http://edoc.bseu.by

USING INFORMATION THEORETICAL APPROACH FOR CONSTRUCTION OF FUZZY DECISION TREES

Vitaly G. Levashenko[1], Anatoly N.Morozevich[2], Elena N. Zaitseva[2] and Stefan Kovalik[1]

[1]Department of Informatics, Faculty of Management Science and Informatics, University of Žilina, Moyzesova 20, 01026, Žilina, SLOVAKIA, levashen @yahoo.com
[2]Department of Information Technologies, Belarus State Economic University, Partizan ave, 26, 220672, Minsk, BELARUS, zaitseva @yahoo.com

ABSTRACT

The construction of fuzzy decision trees is an important way of acquiring imprecise knowledge. Fuzzy ID3 and its variants are efficient methods of making fuzzy decision trees from a group of training examples. This algorithm is used for solving a number of practical problems in the different fields of human activities. The development of this algorithm is introduced in this paper. The new class of properties for such algorithms is developed with this algorithm. The experimental example demonstrates our idea.

1. INTRODUCTION

Last time an information theoretic approach to solve some problems attracts specialists' attention. For instance, enough simple estimations for characteristics of mutual relation between values of attributes, can be obtained by using this approach. In other cases, the information theoretic approach can be an original alternative to traditional methods and approaches, in particular, when solving tasks through decision trees.

The information-theoretic approach [1,2] suggests some advantages for learning in formation patterns from large sets of imperfect data, and it uses tool, based on the Information Theory. The measures of information content, may include proper information, a joint information, a mutual information, a conditional information, and their entropy.

Among the applications of mutual information approach to decision tree design reported already are medical diagnosis, an expert system design tool, character recognition, various control algorithms based on unification of methods of fuzzy logic and information theory. Using an information-theoretic approach to data cleaning is presented in [3]. For analysis of database, the approach has been successfully applied to extracting probabilistic rules from pairs of interdependent attributes in a database [1] and to evaluating reliability of database attributes [4].

In the paper [5] was proposed a general top-down mutual information algorithm to design decision trees. The author shows that the induction of decision trees is an efficient way of learning from examples. Many methods have been developed for constructing decision trees [6]. Fuzzy ID3 algorithm and its variants [7-12] are efficient methods of making fuzzy decision trees. The fuzzy ID3 can generate fuzzy decision trees without much computation. It has the great matching speed and is especially suitable for large-scale learning problems.

The proposed technique to compute information estimation for fuzzy sets includes rules to compute entropy and information quantity for fuzzy relatives. These rules lie on our theoretical investigations in this paper.

The paper is structured as follows. Section 2 – explain the formal notation of the mathematical model. Section 3 contains short information about fuzzy decision trees. Section 4 describes the technique of calculation of information measures. We show proposed algorithms in Section 5. And we demonstrate usage of this algorithm for application example in Section 6.

2. NOTATION

We use following formal notation of the model

X	– a given finite set
N and F(X)	– fuzzy subset and the family of all fuzzy subsets
$R=\{A_1,\dots,A_n\}$	– attributes of a relation R
$\Omega_i = (A_{i,1}, \dots, A_{i,m_i})$	– groups of fuzzy subsets, corresponds to an attribute A_i ($i=1,\dots,n$)
$A_{i,j}$ ($j=1,\dots,m_i$)	– j-th fuzzy subset of attribute A_i
$M(N)=\Sigma_{x\in N} F(x)$	– cardinality of a fuzzy subset N
$I(A_{i,j})$	– proper information of a fuzzy subset $A_{i,j}$ of attribute A_i
$I(A_{i2,j2}, A_{i1,j1})$	– joint information of jointly specified a fuzzy subset $A_{i2,j2}$ of attribute A_{i2} and fuzzy subset $A_{i1,j1}$ of attribute A_{i1}
$I(A_{i2,j2} \vert A_{i1,j1})$	– conditional information of fuzzy subset $A_{i2,j2}$ of attribute A_{i2} given another fuzzy subsets $A_{i1,j1}$ of attribute A_{i1}

$I(A_{i2,j2};A_{i1,j1})$ – mutual information of jointly specified fuzzy subsets $A_{i2,j2}$ and $A_{i1,j1}$ of attributes A_{i1} and A_{i2} respectively

$H(A_i)$ – entropy of attribute A_i as a measure of its uncertainty

$H(A_{i2,j2},A_{i1,j1})$ – entropy of jointly specified a fuzzy subset $A_{i2,j2}$ of attribute A_{i2} and fuzzy subset $A_{i1,j1}$ of attribute A_{i1}

$H(A_{i2,j2}|A_{i1,j1})$ or $(H(A_{i2}|A_{i1,j1}))$ – -conditional entropy of fuzzy subset $A_{i2,j2}$ of attribute A_{i2} (or attribute A_{i2}) given fuzzy subset $A_{i1,j1}$ of attribute A_{i1}

$I(A_{i2};A_{i1})$ – mutual information in attributes A_{i1} about attributes A_{i2}

3. FUZZY LOGIC AND FUZZY DECISION TREES

An experienced expert looking at a 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", „not so possible", „absolutely impossible", etc. The expert can represent these empirical characteristics as numerical values. The usage of Boolean approach or Multi-Valued Logic approach not always gives necessary reliance. It is based on the idea of crisp logical approach: some attribute values are correct and others are not. For example, if the valid range of a numeric attribute is {0; 1; 2; 3}, the value of 1.1 is incorrect from the crisp logical point of view. The limitations of this approach are obvious: a real-world valid range may have "flexible" boundaries.

There is another significant drawback of such crisp approach to data evaluation: in many cases it 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.

For example, we talk about an attribute *Outlook* and we have got 3 possible value of the attribute (*Sunny*, *Cloudy*, *Rain*). As a rule it is impossible to unequivocally evaluate a membership of parameter's significance (*Sunny*=1, *Cloudy*=0 and *Rain*=0). Expert is able to evaluate a membership of a fuzzy subset of a pairs (*Sunny*, *Cloudy*) or (*Cloudy*, *Rain*) as (*Sunny*=0,8, *Cloudy*=0,2 and *Rain*=0) more naturally and more exact.

Thus, using the Fuzzy Logic approach and looking at the possibility degree as a fuzzy measure seems a more appropriate way to deal with real world data.

Consider a directed tree of which each edge links two nodes, the initial node and the terminal node. The former is called the father-node of the latter while the latter is said to be the son-node of the former. The node having not its fathernode is the root whereas the nodes having not any sonnodes are called leaves. Consequences

***Definition* 1.** Let $\Omega_i \subset F(X)$ $(1 \le i \le n)$ be n given groups of fuzzy subsets, with the property $|\Omega_i| > 1$. Fuzzy decision tree is a directed tree satisfying [13]

- each node of the tree belongs to F(X);
- for each not_leaf, L, whose all son-nodes constitute a subset of F(X) denoted by Γ, there exists i $(1 \le i \le n)$ such that $\Gamma = \Omega_i \cap L$;
- each leaf corresponds to one or several values of classification decision.

Each group of fuzzy subsets, Ω_i, corresponds to an attribute and each fuzzy subset corresponds to a value of the attribute.

A fuzzy decision tree is a generalization of the crisp case. The paper [13] gives a comparison between fuzzy decision trees and crisp ones (fig. 1). The fuzzy decision tree, regarded as a generalization of the crisp case, is more robust in tolerating imprecise information.

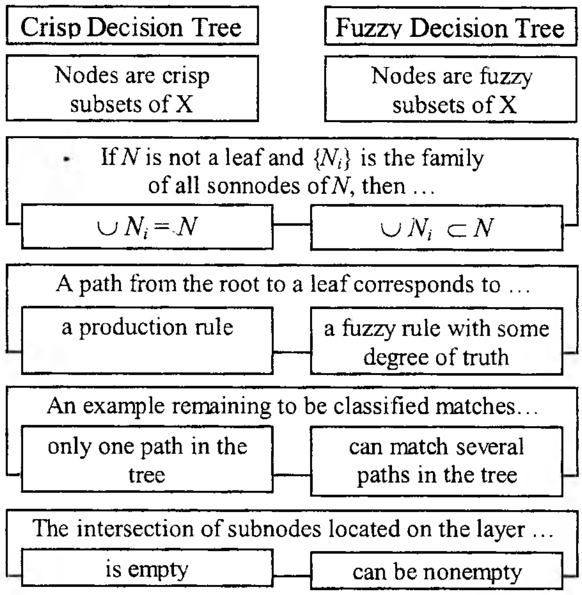

Figure 1. A comparison between the fuzzy decision tree and the crisp case

Let there be N training examples and n attributes $R = (A_1,\ldots,A_n)$. For each attribute A_i $(1 \le i \le n)$ exists group of fuzzy subsets Ω_i. We assume that each group is a set of m_i $(m_i \ge 2)$ values of fuzzy subsets $A_{i,1},\ldots,A_{i,m_i}$. Consequently, $\Omega_i = (A_{i,1},\ldots, A_{i,j},\ldots,A_{i,m_i})$.

A subset $I \subset R$ of input attributes ($|I| \ge 1$): This is a subset of attributes that are assumed to be initial datas and can be used as initial variables for target attributes (see next).

A subset $O \subset R$ of target (output) attributes ($|O| \geq 1$). Our model is aimed to detecting the values of target attributes, based on the information values of these attributes and the input attributes.

The following constraints are imposed on the partition of ours model:

- $I \cap O = \emptyset$, i.e., the same attribute cannot be both an input and a target attribute.
- $I \cup O \subseteq R$; i.e., some attributes are allowed to be neither input, nor target attributes.

***Example* 1**. Consider tabl. 1. Each column corresponds to a fuzzy subset defined on X={1,2,3,...,16}, for instance, sunny = 0.9/1 + 0.8/2 + 0.0/3 +...+ 1.0/16.

Five attributes are as follows $R=(A_1,A_2,A_3, A_4,A_5)$ = {*Outlook*, *Temperature*, *Humidity*, *Wind* and *Games*}.

Outlook=$\{A_{1,1}, A_{1,2}, A_{1,3}\}$={*Sunny*, *Cloudy*, *Rain*},
Temperature=$\{A_{2,1}, A_{2,2}, A_{2,3}\}$={*Hot*, *Mild*, *Cool*},
Humidity =$\{A_{3,1}, A_{3,2}\}$={*Humid*, *Normal*},
Wind = $\{A_{4,1}, A_{4,2}\}$={*Windy*, *Not_windy*}.
Games = $\{A_{5,1}, A_{5,2}, A_{5,3}\}$={*Volleyball*, *Swimming* and *Weight_lifting*}.
I = {*Outlook*, *Temperature*, *Humidity* and *Wind*}.
O ={*Games*}.

Table 1: A small training set

No	*OutLook*			*Temperature*			*Humidity*		*Windy*		*Game**		
	Sunny	*Cloudy*	*Rain*	*Hot*	*Mild*	*Cool*	*Humid*	*Normal*	*Windy*	*Not_win*	*V*	*S*	*W*
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	*0.59*	*0.41*	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	*0.36*	*0.64*	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	*0.57*	*0.43*	0.0

* Three symbols, *V*, *S* and *W*, denote three sports to play: *Volleyball*, *Swimming* and *Weight_lifting*, respectively.

4. INFORMATION-THEORETIC LEARNING

Information Theory [14] suggests a distribution-free modeling of conditional dependency between random variables of both a discrete and a continuous nature. If nothing is known on the causes of a random variable X, its degree of uncertainty can be measured by its unconditional entropy.

4.1. Information in Fuzzy Subset

The main terms and expressions related with notations such as proper information, joint information, condition information and mutual information for a fuzzy subset are introduced in this part. In this part we generalize indicated in paper [15-16] basic information terms of crisp case on case fuzzy subset. These terms allow to evaluate influence of an attributes on another ones.

***Definition* 2**. The proper information of a fuzzy subset $A_{i,j}$ of attribute A_i is

$$I(A_{i,j}) = \log M(A_i)/M(A_{i,j}) \quad \text{bits}, \qquad (1)$$

where log denotes the base 2 logarithm; $(i = 1,\ldots,n; j = 1,\ldots,m_i)$.

***Definition* 3**. The joint (the unconditional) information of a fuzzy subset $A_{i2,j2}$ of attribute A_{i2} and fuzzy subset $A_{i1,j1}$ of attribute A_{i1} is given by

$$I(A_{i1,j1},A_{i2,j2}) = \log \left(M(A_{i1}) / M(A_{i1,j1} \times A_{i2,j2})\right), \qquad (2)$$

***Definition* 4**. The condition information of a fuzzy subset $A_{i2,j2}$ of attribute A_{i2}, given another fuzzy subset $A_{i1,j1}$ of attribute A_{i1} is

$$I(A_{i2,j2}|A_{i1,j1}) = I(A_{i1,j1},A_{i2,j2}) - I(A_{i1,j1}) = \qquad (3)$$

$$= \log \left((M(A_{i1}) \times M(A_{i1,j1})) / M(A_{i1,j1} \times A_{i2,j2})\right).$$

The condition information describes the uncertainty of fuzzy subset $A_{i2,j2}$ when another fuzzy subset $A_{i1,j1}$, is given.

Definition 5. The mutual information of jointly specified fuzzy subset $A_{i2,j2}$ of attribute A_{i2}, and fuzzy subset $A_{i1,j1}$ of attribute A_{i1} satisfies the equations

$$I(A_{i1,j1};\ A_{i2,j2}) = \tag{4}$$

$$= I(A_{i1,j1}) + I(A_{i2,j2}) - I(A_{i2,j2},\ A_{i1,j1}) = I(A_{i2,j2};\ A_{i1,j1})$$

or

$$I(A_{i1,j1};\ A_{i2,j2}) = I(A_{i1,j1}) - I(A_{i1,j1}|A_{i2,j2}) =$$

$$= I(A_{i2,j2}) - I(A_{i2,j2}\mid A_{i1,j1}) = I(A_{i2,j2};\ A_{i1,j1}).$$

The following example is illustrated these formulas.

Example 2. For a small training set (tab.1) used (1)-(4) obtain:

$I(\mathit{Sunny})=$
$= \log(M(\mathit{Outlook})/M(\mathit{Sunny}))=\log(16/6,6)=1,278;$

$I(V) = \log(M(\mathit{Game})/M(V))=\log(16/4,52)=1,824;$

$I(V,\ \mathit{Sunny})=$
$=\log(M(\mathit{Outlook})/M(V\times \mathit{Sunny}))=\log(16/0,11)=3,141;$

$I(V|\ \mathit{Sunny})= I(V,\ \mathit{Sunny})-I(\mathit{Sunny})= 1,824-1,278= 0,546;$

$I(V;\ \mathit{Sunny}) = I(V) + I(\mathit{Sunny}) - I(V,\mathit{Sunny}) =$
$1,824 + 1,278 - 3,141 = -0,038.$

4.2. Entropy and Conditional Entropy

Let's generalize previous evaluations for sets of the values of attribute. In other words we shall calculate their expectation. For this purpose we shall generalize concept information to concept entropy. Consider the using these estimations for fuzzy subsets.

In communication engineering, the entropy of A_{i1} represents the minimum expected number of bits (if we used the base 2 logarithm) required to transmit a value of A_{i1}. The entropy reaches its maximum value, when A_{i1} is uniformly distributed in its domain. Entropy is different from variance by its metric-free nature: it is dependent only on the probability distribution of a random variable and not on its values. This makes the entropy a universal measure of valued and variables.

Definition 6. The entropy of a attribute A_i is

$$H(A_i) = \sum_{j=1}^{m_i} M(A_{i,j})/M(A_i) \times I(A_{i,j}), \tag{5}$$

where $I(A_{i,j})$ see in Eq.(1).

Entropy of output attribute is an average amount of information, which should be extracted for determination of this output attribute. Entropy of input attribute is an average amount of information, which is extracted when we have detected value of this attribute.

Definition 7. The joint entropy of attributes A_{i2} and A_{i1} is

$$H(A_{i1},\ A_{i2}) = \tag{6}$$

$$= \sum_{j1=1}^{m_{i1}} \sum_{j2=1}^{m_{i2}} M(A_{i1,j1}\times A_{i2,j2})/M(A_{i1})\times I(A_{i1,j1}A_{i2,j2}),$$

where $I(A_{i1,j1}A_{i2,j2})$ see in Eq.(2).

Definition 8. The conditional entropy of attribute A_{i2} given another attribute A_{i1} is

$$H(A_{i2}|A_{i1}) = H(A_{i1},\ A_{i2}) - H(A_{i1}). \tag{7}$$

This entropy is average uncertainty of values of attribute A_{i2} given attribute A_{i1}. The conditional entropy of attribute A_{i2} given fuzzy subset $A_{i1,j1}$ of attribute A_{i1} is

$$H(A_{i2}\mid A_{i1,j1}) = \tag{8}$$

$$= \sum_{j2=1}^{m_{i2}} M(A_{i1,j1}\times A_{i2,j2})/M(A_{i1})\times M(A_{i1,j1})\times I(A_{i2,j2}|A_{i1,j1}),$$

where $I(A_{i1,j1}|A_{i2,j2})$ see in Eq.(3).

The association between two fuzzy subset $A_{i1,j1}$ of attribute A_{i1} and fuzzy subset $A_{i2,j2}$ of attribute A_{i2} (the mutual information) is defined by the Information theory as a decrease in entropy of $A_{i2,j2}$ as a result of knowing $A_{i1,j1}$ (and vice versa), namely the uncertainty of value of fuzzy subset $A_{i1,j1}$ of attributes A_{i2} when value of attributes A_{i1} is given.

Definition 9. The mutual information between attributes A_{i2} and A_{i1}

$$I(A_{i1};\ A_{i2}) = H(A_{i2}) - H(A_{i2}|\ A_{i1}) = \tag{9}$$

$$= H(A_{i1})-H(A_{i1}|\ A_{i2})=I(A_{i2};\ A_{i1})$$

or $I(A_{i1};\ A_{i2}) = H(A_{i2}) + H(A_{i1}) - H(A_{i1},A_{i2})=$

$$\sum_{j1=1}^{m_{i1}} \sum_{j2=1}^{m_{i2}} M(A_{i1,j1}\times A_{i2,j2})/M(A_{i1})\times I(A_{i2,j2};\ A_{i1,j1}),$$

where $I(A_{i1,j1};\ A_{i2,j2})$ see in Eq.(4).

Mutual information in attribute A_{i1} about attribute A_{i2} and vice versa, that reflects the influence of attribute A_{i1} on the attribute A_{i2} and conversely, the influence of attribute A_{i2} on attribute A_{i1}.

The difference between mutual information and correlation coefficient resembles the difference between entropy and variance: the mutual information is a metric-free measure, while the correlation coefficient measures a degree of functional (e.g., linear) dependency between values of attributes.

***Example* 3**. For a small training set (tab.1) used (5)-(9) obtain:

$$H(Outlook) = M(Sunny)/M(Outlook) \times I(Sunny) + M(Cloudy)/M(Outlook) \times I(Cloudy) + M(Rain)/M(Outlook) \times I(Rain) = 1{,}5419$$

$$H(Game) = M(V)/M(Game) \times I(V) + M(S)/M(Game) \times I(S) + M(W)/M(Game) \times I(W) = 1{,}5424$$

$$H(Game, Outlook) = \sum_{A_{i1,j1} \in Game} \sum_{A_{i2,j2} \in Outlook} M(A_{i1,j1} \times A_{i2,j2})/M(A_{i1}) \times I(A_{i1,j1}, A_{i2,j2}) = 2{,}898$$

$$H(Game|Outlook) = H(Game, Outlook) - I(Outlook) = 2{,}898 - 1{,}5419 = 1{,}356;$$

$$H(Game|Sunny) = \sum_{A_{i1,j1} \in Game} M(A_{i1,j1} \times A_{i2,j2})/M(A_{i1}) \times I(A_{i1,j1}, A_{i2,j2}) = 0{,}211 + 0{,}212 + 0{,}209 = 0{,}632$$

$$I(Game; Outlook) = H(Game) - H(Game|Outlook) = 1{,}5424 - 1{,}356 = 0{,}186.$$

In more detail see these results in tabl.2.

Table 2. The condition information and condition entropy

	I (*Game* | A_i)			H (*Game* | A_i)			
Game	*V*	*S*	*W*	*V*	*S*	*W*	
A_1 = OutLoook							
Sunny	1,863	1,111	1,932	0,211	0,212	0,209	Σ0,632
Cloudy	1,318	2,545	1,226	0,192	0,158	0,190	
Rain	3,244	3,968	0,268	0,077	0,057	0,050	
					H(*Game* | *Outlook*) =		1,356
A_2 = Temperature							
Hot	1,836	**0,926**	2,370	0,196	0,186	0,175	Σ0,557
Mild	1,408	2,705	**1,090**	0,186	0,145	0,179	Σ0,510
Cool	2,620	4,356	**0,343**	0,114	0,057	0,073	Σ0,244
					H(*Game* | *Temperature*) =		**1,311**
A_3 = Humidity							
Mild	2,247	1,560	1,152	0,166	0,185	0,181	
Cool	1,639	2,128	1,152	0,342	0,317	0,337	
					H(*Game* | *Humidity*) =		1,528
A_4 = Windy							
Windy	2,966	2,128	0,637	0,154	0,198	0,166	
NotWindy	1,365	1,767	1,654	0,315	0,308	0,312	
					H(*Game* | *Windy*) =		1,453

5. The Algorithm of the Fuzzy Decision Tree Design

There are two points in the process of constructing fuzzy decision trees. One is the selection of expanded attributes. They are such attributes that according to values of attributes (which are fuzzy subsets) trees are expanded at the nodes considered. The other is the judgment on leaves. Nodes are usually regarded as leaves if the relative frequency of one class is greater than or equal to a given threshold value.

A general learning algorithm for construction fuzzy decision trees can be described as follows.

Input Data: The small training set *R*. Where $A_i \in I$ (i=1,…,n), $A_k \in O$.

Output data: Fuzzy decision trees.

FDT-Algorithm (*R*)

{ ATTRIB={A_1,…,A_n};
  DoWhile ATTRIB ≠ ∅
  { Calculate $H(A_k|A_i)$ for $\forall\, A_i \in I$ using Eq. (7);
    Select the attribute A_i according to information criterion - minimal entropy $H(A_k|A_i) \rightarrow \min$;
    Assign chosen attribute A_i to current tree node of the level fuzzy decision trees;
    Calculate values of output attributes: $\min_{j=1}(I(A_{2,j})$ for $\forall$ fuzzy subset of A_k);
    ATTRIB = ATTRIB \ A_i;
  }
}

6. Application Example

Let us explain the technique of computations by the example below.

***Example* 4**. *Input Data:* The small training set, contains 4 input attributes {*Outlook*, *Temperature*, *Humidity*, *Wind*} (n=4) and 1 output attribute *Game* (tabl. 1).

Output data: Fuzzy decision trees

FDT-Algorithm

Step 1. Calculate the condition information and condition entropy between input attributes and the output attribute, in accordance with Eqs.(5-9), see tabl. 2.

Step 2. Choose the attribute A_i with minimal entropy:

$\min_{i=1,\dots,4}(H(Game|A_i)=\min(1{,}36;1{,}31;1{,}52;1{,}45)=1{,}31$

$H(Game|Temperature) = \min$).

We have to choise A_2 = *Temperature*.

Step 3. Assign chosen attribute A_2=*Temperature* to current tree node of the fuzzy decision tree.

Step 4. Calculate values of output attributes for current level.

$\min_{j=1}(I(A_{2,j})_{\text{for } \forall \text{ fuzzy subset of } A_k}) = \{0{,}926;1{,}090;0{,}343\};$

$Val_{Tempereature=Hot}(S) = 2^{-I(S|Hot)} = 2^{-0{,}926} = 0{,}526;$
$Val_{Tempereature=Mild}(W) = 2^{-I(W|Mild)} = 2^{-1{,}090} = 0{,}470;$
$Val_{Tempereature=Cool}(W) = 2^{-I(W|Cool)} = 2^{-0343} = 0{,}782;$

Build the first level of fuzzy decision tree (see fig.2)

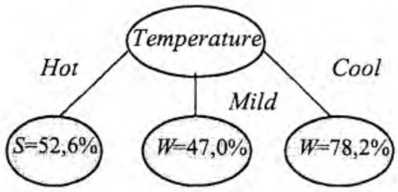

Figure 2. First level of fuzzy decision tree

Step 5. Repeat *Step*.1 - *Step*.4 for another attributes {*Outlook*, *Temperature*, *Humidity* and *Wind*}. The results are given in tab's.3-5 and fig.3.
End

Table 3. The condition information and condition entropy for the attribute *Temperature* = *Hot*.

	I(*Game*|*Hot*,A*i*)			H(*Game*|*Hot*,A*i*)			
Game	*V*	*S*	*W*	*V*	*S*	*W*	
A1 = *Outlook*							
Sunny	1,830	**0,680**	3,408	0,123	0,101	0,077	**Σ0,301**
Cloudy	1,365	**1,025**	3,051	0,046	0,043	0,032	**Σ0,121**
Rain	3,184	2,614	**0,461**	0,020	0,024	0,019	**Σ0,063**
H(*Game* | *Temperature* = *Hot*, *Outlook*) =							**Σ0,484**
A3 = *Humidity*							
Mild	2,383	0,682	2,435	0,083	0,077	0,082	
Cool	1,471	1,191	2,313	0,106	0,104	0,093	
H(*Game*| *Temperature* = *Hot*, *Humidity*) =							0,545
A4 = Windy							
Windy	2,623	0,916	1,701	0,056	0,064	0,069	
Not_windy	1,548	0,931	2,905	0,1324	0,122	0,097	
H(*Game*| *Temperature* = *Hot*, *Windy*) =							0,540

Table 4. The condition information and condition entropy for the attribute *Temperature* = *Mild*

	I(*Game*|*Mild*,A*i*)			H(*Game*|*Mild*,A*i*)			
Game	*V*	*S*	*W*	*V*	*S*	*W*	
A1 = *Outlook*							
Sunny	1,539	2,160	1,211	0,071	0,065	0,070	
Cloudy	1,238	3,179	1,103	0,087	0,059	0,086	
Rain	1,706	3,232	0,768	0,025	0,017	0,022	
H(*Game*| *Temperature* = *Mild*, *Outlook*) =							0,502
A3 = *Humidity*							
Mild	1,322	3,582	0,953	0,034	0,019	0,032	
Cool	1,428	2,564	1,123	0,152	0,124	0,147	
H(*Game*| *Temperature* = *Mild*, *Humidity*) =							0,508
A4 = *Windy*							
Windy	2,534	3,760	**0,408**	0,069	0,044	0,048	**Σ0,161**
Not_windy	**0,878**	2,194	2,075	0,092	0,092	0,095	**Σ0,279**
H(*Game*| *Temperature* = *Mild*, *Windy*) =							**Σ0,440**

Table 5. The condition information and condition entropy for the attribute *Temperature*=*Cool*

	I(*Game*|*Cool*,A*i*)			H(*Game*|*Cool*,A*i*)			
Game	*V*	*S*	*W*	*V*	*S*	*W*	
A1 = *Outlook*							
Sunny	2,619	1,737	**0,515**	0	0,021	0,014	**Σ0,035**
Cloudy	1,411	6,381	**0,708**	0,058	0,008	0,047	**Σ0,114**
Rain	5,515	8,404	**0,032**	0,014	0	0,004	**Σ0,018**
H(*Game*| *Temperature* = *Cool*, *Outlook*) =							**Σ0,167**
A3 = *Humidity*							
Mild	2,983	**4,356**	0,195	0,0392	0	0,018	
Cool	2,430	3,652	0,444	0,074	0,048	0,054	
H(*Game*| *Temperature* = *Cool*, *Humidity*) =							0,508
A4 = Windy							
Windy	5,069	3,314	0,201	0,018	0,039	0,021	
Not_windy	1,910	6,848	0,463	0,077	0,009	0,050	
H(*Game*| *Temperature* = *Cool*, *Windy*) =							0,214

A fuzzy decision tree constructed by using this algorithm is shown as fig. 3.

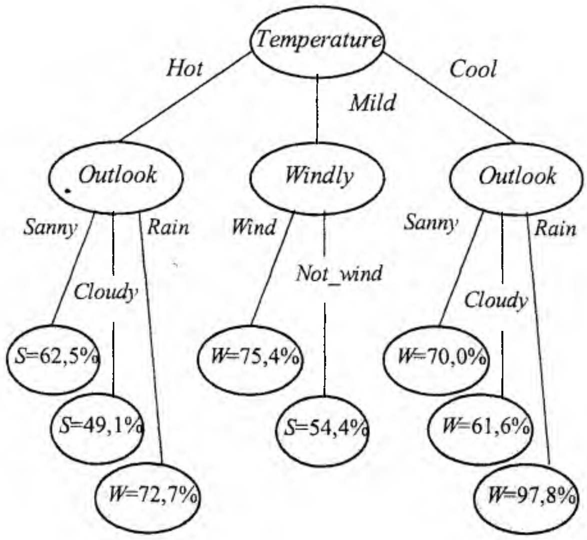

Figure 3.Construction fuzzy decision trees

CONCLUSION

This paper investigates the buildind of fuzzy decision trees. The actual task for existing algorithms for constructing fuzzy decision trees is the correct selection of attributes. One of the ways is attempt to obtain a small-scale tree via the expanded attribute selection and to improve the classification accuracy for unknown cases. It is possible that the reduction of decision tree scale results in the improvement of the classification accuracy for unknown cases. An important problem is whether or not there exists an exact algorithm for constructing the smallest-scale fuzzy decision tree.

Centering on this optimal problem, this paper discusses the representation of fuzzy decision trees and presents an algorithm for constructing fuzzy decision trees and shows the its accuracy.

The interest on using information theory methods was paid comparatively recently to solve problems of real world description. The information approach makes wider our understanding, but its application is connected with some difficulties. This approach seems to be promising to solve the problem of comparison of optimization algorithms for logic functions by evaluating each of the strategies through information estimations. An unique possibility of the information approach to form a *prognosis* for the searching algorithms is not studied yet. Applying this property can strongly affect on the searching algorithms.

In this paper we systematically explain the ways to adapt and interpret the information theory methods to construct fuzzy decision trees. We illustrate various algorithmic, computational aspects by a number of examples.

We stand on the position that a fuzzy set of attributes has got information, which can be estimated numerically. The methods of the information theory allow us for such estimations.

We have developed the ways to compute information estimations for fuzzy sets. We propose the technique to compute information estimations for fuzzy sets, which is simple to understand and apply. The approach outlined in this paper is a basis for further work.

7. References

[1]. R.M. Goodman, J.W. Miller and P. Smyth, An Information Theoretic Approach to Rule-based Connectionist Expert Systems, in: D.S. Touretzky (Ed.), *Advances in Neural Information Processing Systems I, Morgan Kaufmann, San Mateo*, CA, 1988, pp. 256-263.

[2]. Hartmann C.R.P., Varshney P.K., Mehrotra K.G., and Gerberich C.L., Application of Information Theory to the Construction of Efficient Decision Trees, *IEEE Trans. on Information Theory,* IT-**28**, 1982, pp.565-577.

[3]. Guyon, N. Matic and V. Vapnik, Discovering Informative Patterns and Data Cleaning, in: U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy (Eds.), *Advances in Knowledge Discovery and Data Mining, AAAI=MIT Press, Menlo Park*, CA, 1996, pp.18-203.

[4]. O. Maimon, A. Kandel and M. Last, Fuzzy Approach to Data Reliability, in: O. Pons, A. Vila, J. Kacprzyk (Eds.), *Knowledge Management in Fuzzy Databases, Springer*, Berlin, 1999.

[5]. J.R. Quinlan, Induction of Decision Trees, *Machine Learning*, **1**, 1986, pp. 81-106.

[6]. S.R. Safavian and D. Landgrebe, A Survey of Decision Tree Classifier Methodology, *IEEE Trans.Sys,Man Cybernet.* **21,** 1991, pp.660-674.

[7]. K.J. Cios and L.M. Sztandera, Continuous ID3 Algorithm with Fuzzy Entropy Measures, *Proc. IEEE Internat. Conf. on Fuzzy Systems*, San Diego, CA, 1992, pp. 469-476.

[8]. H.Ichihashi et al., Neuro-Fuzzy ID3, *Fuzzy Sets and Systems*, **81**, 1996, pp.151-167.

[9]. T.Tani and M. Sakoda, Fuzzy Modeling by ID3 Algorithm and Its Application to Prediction, Proc. *IEEE Internat. Conf. on Fuzzy Systems*, San Diego, CA, 1992, pp. 923-930.

[10]. R.L.P.Chang and T. Pavliddis, Fuzzy Decision Tree Algorithms, *IEEE Transactions on Systems, Man and Cybernetics*, 7, 1977, pp. 28-35.

[11]. R.Weber, Fuzzy-ID3: a Class of Methods for Automatic Knowledge Acquisition, *Proc. 2nd Int.Conf. on Fuzzy Logic and Neural Networks,* Iizuka, Japan, 1992, pp. 265-268.

[12]. Y. Yuan and M.J. Shaw, Induction of Fuzzy Decision Trees, *Fuzzy Sets and Systems*, **69**. 1995, pp.125-139.

[13]. Xizhao Wang, Bin Chen, Guoliang Qian. Feng Ye, On the optimization of Fuzzy Decision Trees, *Fuzzy Sets and Systems*, **112**. 2000, pp.117-125.

[14]. T.M. Cover, Elements of Information Theory. *Wiley*, New York, 1991.

[15]. D.A.Simovici, V.Shmerko, V.Cheushev and S.Yanushkevich, Information Estimation for Logic Functions, *Proc. Int. Conf. on Applications of Computer Systems,* Szczecin. Poland, 1997, pp.287-300.

[16]. A.Kurbatsky and V.Cheushev, Infornation Approach for Analysis and Optomization in Decission Making Support Systems. *Institute of Engineering Cybernetics*, Minsk, 1999, (In Russian).