže spracovať ľahko, ale je tažké ich spracovať pomocou techniky – počítača. Tento problém umožňuje popísať fuzzy logika. Fuzzy rozhodovacie stromy sú populárnou metódou pre fuzzy klasifikáciu. Na základe teórie informácie zavádzame pojem kumulatívneho informačného odhadu. Tento pojem používame pre vytváranie rôznych kriterií pre tvorbu rozhodovacích stromov. Použitie týchto kriterií nám umožňuje vytváranie nových typov stromov. V tomto článku zavádzame pojem stabilného usporiadaného fuzzy rozhodovacieho stromu (Stable Ordered Fuzzy Decision Tree - FDT). Tento strom je orientovaný na paralelné a stabilné spracovanie vstupných atribútov s rozličnou cenou. Použitie FDT umožňuje realizovať sub-optimálnu klasifikáciu. Takáto klasifikácia určuje sekvenciu vstupných atribútov s minimálnou cenou ich vyhodnotenia. Uvádzame tiež proces transformácie z FDT na množinu fuzzy pravidiel. Výsledky tohoto článku môžu byť využité pri návrhu fuzzy systémov pre podporu rozhodovania alebo expertných systémov, založených na množine fuzzy pravidiel v tvare „ak x = A a y = B potom z = C“.

Kľúčové slová: klasifikácia, kumulatívne informačné odhady, stabilné fuzzy rozhodovacie stromy

Summary

CLASSIFICATION BASED ON STABLE FUZZY DECISION TREE METHOD

Vitaly LEVASHENKO, Štefan KOVALÍK, Karol MATIAŠKO

Decision tree induction is one of useful approaches for extracting classification knowledge from set instances. Considerable part of these instances obtains from formal analysis and modeling of human activities, which has fuzzy nature. It is often the case that real-world tasks can be handled easily by humans, they are often too difficult to be handled by machines. Fuzzy logic allows us to describe this problem. Fuzzy decision tree is a very popular method for fuzzy classification. We introduced term of cumulative information estimations based on Theory of Information approach. We used these cumulative estimations for synthesis of different criteria of decision tree induction. Usage these criteria allow us to produce new type of trees. In this paper we introduce Stable Ordered Fuzzy Decision Tree (FDT). The tree is oriented to parallel and stable processing of input attributes with differing cost. Usage this FDT allows us to realize a sub-optimal classification. Such classification detect a sequence of checks of input attributes with minimize the check-up cost. Also we introduce transformation process from FDT to fuzzy rules set. The results of this paper may be used for design of fuzzy decision-making or expert systems, which based on fuzzy rules set “if x is A and y is B then z is C”

Key words: classification, cumulative information estimations, stable ordered fuzzy decision trees.

Zusammenfassung

KLASSIFIKATION AUSGEFÜHRT AUS DER METHODE VON STABILEN ENTSCHEIDUNGBÄUMEN

Vitaly LEVASHENKO, Štefan KOVALÍK, Karol MATIAŠKO

Die Aufbau den Entscheidungsbäumen ist eine von besten Data-Minning Verfahren, die für die Gewinnung des Wissens aus Instanzen genutzt werden. Ein grosser Teil von diesen Instanzen darstellt die formale Analyse und die Modellierung von Aktivitäten der Menschen, welche die Fuzzy-Character haben. Menschen können leicht sehr viele Aufgaben erledigen, welche oft nur sehr schwierig Automaten erledigen können. Die Fuzzy-Logic ermöglicht uns diese Probleme beschreiben. Fuzzy - Entscheidungsbaum ist eine sehr populäre Methode für die Fuzzy-Klassifizierung. Wir einführen die kumulative information Abschätzung, die aus der Informationsteorie ausgeführt ist. Wir verwenden diese kumulative Abschätzungen für die Synthese von verschiedenen Kriterien der Induktion von Entscheidungsbämen. Mit Hilfe von der Verfügung diesen Kriteria wir können neue Variante Bäumen entwickeln. In diesem Referat wir wollen stabile geordnete Fuzzy - Entscheidungsbäme (Stable Ordere Fuzzy Decision Tree -FDT) vorstellen. Diese Bäume sind nach der parallelen und stabilen Bearbeitung von Eingangsattributen mit der minimalen Bewertung orientiert.

Mit Hilfe von FDT wir können suboptimale Klassifikation machen. Diese Klassifikation bestimmt die Folge die Eingangsattributen mit der minimalen Kosten der Bewertung.

Wir beschreiben auch die transformation aus FDT nach die Gruppe Fuzzyregeln. Das Ergebnis von diesen Referat kann beim Aufbau von Fuzzy - Entscheidungsfindung systemen oder von Expertensystemen benutzen sein.

Key words: Klassifizierung, kumulative Informationsabschätzung, stabile geordnete Fuzzy - Entscheidungsbaum.