Summary of making fuzzy rules from decision trees based on cumulative information and classification ambiguity

Jan Bohacik[*], Katedra informatiky FRI ŽU in Žilina, Slovakia

Karol Matiaško, Katedra informatiky FRI ŽU in Žilina, Slovakia

Rosaldo Rossetti, Departamento de Engenharia Informatica FEUP in Porto, Portugal

Abstract: Two methods of making fuzzy rules on the basis of decision trees are summarized in a united terminology using notions of fuzzy logic. A comparison of these methods is given. These methods can be used in the Data Mining stage of the Knowledge Discovery in Databases and their outcomes as knowledge base for expert systems in transportation.

Key words: fuzzy rules, decision tree, classification tree, cumulative information, classification ambiguity, notions of fuzzy logic

1. Introduction

Classification is often solved in the Data Mining stage of the Knowledge Discovery in Databases. Given a set of training instances (known instances) where each instance is described by attributes and classified into a class, the task is to build a model that predicts the class of an unseen instance. This model can be used in systems where a decision is required on the basis of known data. Systems of this kind are described more in (9), (8). An example of using in transportation is a prediction of accidents on the basis of risk factors (3). One of the most popular classification models is the classification rules model. Classification rules are popular as they are in a form which is natural for human beings.

In traditional classification rules classification, an attribute is either categorical or numerical. Numerical attributes need to be discretized. During discretization, there is some loss of information. This loss is caused by making precise boundaries. It can be lowered when notions of fuzzy logic, i.e. fuzzy sets, possibility distribution and linguistic variables, are used. When notions of fuzzy logic are used in distcretization, it is called fuzzification. Fuzzification leads to changes in traditional algorithms in the Knowledge Discovery in

* Ján Boháčik with all Slovak diacritics.

Databases. In the case of classification rules, fuzzy classification rules are used. Fuzzy classification rules are called fuzzy rules hereafter.

The focus of this paper is fuzzy rules made on the basis of decision trees. Two fuzzy rule methods are described in a united terminology using notions of fuzzy logic. One is based on work (10) (called decision tree based on classification ambiguity in this paper), the other is based on work (7) (called decision tree based on cumulative information in this paper). The former uses minimization of classification ambiguity as the criterion for choosing a linguistic variable for a node, the latter uses maximization of cumulative information.

In the following definitions, there are some basic definitions from fuzzy mathematics.

Definition 1:

Let some universe $\boldsymbol{U}$ be given, a *fuzzy subset* $\boldsymbol{M}$ of the set $\boldsymbol{U}$ is a map:

$$\boldsymbol{M}\colon \boldsymbol{U} \rightarrow [0,1],$$

where the value of $\boldsymbol{M}(\boldsymbol{x})$ for each $\boldsymbol{x} \in \boldsymbol{U}$ is interpreted as the degree to which the $\boldsymbol{x}$ is an element of fuzzy subset $\boldsymbol{M}$ (i.e. *membership degree*), or equally, as the truthfulness of the statement "$\boldsymbol{x}$ is an element of fuzzy subset $\boldsymbol{M}$".

Definition 2:

Let $\boldsymbol{M}$ be a fuzzy set defined on the universe $\boldsymbol{U}$. *Fuzzy set* $\boldsymbol{M}$ *at significance level* α (marked $\boldsymbol{M}^{\alpha}$) is defined as follows:

$$\boldsymbol{M}^{\alpha}(\boldsymbol{x}) = \begin{cases} \boldsymbol{M}(\boldsymbol{x}) & ;\text{ if } \boldsymbol{M}(\boldsymbol{x}) \geq \alpha \\ 0 & ;\text{ if } \boldsymbol{M}(\boldsymbol{x}) < \alpha \end{cases}.$$

Definition 3:

Cardinality of fuzzy set $\boldsymbol{M}$ defined on the universe $\boldsymbol{U}$ is specified as follows:

$$\#(\boldsymbol{M}) = \sum_{\boldsymbol{x} \in \boldsymbol{U}} \boldsymbol{M}(\boldsymbol{x}).$$

Definition 4:

A *linguistic term* is a (lexical) name associated with a fuzzy set $\boldsymbol{M}$ which is defined on a universe $\boldsymbol{U}$. A *linguistic variable* is a set of linguistic terms. The fuzzy sets these terms are associated with are defined on one universe $\boldsymbol{U}$.

When $\boldsymbol{M}(\boldsymbol{x})$ is replaced with $\boldsymbol{M}^{\alpha}(\boldsymbol{x})$ in the formula in Definition 3, symbol $\#_{\alpha}(\boldsymbol{M})$ is used for denoting the value of the sum of this formula. Let linguistic terms "*sunny*", "*cloudy*", "*rainy*" are associated with fuzzy subsets of universe $\boldsymbol{U}$. Let "*Outlook*" be a linguistic variable defined as follows:

$$Outlook = \{sunny, cloudy, rainy\}.$$

It is said that linguistic terms "*sunny*", "*cloudy*", "*rainy*" are associated with (are defined for) linguistic variable "*Outlook*". Membership degree to which $\boldsymbol{x}$ is an element of the fuzzy set associated with "*sunny*" is symbolized by $sunny(\boldsymbol{x})$. Similarly, if $\boldsymbol{M}$ is the fuzzy set associated with linguistic term "*sunny*", $\#(sunny)$ (resp. $\#_{\alpha}(sunny)/sunny^{\alpha}$) is often used instead of $\#(\boldsymbol{M})$ (resp. $\#_{\alpha}(\boldsymbol{M})/\boldsymbol{M}^{\alpha}$). If linguistic term "*sunny*" is chosen from linguistic terms predefined for the variable "*Outlook*", it is denoted by "*Outlook* is *sunny*". New linguistic terms can be derived from linguistic terms defined for a variable when operators such as AND or OR are used and when the corresponding membership degrees are computed with t-norm and s-norm. So e.g., if linguistic term "*sunny* OR *cloudy*" is derived from linguistic terms predefined for the variable "*Outlook*", it is denoted by "*Temperature* is *sunny* OR *cloudy*". Symbol $\underset{\boldsymbol{x}\in \boldsymbol{X}}{\operatorname{argmax}}\{\mathrm{f}(\boldsymbol{x})\}$, where $\boldsymbol{x}$ is a linguistic term and $\boldsymbol{X}$ is a linguistic variable, stands for a set of all $\boldsymbol{x}\in\boldsymbol{X}$ which $\mathrm{f}(\boldsymbol{x})$ has the maximal value for.

Definition 5:

Let $\boldsymbol{U}$ be the set of all possible instances and let all $\boldsymbol{e}\in\boldsymbol{U}$ be described by linguistic variables $\boldsymbol{A} = \{A_1; \ldots; A_k; \ldots A_n\}$. A *linguistic condition* E is a linguistic term associated with a subset terms(E) of linguistic terms defined for linguistic variables in $\boldsymbol{A}$. Its lexical name is a connection of terms in terms(E) with conjunction "AND". For any possible linguistic variable in $\boldsymbol{A}$ there is at most one linguistic term from linguistic terms defined for this linguistic variable. E is associated with a fuzzy set whose membership degrees

$E(\boldsymbol{e})$, $\boldsymbol{e}\in\boldsymbol{U}$, are defined as follows: $E(\boldsymbol{e})=1$ if terms(E)=$\emptyset$, otherwise the value of $E(\boldsymbol{e})$ is the result of t-norm applied on all $a_{k,l}\in\text{terms}(E)$.

Linguistic term "$a_{k,l}$" in terms(E) of a linguistic condition E, $a_{k,l}\in A_{\boldsymbol{k}}\in \boldsymbol{A}$, is equally replaced by "A_k is $a_{k,l}$" and vice versa. For example, linguistic condition "$a_{3,3}$ AND $a_{5,2}$" equals "A_3 is $a_{3,3}$ AND A_5 is $a_{5,2}$". The following symbols are also used. $E=\emptyset/E\neq\emptyset$ means terms(E)=$\emptyset$ /terms(E)$\neq\emptyset$. If there is a linguistic term/no linguistic term defined for linguistic variable A_k in terms(E), we write $A_k\in E/A_k\notin E$. Symbol $E\setminus A_k$ means removing the linguistic term $a_{k,l}\in A_{\boldsymbol{k}}$ from terms(E) if there is such. $E\cup a_{k,l}$ where $A_k\notin E$ is a linguistic condition "E AND A_k is $a_{k,l}$" if $E\neq\emptyset$ and "A_k is $a_{k,l}$" if $E=\emptyset$. Its $(E\cup a_{k,l})(\boldsymbol{e})$=$\mathbf{T}(E(\boldsymbol{e}),a_{k,l}(\boldsymbol{e}))$, $\mathbf{T}$ is t-norm. $E\cup c_j$ where $c_j\in C$ is a linguistic term "E AND C is c_j" if $E\neq\emptyset$ and "C is c_j" if $E=\emptyset$. Its $(E\cup c_j)(\boldsymbol{e})$=$\mathbf{T}(E(\boldsymbol{e}),c_j(\boldsymbol{e}))$, $\mathbf{T}$ is t-norm.

The paper is organized as follows. In Section 2 and Section 3, the decision tree based on classification and on cumulative information are described respectively. Section 4 contains a comparison of these methods. Section 5 concludes this paper.

2. Decision tree based on classification ambiguity

The input parameters in Table. 1 are supposed for the method based on classification ambiguity (10). This table is a result of summary done after experimenting with the method.

Table. 1: Input parameters of the method based on classification ambiguity.

Input parameter	The meaning of the input parameter
α	*Significance level*. It serves as a filter of insignicant membership degrees in instances $\boldsymbol{e}\in\boldsymbol{V}$. By this, ambiguity can be eliminated and the importance of higher values of $a_{k,l}(\boldsymbol{e})$ and $c_j(\boldsymbol{e})$, $\boldsymbol{e}\in\boldsymbol{V}$, can be increased. $\alpha\in[0,1]$.
β	*Degree-of-truthfulness threshold.*. It controls the minimal truthfulness of fuzzy rules made from the fuzzy decision tree. Lower value leads to simpler fuzzy rules, i.e. to fuzzy rules with lower number of "*Linguistic variable* is *linguistic term*" in their conditions. On the other hand, the accuracy of determining the

	values of $c_j(\boldsymbol{e})$ for all $c_j \in C$ is often lowered when they are used. If the value of β increases to a certain level, the increasing of accuracy of determining $c_j(\boldsymbol{e})$, $c_j \in C$, stops. It sometimes happens there are not any rules with a high value of β. $\beta \in [0,1]$.
ν	*Simplify*. It is a boolean parameter. $\nu \in \{true, false\}$. If its value is *true*, made fuzzy rules are simplified after the transformation from the decision tree. Simplification consists in decreasing the number of "*Linguistic variable* is *linguistic term*" in the conditions of the fuzzy rules while β of each fuzzy rule is kept. By this, the fuzzy rules become more general. There is also a higher probability that when they are used, the values of $c_j(\boldsymbol{e})$ for all $c_j \in C$ are determined more correctly for more instances $\boldsymbol{e} \in \boldsymbol{U}$.
ω	*Keep the current degree of truthfulness***.** It is a boolean parameter. $\omega \in \{true, false\}$. If its value is *true*, no fuzzy rule is replaced with a simplified fuzzy rule with a lower value of degree of truthfulness comparing to the original fuzzy rule when the fuzzy rules are simplified. If the value of ν is *false*, the value of parameter ω is irrelevant.
A	*Input linguistic variables*. Linguistic variables that describe instances $\boldsymbol{e} \in \boldsymbol{U}$, where $\boldsymbol{U}$ is the universe of all possible instances in the task.
C	*Class linguistic variable*.
$\boldsymbol{V}$	*Known instances (learning set)***.** A set of instances $\boldsymbol{e} \in \boldsymbol{U}$ which membership degrees associated with linguistic terms $a_{k,l} \in A_k \in \boldsymbol{A}$ and $c_j \in C$ are known for.

Input parameters and their meanings give a user perception about the method. The method itself uses measures described now. Table. 2 contains a summary of measures important for the described method and their meanings. Measures are formulated with t-norm defined as $\mathbf{T}(a,b) = \min\{a,b\}$.

Measure $\mathbf{DTF}_\alpha(E;c_j;\boldsymbol{V})$ (*degree of truthfulness*) is based on (5) and it is counted for fixed linguistic condition E, known instances $\boldsymbol{V}$ and significance level α. The definition of measure $\mathbf{DTF}_\alpha(E;c_j;\boldsymbol{V})$ is as follows:

$$\mathbf{DTF}_\alpha(E;c_j;\boldsymbol{V})=\frac{\sum_{e\in V}\min\{c_j^\alpha(\boldsymbol{e}),E^\alpha(e)\}}{\sum_{e\in V}E^\alpha(\boldsymbol{e})}\text{ for all }c_j\in C.$$

$\mathbf{DTF}_\alpha(E;c_j;\boldsymbol{V})$ for all $c_j\in C$ can be considered to be the possibility distribution, for known instances $\boldsymbol{V}$, that the fuzzy set associated with $E\cup c_j$ at significance level α is a subset of the fuzzy set associated with $c_j\in C$ at significance level α (10).

Let us define $\pi_\alpha(E;c_j;\boldsymbol{V})$ on the basis of (10) as the possibility of classifying an instance to linguistic term $c_j\in C$ for given significance level α, linguistic condition E and known instances $\boldsymbol{V}$ as follows:

$$\pi_\alpha(E;c_j;\boldsymbol{V})=\frac{\mathbf{DTF}_\alpha(E;c_j;\boldsymbol{V})}{\max_{c_k\in C}\{\mathbf{DTF}_\alpha(E;c_j;\boldsymbol{V})\}}\text{ for all }c_j\in C.$$

Let π_E be the possibility distribution on C which values $\pi_\alpha(E;c_j;\boldsymbol{V})$ for all $c_j\in C$ are ordered in non-increasing order in and let first respective value (i.e. the highest value) be represented by $\pi_E(1)$, the second respective value by $\pi_E(2)$, …, and the last lowest value by $\pi_E(\#(C))$. Let $\pi_E(\#(C)+1)=0$. Then on the basis of (10) and (4), classification ambiguity of known instances in $\boldsymbol{V}$ classified into terms $c_j\in C$ if a linguistic condition $\boldsymbol{E}$ is known is defined as follows. If $\#(C)=1$ then $\mathbf{AMB}_\alpha(E;C;\boldsymbol{V})=0$. If $\#(C)>1$ then its definition is:

$$\mathbf{AMB}_\alpha(E;C;\boldsymbol{V})=\sum_{i=2}^{\#(C)}(\pi_E(i)-\pi_E(i+1))\cdot\ln i.$$

For measuring classification ambiguity in known instances $\boldsymbol{e}\in\boldsymbol{V}$ when linguistic variable $A_k\in\boldsymbol{A}$ is considered and a linguistic condition E is known, measure $\mathbf{AMB}_\alpha(A_k;E;C;\boldsymbol{V})$ is used. On the basis of (10), for $A_k\notin E$, it is defined as:

$$\mathbf{AMB}_\alpha(A_k;E;C;\boldsymbol{V})=\sum_{a_{k,l}\in A_k}\frac{\#(E\cup a_{k,l})}{\sum_{a_{k,l}\in A_k}\#(E\cup a_{k,m})}\cdot\mathbf{AMB}_\alpha(E\cup a_{k,l};C;\boldsymbol{V}).$$

Table. 2: Sense of measures for the method based on classification ambiguity.

Measure	Sense for the method
$\mathbf{DTF}(E;c_j;\boldsymbol{V})$	$\mathbf{DTF}(E;c_j;\boldsymbol{V})$ = $\mathbf{DTF}_0(E;c_j;\boldsymbol{V})$. Degree of truthfulness of the

Measure	Sense for the method
	rule "**IF** E **THEN** C is c_j" for known instances $\boldsymbol{e}\in\boldsymbol{V}$, $c_j\in C$. It is used as one of ending conditions when a decision tree is build and as a measure when fuzzy rules are simplified.
$\mathbf{AMB}_\alpha(E;C;\boldsymbol{V})$	Classification ambiguity of known instances $\boldsymbol{e}\in\boldsymbol{V}$ into class linguistic terms $c_j\in C$ if linguistic fuzzy condition E is known. It is used for counting a condition for deciding about excluding the branch that is not processed when a decision tree is build.
$\mathbf{AMB}_\alpha(A_k;E;C;\boldsymbol{V})$	Classification ambiguity of known instances $\boldsymbol{e}\in\boldsymbol{V}$ when linguistic variable $A_{\boldsymbol{k}}$ is considered and linguistic condition E is known. It is used for deciding about associating linguistic variable $A_{\boldsymbol{k}}\notin E$ with the node of a decision tree which is connected to the end of the branch corresponding to E.

The actual making of fuzzy rules consists of several partial activities. First, the decision tree is build. Then it is transformed into fuzzy rules. These fuzzy rules are either kept or simplified according to the value of parameter ν. Particular partial activities for making respective rules are expressed in series of tables Table. 3, Table. 4 and Table. 5.

Table. 3: Making a decision tree with reduction of classification ambiguity.

decision tree := makeTree(α; β; $\boldsymbol{A}$; C; $\boldsymbol{V}$)	
Step 1	Make the root and associate linguistic variable $minA_k := \underset{A_k\in\mathbf{A}}{\operatorname{argmin}}\{\mathbf{AMB}_\alpha(A_k;\varnothing)\}$ with it. Make a branch for each $a_{k,l}\in minA_k$, connect them with the root, associate them with the particular $a_{k,l}$ and consider them unprocessed.
Step 2	If there is no unprocessed branch, END. Otherwise, choose one of the unprocessed branches and consider it the current branch. Make linguistic condition E for the current branch. Linguistic condition E consists of all "Linguistic variable associated with the node is linguistic term associated with the branch" from the root to the current branch connected with operator AND.

decision tree := makeTree(α; β; $\boldsymbol{A}$; C; $\boldsymbol{V}$)	
Step 3	Set $\max := \max_{c_j \in C} \{ \mathbf{DTF}(E; c_j; \boldsymbol{V}) \}$ and $c_{\max} := \operatorname{argmax}_{c_j \in C} \{ \mathbf{DTF}(E; c_j; \boldsymbol{V}) \}$. If $\max > \beta$ then make the leaf which is associated with linguistic term $c_{\max}$, connect it with the current branch, consider the current branch processed and go to the **Step 2**. Otherwise, go to the **Step 4**.
Step 4	If there is not any $A_k \in \boldsymbol{V}$, $A_k \notin E$, consider the current branch processed and go to the **Step 2**. Otherwise, set the value of $\text{E\_AMB} := \mathbf{AMB}_{\alpha}(E; C; \boldsymbol{V})$, $\text{minAMB} := \min_{A_k \in \boldsymbol{V},\, A_k \notin E} \{ \mathbf{AMB}_{\alpha}(A_k; E; C; \boldsymbol{V}) \}$, $minA_k := \operatorname{argmin}_{A_k \in \boldsymbol{V},\, A_k \notin E} \{ \mathbf{AMB}_{\alpha}(A_k; E; C; \boldsymbol{V}) \}$. If $\text{minAMB} < \text{E\_AMB}$, go to the **Step 5**. Otherwise, consider the current branch processed and go to the **Step 2**.
Step 5	Make a node which linguistic variable $minA_k$ is associated with. Make a branch coming from this node for each $a_{k,l} \in A_k$ and consider them unprocessed. Connect this node with the current branch and consider this branch processed. Go to the **Step 2**.

Table. 4: Transformation of the tree made according to Table. 3 to fuzzy rules.

{fuzzy rule}:= makeFuzzyRules(decision tree)	
Step 1	For each leaf ^{i} of the decision tree, mark the linguistic term associated with leaf ^{i} as c^i. For each leaf ^{i}, take the branch coming to this leaf and make linguistic condition E_i for this branch. This linguistic condition E_i contains expressions "*Linguistic variable of the node* is *linguistic term of the branch*" from the root to that branch and they are connected with operator AND.
Step 2	For each leaf ^{i} make fuzzy rule "**IF** E_i **THEN** C is c^i".

Table. 5: Simplification of fuzzy rules made according to Table. 4.

{fuzzy rule}:=simplifyFuzzyRules({fuzzy rule} from Table. 4; β; ω; $\boldsymbol{A}$; C; $\boldsymbol{V}$)	
Step 1	For each rule "**IF** E_i **THEN** C is c_j" that has more than one linguistic variable

{fuzzy rule}:=simplifyFuzzyRules({fuzzy rule} from Table. 4; β; ω; $\boldsymbol{A}$; C; $\boldsymbol{V}$)	
	$A_k \in \boldsymbol{A}$ in its condition do: Set $\max := \max\limits_{A_k \in E_i} \{\mathbf{DTF}(E_i \setminus A_k; c_j; \boldsymbol{V})\}, maxA_k := \operatorname{argmax}\limits_{A_k \in E_i} \{\mathbf{DTF}(E_i \setminus A_k; c_j; \boldsymbol{V})\}$, $\text{dtf} := \mathbf{DTF}(E_i; c_j; \boldsymbol{V})$. If $\{\omega = \text{true and } \max \geq \text{dtf}\}$ or $\{\omega = \text{false and } \max \geq \beta\}$ then replace "**IF** E_i **THEN** C is c_j" with „**IF** $E_i \setminus maxA_k$ **THEN** C is c_j".
Step 2	Ak došlo v kroku **Krok 1** k aspoň jednej náhrade fuzzy pravidla, opakuj krok **Krok 1**. V opačnom prípade odstráň opakujúce sa fuzzy pravidlá.

Classification of an instance $\boldsymbol{e} \in \boldsymbol{U}$ means determining values of $c_j(\boldsymbol{e})$ for all $c_j \in C$. Fuzzy rules made according to Table. 4 or Table. 5 can be used for classification of an $\boldsymbol{e} \in \boldsymbol{U}$ when the algorithm in Table. 6 is employed.

Table. 6: Classification of an instance e∈U with fuzzy rules made according to Table. 4/5.

$\{c_j(\boldsymbol{e})\}$: = classify({fuzzy rule}from Table. 4/5; $\boldsymbol{e}$; C) for $\boldsymbol{e} \in \boldsymbol{U}$ whose $a_{k,l}(\boldsymbol{e})$, $a_{k,l} \in A_k \in \boldsymbol{A}$, are known	
Step 1	If made fuzzy rules do not contain at least one fuzzy rule "**IF** E_i **THEN** C is c_j", it is better to repeat the process of making fuzzy rules. It is recommended to use $\boldsymbol{V}$ with greater $\#(\boldsymbol{V})$. If there is not such a $\boldsymbol{V}$ available, it is possible to set $c_j(\boldsymbol{e}) := 0$ for c_j without any "**IF** E_i **THEN** C is c_j".
Step 2	For each fuzzy rule "**IF** E_i **THEN** C is c_j", compute the membership degree of the linguistic condition E_i. These degrees have the following mark: ${}^{c_j}E_i(\boldsymbol{e})$.
Step 3	Divide fuzzy rules into $\#(C)$ groups marked ^{c_j} on the basis of linguistic terms $c_j \in C$ in their conclusions.
Step 4	For each group ^{c_j}, compute the maximum of values of all ${}^{c_j}E_i(\boldsymbol{e})$. The result of this computation is the value of $c_j(\boldsymbol{e})$.

3. Decision tree based on cumulative information

The input parameters in Table. 7 are supposed for the method based on maximization of cumulative information (7). This table is a result of summary done after experimenting with the method.

Table. 7: Input parameters of the method based on cumulative information.

Input parameter	The meaning of the input parameter
α	*Frequency-of-branch threshold*. It controls the growth of the decision tree on the basis of the frequency of branch. The higher the value of α is, the lower the height of the decision tree is (or equally the lower the number of "*Linguistic variable* is *linguistic term*" in conditions of made fuzzy rules is). $\alpha \in [0,1]$.
β	*Frequency-of-class threshold***.** It controls the growth of the decision tree on the basis of the frequency of class linguistic term. The lower the value of β**,** the lower the height of the decision tree is (or equally the lower the number of "*Linguistic variable* is *linguistic term*" in conditions of made fuzzy rules is). Increasing the value of α and decreasing the value of β can lead to a potentially better classification of a higher number of unknown instances $\boldsymbol{e} \in \boldsymbol{U} \setminus \boldsymbol{V}$. However, it can lead to a less accurate classification of especially known instances $\boldsymbol{e} \in \boldsymbol{V}$. $\beta \in [0,1]$.
CRIT	*Criterion for association of a linguistic variable* $A_k \in \boldsymbol{A}$ with a node when the decision tree is built.
$\boldsymbol{A}$	*Input linguistic variables*. Linguistic variables that describe instances $\boldsymbol{e} \in \boldsymbol{U}$, where set $\boldsymbol{U}$ is the universe which it is worked in the task of making classical fuzzy classification rules with.
C	*Class linguistic variable*.
$\boldsymbol{V}$	*Known instances (learning set)***.** A set of instances $\boldsymbol{e}$ from $\boldsymbol{U}$ which the values of membership degrees associated with linguistic terms $a_{k,l} \in A_k \in \mathbf{A}$ and $c_j \in C$ are known for.

Input parameters and their meanings give a user perception about the method. The method itself uses measures described now. They are formulated with t-norm $\mathbf{T}(a,b)=a\cdot b$.

Cumulative information of linguistic condition E (linguistic term $E\cup c_j$, $c_j\in C$) is marked as $\mathbf{II}(E;\boldsymbol{V})$ ($\mathbf{II}(E\cup c_j;\boldsymbol{V})$) and defined as:

$$\mathbf{II}(E;\boldsymbol{V})=\begin{cases}-\log_2\#(E) & ;\text{if } E\neq\emptyset\\ -\log_2\#(\boldsymbol{V}) & ;\text{if } E=\emptyset\end{cases}=\begin{cases}-\log_2\sum_{\boldsymbol{e}\in\boldsymbol{V}} E(\boldsymbol{e}); & \text{if } E\neq\emptyset\\ -\log_2\#(\boldsymbol{V}) & ;\text{if } E=\emptyset\end{cases}$$

$$(\mathbf{II}(E\cup c_j;\boldsymbol{V})=-\log_2\#(E\cup c_j)=-\log_2\sum_{\boldsymbol{e}\in\boldsymbol{V}} E(\boldsymbol{e})\cdot c_j(\boldsymbol{e}))\,.$$

Cumulative information is a part of the formula for computing information that describes vagueness of linguistic condition E ($E\cup c_j$, $c_j\in C$). The lower the value of information, the higher the value of vagueness of E ($E\cup c_j$, $c_j\in C$). It is defined as follows:

$$\mathbf{I}(E;\boldsymbol{V})=\log_2\#(\boldsymbol{V})+\mathbf{II}(E;\boldsymbol{V}),$$

$$\mathbf{I}(E\cup c_j;\boldsymbol{V})=\log_2\#(\boldsymbol{V})+\mathbf{II}(E\cup c_j;\boldsymbol{V}).$$

Conditional information (conditional cumulative information) of $c_j\in C$ provided that E is known $\mathbf{I}(c_j/E;\boldsymbol{V})$ ($\mathbf{II}(c_j/E;\boldsymbol{V})$). It describes the vagueness of linguistic term c_j if linguistic condition E is known. It is defined as:

$$\mathbf{I}(c_j/E;\boldsymbol{V})=\mathbf{II}(c_j/E;\boldsymbol{V})=\mathbf{II}(E\cup c_j;\boldsymbol{V})-\mathbf{II}(E;\boldsymbol{V}).$$

For determining the amount of information which is obtained about linguistic variable C if values of $E(\boldsymbol{e})$, $\boldsymbol{e}\in\boldsymbol{V}$, and $a_{k,l}(\boldsymbol{e})$, $a_{k,l}\in A_k\notin\boldsymbol{E}$, $\boldsymbol{e}\in\boldsymbol{V}$, are known, mutual information $\mathbf{I}(C;E;A_k;\boldsymbol{V})$ is used. It is defined as follows:

$$\mathbf{TMP}(c_j;a_{k,l};E;\boldsymbol{V})=\mathbf{II}(E\cup c_j;\boldsymbol{V})+\mathbf{II}(E\cup a_{k,l};\boldsymbol{V})-\mathbf{II}(\boldsymbol{E}\cup a_{k,l}\cup c_j;\boldsymbol{V})-\mathbf{II}(E;\boldsymbol{V}),$$

$$\mathbf{I}(C;\boldsymbol{E};A_k;\boldsymbol{V})=\sum_{c_j\in C}\sum_{a_{k,l}\in A_k}\#(\boldsymbol{E}\cup a_{k,l}\cup c_j)\cdot\mathbf{TMP}(c_j;a_{k,l};E;\boldsymbol{V}).$$

Cumulative entropy of linguistic variable $A_k\in\boldsymbol{A}$ marked with $\mathbf{HH}(A_k;\boldsymbol{V})$ is defined as:

$$\mathbf{HH}(A_k;\boldsymbol{V})=\sum_{a_{k,l}\in A_k}\#(a_{k,l})\cdot\mathbf{II}(a_{k,l};\boldsymbol{V}).$$

This formula is a part of the definition of Shannon entropy $\mathbf{H}(A_k;\boldsymbol{V})=\dfrac{\mathbf{HH}(A_k;\boldsymbol{V})}{\#(\boldsymbol{V})}$. Shannon entropy of linguistic variable $A_k\in\boldsymbol{A}$ represents the average value of information which is obtained if values of $a_{k,l}(\boldsymbol{e})$, $a_{k,l} \in A_k$, are got to know. Table. 8 contains a summary of measures important for the described method and their meanings.

Table. 8: Sense of measures for the method based on cumulative information.

Measure	Sense for the method
$\mathbf{CRIT}(C; A_k; E; \boldsymbol{V})$	**CRIT** represents the criterion for association of linguistic variable $A_k \in \boldsymbol{A}$ with a node when the decision tree is build. There are several criteria of this kind, e.g.: $\mathbf{CRIT}(C; A_k; E; \boldsymbol{V}) = \mathbf{I}(C; E; A_k; \boldsymbol{V}) \rightarrow \max$, $\mathbf{CRIT}(C; A_k; E; \boldsymbol{V}) = \frac{\mathbf{I}(C; E; A_k; \boldsymbol{V})}{\mathbf{HH}(A_k; \boldsymbol{V})} \rightarrow \max$. The latter criterion is some kind of relative value of the former. The former criteria gives preference to linguistic variables with higher cardinality.
$\mathbf{II}(E; \boldsymbol{V})$	It serves as one of ending criteria when the decision tree is built. It is used for deciding about assigning a leaf to the branch corresponding to E on the basis of the validity of $\mathbf{II}(E; \boldsymbol{V}) \geq -\log_2(\alpha \cdot \#(\boldsymbol{V}))$.
$\mathbf{I}(c_j / E; \boldsymbol{V})$	It serves as one of ending criteria when the decision tree is built on the basis of the validity of the following: $\min_{c_j \in C}\{\mathbf{I}(c_j / \boldsymbol{E})\} \leq -\log_2(\beta)$.
$\mathbf{F}(c_j / E; \boldsymbol{V})$	$\mathbf{F}(c_j / E; \boldsymbol{V}) = 2^{-\mathbf{I}(c_j / E; \boldsymbol{V})}$. The value of this parameter for all $c_j \in C$ forms extra criteria for the rule (ECR) with a linguistic condition E. ECRs are a part of the formulas for counting $c_j(\boldsymbol{e})$, $\boldsymbol{e} \in \boldsymbol{U}$ whose $a_{k,l}(\boldsymbol{e})$ for all $a_{k,l} \in A_k \in \boldsymbol{A}$ are known. $\mathbf{F}(c_j / E; \boldsymbol{V})$ is interpreted as the *frequency of class* $c_j \in C$ for the fuzzy rule with linguistic condition E.

Measures showed in Table. 8 form the basis for describing the method based on cumulative information. It starts with making a decision tree according to Table. 9. It continues with a transformation of this decision tree to fuzzy rules according to Table. 10.

Table. 9: Making a decision with maximization of cumulative information.

decision tree := makeTree(α; β; **CRIT**; A; C; V)	
Step 1	Make the root and associate linguistic variable $maxA_k := \underset{A_k \in A}{\text{argmax}} \{\mathbf{CRIT}(C; A_k; \varnothing; V)\}$ with it. Make a branch for each $a_{k,l} \in maxA_k$, connect them with the root, associate them with the particular $a_{k,l}$ and consider them unprocessed.
Step 2	If there is no unprocessed branch, END. Otherwise, choose one of the unprocessed branches and consider it the current branch. Make linguistic term E for the current branch. E consists of all "*Linguistic variable associated with the node* is *linguistic variable associated with the branch*" from the root to the current branch connected with operator AND.
Step 3	Set $\text{branchII} := \mathbf{II}(E; V)$ and $\text{minClassI} := \underset{c_j \in C}{\min}\{\mathbf{I}(c_j / E; V)\}$. If $\{\text{branchII} \geq -\log_2(\alpha * \#(V))\}$ or $\{\text{minClassI} \leq -\log_2(\beta)\}$ or $\{(A \setminus E) = \varnothing\}$, go to **Step 4**, otherwise go to **Step 5**.
Step 4	Make a leaf, connect it with the current branch and consider this branch processed. Associate values $\text{F}(c_j) := \mathbf{F}(c_j / E; V)$ for all $c_j \in C$ and linguistic term $\underset{c_j \in C}{\text{argmax}}\{\text{F}(c_j)\}$ with the made leaf. Go to **Step 2**.
Step 5	Make a node, connect it with the current branch and associate linguistic variable $maxA_k := \underset{A_k \in V,\ A_k \notin E}{\text{argmax}} \{\mathbf{CRIT}(C; A_k; E; V)\}$. Consider the current branch processed. Make a branch for each $a_{k,l} \in maxA_k$, connect them with the made node, associate them with particular $a_{k,l}$ and consider them unprocessed. Go to **Step 2**.

Table. 10: Transformation of a decision tree from Table. 9 to fuzzy rules and ECRs.

{fuzzy rule$_i$, ECR$_i$}:= makeFuzzyRules+ECRs(decision tree from Table. 9)	
Step 1	For each leaf i of the decision tree, mark the linguistic term associated with it as c^i. For each leaf i, take the branch going to it and make linguistic condition E_i for this branch. E_i consists of all "*Linguistic variable associated with the node* is *linguistic term associated with the branch*" from the root to the branch and they are connected with operator AND. Set ECR_i := {$\mathrm{F}_i(c_j)$ | $\mathrm{F}_i(c_j)$ which were associated with leaf i} for each leaf i.
Step 2	Make a fuzzy rule in the form of "**IF** E_i **THEN** C is c^i" for each E_i.

The fuzzy rules coming from the algorithm in Table. 10 are used for classification of instances $\boldsymbol{e}\in\boldsymbol{U}$, i.e. for determining the values of $c_j(\boldsymbol{e})$, $c_j\in C$, $\boldsymbol{e}\in\boldsymbol{U}$, according to Table. 11.

Table. 11: Classification of instance $\boldsymbol{e}\in\boldsymbol{U}$ with fuzzy rules and ECR made according to Table. 10.

$\{c_j(\boldsymbol{e})\}$: = classify({fuzzy rule$_i$, ECR$_i$} from Table. 10; $\boldsymbol{e}$; C) for $\boldsymbol{e}\in\boldsymbol{U}$ whose $a_{k,l}(\boldsymbol{e})$, $a_{k,l}\in A_k\in\boldsymbol{A}$, are known	
Step 1	For each fuzzy rule "**IF** E_i **THEN** C is c_j", compute $E_i(\boldsymbol{e})$.
Step 2	Set $c_j(e):=\sum_{\forall i} E_i(\boldsymbol{e})\cdot\mathrm{F}_i(c_j)$ where $\mathrm{F}_i(c_j)\in\mathrm{ECR}_i$, $c_j\in C$.

4. Comparison of the methods

Two methods described in the previous chapters have been implemented in programming language Java within software library Fuzzy Rule Miner technically introduced in (2). This software library was used for testing the methods on several databases from the UCI Repository (1). The names of these databases are Ecoli Database (Ecoli), Haberman's Survival Data (Haberman) and Wine Recognition Database (Wine). They contain categorical and numerical attributes. These attributes and their values were fuzzified into linguistic variables and membership degrees according to (6). When algorithms were tested, each database was randomly divided 100 times into a learning set of instances and a testing set of instances. The learning and testing set of instances were always composed of 70% and 30% of the database, respectively. For each division, the learning set of instances was used for making fuzzy rules and the instances in the testing set of instances were classified with the fuzzy rules. Then the original class values of the instances were compared with the made ones. Finally, the error rate for a database and an algorithm was computed as the ratio of the number of misclassification combinations to the total number of combinations.

Table. 12: Experimental results.

Database/Method	MCA	MCI
Ecoli	0.3112	0.2022
Haberman	0.2614	0.2624
Wine	0.08170	0.04566
Final error rate	0,2181	0,1701

The experimental results for the two methods on chosen databases are in Table. 12. MCA is method based on classification ambiguity, MCI means method based on cumulative information. For each method and each database, the error rate is in the table. The last row of the table, marked Final error rate, contains the average error rates for particular methods. The final error rate of MCI is lower than the final error rate of MCA. It means the method based on cumulative information achieves better results than the method based on classification ambiguity in the conducted experiments.

5. Conclusions

The contribution of this paper is:

- Summary and a united description of two methods for making fuzzy rules on the basis of decision trees, using notions of fuzzy logic (fuzzy sets, linguistic variable, possibility distribution),
- Experimental comparison of these two methods (the method based on classification ambiguity and the method based on cumulative information) on databases from the UCI Repository. The comparative criterion is error rate. The latter method seems to have better results.

6. Bibliography

(1) Asuncion, A., Newman, D., J.: UCI Machine Learning Repository [http://www.ics.uci.edu/ mlearn/MLRepository.html], Irvine, University of California, Department of Information and Computer Science, CA, 2007.

(2) Bohacik, J., Matiaško, K., Levashenko, V.: Software for Making and Using Fuzzy Rules, Journal of ELECTRICAL ENGINEERING, Vol. 57, 2006, pp. 85-88, ISSN 1335-3632.

(3) Chang, L.-Y., Chen W.-Ch.: Data mining of tree-based models to analyze freeway accident frequency, Journal of Safety Research, Vol. 36, No. 4, 2005, pp. 365-375, ISSN: 0022-4375.

(4) Higashi, M., Klir, G. J.: Measures of uncertainty and information based on possibility distributions, International Journal General Systems, Vol. 9, No. 1, 1983, pp. 43-58, ISSN: 0308-1079.

(5) Kosko, B.: Fuzzy entropy and conditioning, Information Sciences, Vol. 40, No. 2, 1986, pp. 165-174, ISSN: 0020-0255.

(6) Lee, H.-M., Chen, C.-M., Chen, J.-M., Jou, Y.-L.: An efficient fuzzy classiffier with feature selection based on fuzzy entropy, IEEE Transaction on Systems, Man, and Cybernetics - Part B: Cybernetics, Vol. 31, No. 3, 2001, pp. 426-432, ISSN: 1083-4419.

(7) Levashenko, V., Zaitseva, E.: Usage of new information estimations for induction of fuzzy decision trees. In: Proc. of the 3rd IEEE International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2002), Springer, Manchester, pp. 493-499, 2002, ISSN: 3-540-44025-9.

(8) Trnka A.: Proposal of application datawarehouses into control process. In: Proc. of the International Doctoral Seminar, AlumniPress, Smolenice /May 17-19, 2009/, pp. 343-348, 2009, ISBN 978-80-8096-088-9.

(9) Trnka A., Tanuska P.: Datawarehousing and data mining methodologies as a one of the possible benefits in control process. In: Aktualnye problemy i innovacii v ekonomike, upravlenii, obrazovanii, informacionnych technologijach : materialy mezdunarodnoj naucnoj konferencii (12-15 maja 2009 goda, Stavropol-Kislovodsk) - Vol. 5, No. 3, pp. 84-87., 2009, ISSN 2074-1685.

(10) Yuan, Y., Shaw, M. J.: Induction of fuzzy decision trees, Fuzzy Sets and Systems, Vol. 69, No. 2, pp. 125-139, 1995, ISSN: 0165-0114.

7. Address(es) of the author(s)

Jan Bohacik (Ján Boháčik), Ing., PhD.
Katedra informatiky
Fakulta riadenia a informatiky
Žilinská univerzita
Univerzitná 8215/1
010 26 Žilina
SLOVAKIA
Jan.Bohacik@fri.uniza.sk

Karol Matiaško, prof., Ing., PhD.
Katedra informatiky
Fakulta riadenia a informatiky
Žilinská univerzita
Univerzitná 8215/1
010 26 Žilina
SLOVAKIA
Karol.Matiasko@fri.uniza.sk

Rosaldo Rossetti, PhD.
Departamento de Engenharia Informatica
Faculdade de Engenharia da Universidade do Porto
Rua Dr. Roberto Frias
4200-465 Porto
PORTUGAL
rossetti@fe.up.pt

This work is made with the support of

Centre of excellence for systems and services of intelligent transport

ITMS code of the project 26220120028

University of Žilina in Žilina

Operačný program VÝSKUM a VÝVOJ

ERDF - Európsky fond regionálneho rozvoja

Projekt je spolufinancovaný zo zdrojov ES

EURÓPSKA ÚNIA ŠTRUKTURÁLNE FONDY

Accepted for publication in March 2010 by the publisher of the Journal of Information Technologies (ISSN 1337-7469), i.e. by Katedra aplikovanej informatiky FPV UCM in Trnava.