UDC 004.614.2

Vitaly LEVASHENKO [1]**,** Professor,
Patrik HRKUT [1]**,** Assoc. Professor,
Andriy KOVALENKO [2]**,** Assoc. Professor
Lyliya KURMASHEVA [3]**,** PhD student
[1] University of Žilina, Univerzitná 8215/1, 010 26 Žilina, Slovakia, e-mail: vitaly.levashenko @fri.uniza.sk
[2] Kharkov State Technical University of Radio electronics, Nauky Ave. 14, Kharkov, Ukraine,
e-mail: andriy.kovalenko@nure.ua
[3] Manash Kozybayev North Kazakhstan university, Pushkin street, 86, Petropavlovsk, Kazakhstan,
e-mail: kharytonova-lv@ukr.net

FUZZY DECISION TREES IN MEDICAL DECISION MAKING SUPPORT SYSTEM

Abstract. Decision Making Support System is used widely in medicine now because decisions play an important role in medicine, especially in medical diagnostic processes Decision Making Support Systems helping physicians are becoming a very important part in medical decision making, particularly in those situations where decision must be made effectively and reliably. Since conceptual simple decision making models with the possibility of automatic learning should be considered for performing such tasks, decision trees are a very suitable candidate. In this paper Fuzzy Decision Trees are proposed for the application in Medical Decision Making Support System. Induction of these Fuzzy Decision Trees is based on cumulative information estimates. Comparisons with different methods show it is a promising solution.
Key words: Machine Learning, Fuzzy Decision Trees, Medicine.

Introduction

Information technology is of vital importance in the realm of medicine. As a consequence of aging population and an increasing morbidity there are more and more patients with different diseases. That leads to a lack of professionals who can treat them and to escalating costs. An interesting solution appears special devices with an intelligent clinical Decision Making Support System (DMSS) that can be used by non-professionals [1, 2, 3]. Such devices can indicate preventive diagnosis with next professional support by medical personal. The clue to successful treatment of a lot of diseases is correct diagnosis, which is impossible without the appropriate interpretation of the entire spectrum of diagnostic facts including anamnesis, physiological parameters, and diagnostic imaging. Data mining is a process of extracting implicit, potential, novel, useful and intelligible patterns from mass data of data sets, databases or data warehouse, etc. The technologies of classification, estimation, prediction, affinity grouping, association rules, clustering, description and visualization are covered in data mining, which is widely used in the fields of medicine [4]. Therefore, methods of Data Mining are useful and applicable for design medical DMSS (MDMSS).

Since conceptual simple decision making models with the possibility of automatic learning should be considered for performing such tasks, decision trees are a very suitable candidate. They have been already successfully used for many decision making purposes [5].

A decision tree is a graphic model of a decision process, and it is usually used as a decision support tool or classifier. A decision trees is one of the best ways to analyze a decision, as it is visualized and simple to understand and interpret. It’s possible consequence includes chance event outcomes, resource costs or utility. But in real application and first of all in medicine the input data for the analysis isn’t defined exact and has some ambiguity. The data cleaning causes the sacrifice of useful information in this case [3]. The development of methods based on fuzzy data application is actual problem in the design of MDMSS [6].

In this paper we propose to investigate MDMSS based on Fuzzy Logic and Fuzzy Decision Trees (FDT), as an efficient alternative to crisp classifiers that are applied independently. This cooperation tries to soften the accuracy/interpretability tradeoff. Many FDT induction algorithms have been introduced. There are different medical applications of FDT for building of rules for classification [7]. Authors in paper [7] use fuzzy ID3 algorithm, that doesn't allow FDT building with parallel structure. FDT in [5] is satisfactory for completely specified initial data. In [8] the ordered FDT have been proposed that permit to find a sequence of rules, which analyze input attributes in order that is both cost effective and guarantees a desired level of accuracy. Every node of one level of such FDT associates with similar attribute. In this paper authors have been used a special cumulative information estimates of fuzzy sets. The applications of these estimations in algorithms for FDT induction permit to construct trees with different properties [8, 9, 10].

In this paper we develop the application of cumulative information estimations for FDT induction. These FDT are used for the analysis and the classification of medical data. The classification result based on these FDT is compared with other methods of Data Mining (as C4.5, CART, naïve Bayes classifiers and *k*-nearest neighbor). The experimental investigation and comparison have been implemented for 17 typical data sets from UCI Machine Learning Repository. This comparison shown that the FDT induction based on cumulative information estimations outperformed most other methods of Data Mining. This provides some justification for using this type of FDT in a DMSS.

The paper is organized as follows. Section 2 contains brief information about used MDMSS and representation of fuzzy data. Section 3 shows FDT and Fuzzy classification rules under a simple example. Section 4 demonstrates the experimental results of proposed FDT.

Decision making support system based on fuzzy logic

1. *Decision making support system*

MDMSS are systems that used in medicine for the tasks of diagnosis, prognosis, treatment planning and decision support. Such system creates an information environment, supported by sophisticated and robust search, optimization, and matching techniques for heterogeneous information (images and clinical data) in form of electronic health records. MDMSSs have been implemented for specific areas in medicine or diseases. Some of these systems have similar conception and based on the identical mathematical background. We use conception with comparison of new case with previous cases and selection most similar as decision (Fig. 1). Thus the classification is principal problem of this conception based on special rules that agrees with Block of Compare new case and ontology. The mathematical background of this block is Fuzzy Classification rules that are formed by FDT. The block for Preparation of initial data implements transformation of the input data to the fuzzy data. This procedure is named as fuzzification. The result presentation is interpretation of the decision by the de-fuzzification procedure.

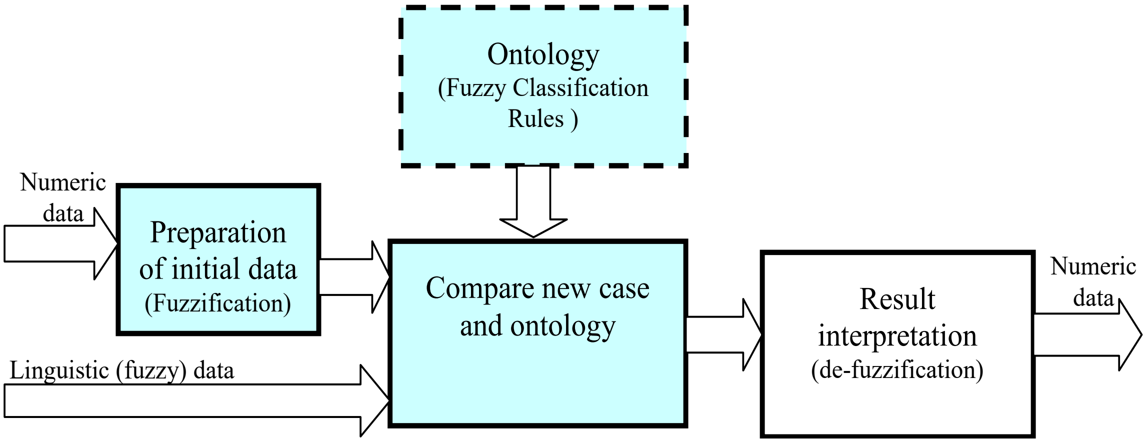

Fig. 1. Decision Making Support System

The decision making procedure corresponds to the recognition (classification) of the new case and is the process of moving from concrete examples to general models, where the goal is to learn how to classify objects by analyzing a set of instances (already solved cases) whose classes are known. Instances are typically represented as attribute-value vectors. One of possible solutions for such classification is implemented by Decision Trees. A decision tree is formalism for expressing such mappings and consists of tests of attribute nodes linked to two or more sub-trees and leafs or decision nodes labeled with a class which means the decision. A test node computes some outcome based on the attribute values of an instance, where each possible outcome is associated with one of the sub-trees. An instance is classified by starting at the root node of the tree. If this node is a test, the outcome for the instance is determined and the process continues using the appropriate subtree. When a leaf is eventually encountered, its label gives the predicted class of the instance. The FDT is one of possible types of decision trees that permits to operate by fuzzy data (attributes).

The process of construction of FDT is based on the use of a fuzzy partition for each numerical attribute. An automatic method of construction of such a partition from a set of precise values could be used in order to obtain automatically a set of fuzzy values for each numerical attribute. Fuzzy data are used in situations that are especially difficult or ambiguous, and unsolvable by other types of logic. Fuzzification transforms precise input into corresponding fuzzy input [12]. The interpretability of a fuzzy system – especially if applied in data analysis – is one of its key advantages.

Therefore, the considered conception of MDMSS (Fig. 1) can be implemented based on the fuzzy classification rules.

2. *Fuzzy Logic*

Fuzzy logic is a popular approach to capture vagueness of information. The basic idea is to use instead the "crisp" 1 and 0 values the values of the interval [0, 1] indicating a degree of truth or confidence.

A fuzzy set F with respect to a universe U is characterized by a membership function μ_F: $U \to [0.1]$, which assign a F-membership degree, $\mu_F(u)$, to each element u in U. $\mu_F(u)$ gives an estimation that u belongs to the fuzzy set F [11].

For example, consider attribute A_i that is *Age*. This attribute has 3 fuzzy partitions $A_{i,1}$ (*young*), $A_{i,2}$ (*adult*), $A_{i,3}$ (*old*) (with range [0, 1]) as it is depicted in Fig. 2. The real value $u \in U$ of this attribute A_i is interpreted as: $\mu_{young}(u) = 0.7$, $\mu_{adult}(u) = 0.3$, and $\mu_{old}(u) = 0$.

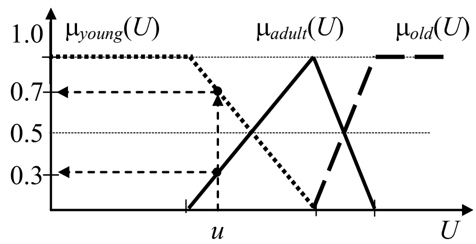

Fig. 2. Fuzzy membership functions of an attribute A_i (*Age*)

Thus, the fuzzification of the initial data is performed by analyzing the corresponding values of a membership function. Here, each attribute value can be seen as likelihood estimate. In this paper we analyze a particular case when the sum of membership values of all partitions equals to 1. For these purposes, we use one of the algorithms to transform from numeric to triangular fuzzy data, presented in [12].

A typical classification problem can be described as follows [13]. A universe of objects $U=\{u\}$ is described by N training examples and n input attributes $A=\{A_1, ..., A_n\}$. Each attribute A_i $(1 \le i \le n)$ measures some feature presented by a group of discrete linguistic terms. We assume that each group is a set of m_i $(m_i \ge 2)$ values of fuzzy subsets $\{A_{i,1}, \dots, A_{i,j}, \dots, A_{i,m_i}\}$. We assume that each object u in the universe is classified by a set of classes $\{B_1, ..., B_{m_b}\}$. This set describes the class attribute B. The class attribute B has to determine by values of attributes A_i.

Let us consider the simplified example. In this example we use only four input attributes: A_1, A_2, A_3, A_4 and one output (class) attribute B [13]. Each attribute has the 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 $B = \{B_1, B_2, B_3\}$. Let the instances be the ones presented in Table 1. Let the costs of different attributes be the ones on the lowest row of this Table 1. The cost of input attribute A_i denoted as $Cost(A_i)$ is a priory given value. This value describes time and other cost required to determine the value of this attribute in during the classification of new instance. Our goal is find a method for transform values of input attributes into the value of output attribute with minimal resources: *sum Cost* $(A_i) \to$ *minimum.*

Table 1

A training set

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				

Fuzzy decision trees induction

There are different approaches to induct FDT [13, 14, 15]. The principal goal of these approaches for FDT induction is selection of expanded attributes and determination of the leaf node. The key points of approaches for induction of FDT are (*a*) a heuristic for selecting expanded attributes and (*b*) a rule for transform nodes into leaves. An expanded attribute is an attribute that according to the values of the attribute tree expands the node considered. The cumulative information estimates allow defining criterion of expanded attributes selection to induct FDT with different properties. These FDT were detail considered in [8].

The selection criterion of expanded attributes A_{i_q} for induction of non-ordered FDT is defined as:

$$i_q = \arg\max \frac{\mathbf{I}(B; A_{i_1,j_1}, \ldots, A_{i_{q-1},j_{q-1}}, A_{i_q})}{\mathrm{Cost}(A_{i_q})}, \tag{1}$$

where $A_{i_1,j_1},\ldots,A_{i_{q-1},j_{q-1}}$ are values of input attributes $A_{i_1},\ldots,A_{i_{q-1}}$ of path from root node to examined attribute; A_{i_q} is the attribute that isn't in this path.

Maximum value of cumulative mutual information (1) allows to select expanded attributes A_{i_q} between other attributes.

There are two tuning parameters α and β used in the algorithm [8]. Expanding a tree branch is stopped when either the frequency *f* of the branch is below α or when more than β percent of instances left in the branch has the same class label. Thus these values are key parameters for deciding have we already approached to leaf node or should we need to continue expanding the branch.

The Non-ordered FDT inducted according (1) for data in Table 1 ($\beta = 0.75$ and $\alpha = 0.16$) is in Fig. 3. The transformation from entropy into information shown at this figure. The sum of cumulative conditional entropy and cumulative mutual information is constant. It is true for each branches of FDT. This is fully consistent with the law of conservation of information.

In the all type of FDT, each non-leaf node is associated with an attribute $A_i \in A$. 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 $L = \{l_1, ..., l_r, ..., l_R\}$. There is also a vector of values $\mathbf{F}^r = [F_1{}^r;..., F_{j_b}{}^r;..., F_{m_b}{}^r]$ for each r-th leaf l and each j_b-th class B_{j_b}. Each value $F_{j_b}{}^r$ means the certainty degree of the class B_{j_b} attached to the leaf node l_r.

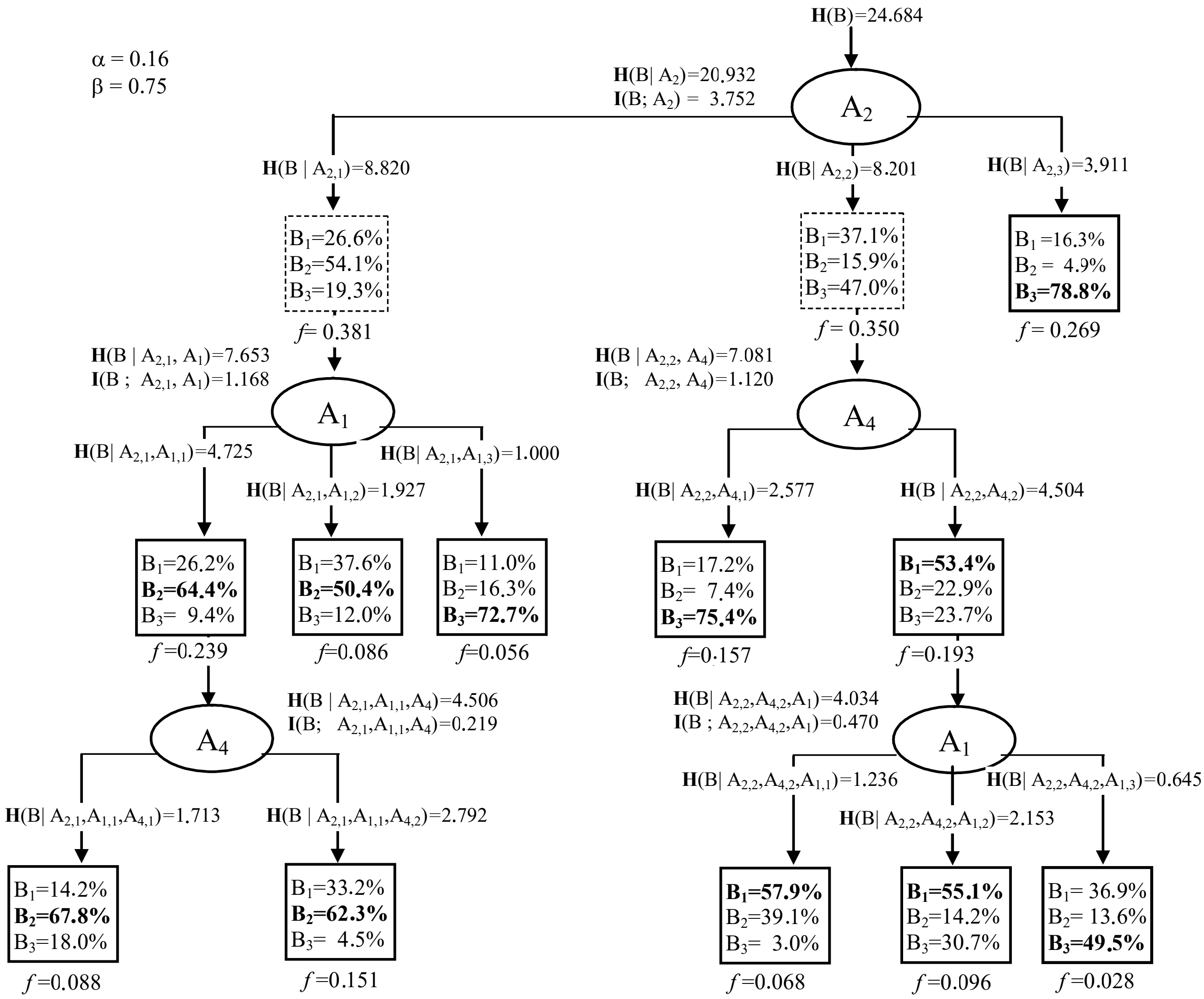

Fig. 3. Non-ordered FDT inducted for the data in Table 1

In fuzzy cases, a new instance e may be classified into different classes with different degrees. Then, each leaf $l_r \in L$ corresponds to one (r-th) classification rule. The condition part of the classification rule is a group of conditions presented in theform "attribute is attribute's value" and those conditions 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 the values of class attribute B with their truthfulness vector $\mathbf{F}^r$ values.

Let's consider the path $P_r(e) = \{[A_{i_1,j_1}(e)]^r, \ldots, [A_{i_s,j_s}(e)]^r, \ldots, [A_{i_S,j_S}(e)]^r\}$ from the FDT root to the r-th leaf. This path $P_r(e)$ consist of S nodes which are associated with attributes $A_{i_1},.., A_{i_s},\ldots, A_{i_S}$ and respectively their S outgoing branches associated with the values $A_{i_1,j_1},.., A_{i_s,j_s},\ldots, A_{i_S,j_S}$. Then the r-th rule has the following form:

$$\text{IF } (A_{i_1} \text{ is } A_{i_1,j_1}) \textit{ and } \ldots \textit{ and } (A_{i_S} \text{ is } A_{i_S,j_S}) \text{ THEN B (with truthfulness } \mathbf{F}^r).$$

Our approach uses several classification rules for classification of a new instance e. That's why, there may be several paths whose all outgoing node's branches are associated with values

$A_{i_S,j_S}(e)$ greater 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 $W_r(e)$. The weight is for instance e and the r-th rule is given by the rule $W_r(e) = \prod_{s=1}^{S} [A_{i_s,j_s}(e)]^r$, where $[A_{iS,jS}(e)]^r$ is the value of the attribute $A_{iS,}$ for the new instance e. The weight $W_r(e)$ is equal 0 if there is a attribute's value A_{i_S,j_S} whose membership function equals 0. Values of class attribute B for the new instance e are:

$$\mu_B(e) = \sum_{r=1}^{R} W_r(e) \times \mathbf{F}^r , \qquad (2)$$

where $\mathbf{F}^r$ is the truthfulness of the r-th rule.

Below the transformation process of the FDT into fuzzy rules and these rules are used for classification are described by example for the non-ordered FDT in Fig.3. The FDT in Fig. 3 has $R = 9$ leaves. Let a new instance e have following attribute values: $A_1 = \{A_{1,1}; A_{1,2}; A_{1,3}\} = \{0.9; 0.1; 0.0\}$, $A_2 = \{A_{2,1}; A_{2,2}; A_{2,3}\} = \{1.0; 0.0; 0.0\}$, $A_3 = \{A_{3,1}; A_{3,2}\} = \{0.8; 0.2\}$ 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 the 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 with $\mathbf{F}^1 = [0.142; 0.678; 0.180]$;
$r = 2$: IF A_2 is $A_{2,1}$ *and* A_1 is $A_{1,1}$ *and* A_4 is $A_{4,2}$ THEN B with $\mathbf{F}^2 = [0.332; 0.623; 0.045]$;
$r = 3$: IF A_2 is $A_{2,1}$ *and* A_1 is $A_{1,2}$ THEN B with $\mathbf{F}^3 = [0.376; 0.504; 0.120]$;
...
$r = 9$: IF A_2 is $A_{2,3}$ THEN B with $\mathbf{F}^9 = [0.163; 0.049; 0.788]$.

The weights $W_r(e)$ ($r = 1, \ldots, 9$) are: $W_1(e) = 1.0 \times 0.9 \times 0.4 = 0.36$, $W_2(e) = 1.0 \times 0.9 \times 0.6 = 0.54$, $W_3(e) = 1.0 \times 0.1 = 0.10$ and all the other $W_r(e)$ are equal 0.

We obtain with (2) for this FDT:

$$\mu_{B_1}(e) = 0.1142 \times 0.36 + 0.678 \times 0.54 + 0.180 \times 0.1 = 0.268;$$

Similarly, $$\mu_{B_2}(e) = 0.332 \times 0.36 + 0.623 \times 0.54 + 0.045 \times 0.1 = 0.631;$$

$$\mu_{B_3}(e) = 0.376 \times 0.36 + 0.504 \times 0.54 + 0.120 \times 0.1 = 0.101.$$

The values of class attribute $B = \{B_1, B_2, B_3\} = \{0.268; 0.631; 0.101\}$ for the new 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.

Experimental Results

The main purpose of our experimental study is to compare proposed FDT with other well-known classification methods. All algorithms are coded in CPP. We used data for medical diagnosis to form classification rules based on non-ordered FDT only. The software for experiment includes 4 basic blocks (Fig. 4). The experiments have been carried out on UCI Machine Learning Repository benchmarks (dataset). We choose 17 medical datasets from this Repository. The main criterion of choice is type of class attribute. We have to choice datasets with discrete class attribute only.

We had divided initial dataset into 2 parts. The first part (70% from initial dataset) was used for building classification models. The second part (30% from initial dataset) was used for verification of the classification models. This process was repeated 1000 times, and average estimations were produced.

A description of datasets is shown in Table 2. Columns [Dataset], [TS], [NoIA] and [NoOA] describe initial parameters of datasets: Total sets, Number of initial attributes and Number of classes. The column labeled [iErrors] gives naïve initial error, when we choose the class value with maximum frequency.

A fragment of our results is shown in Table 3.

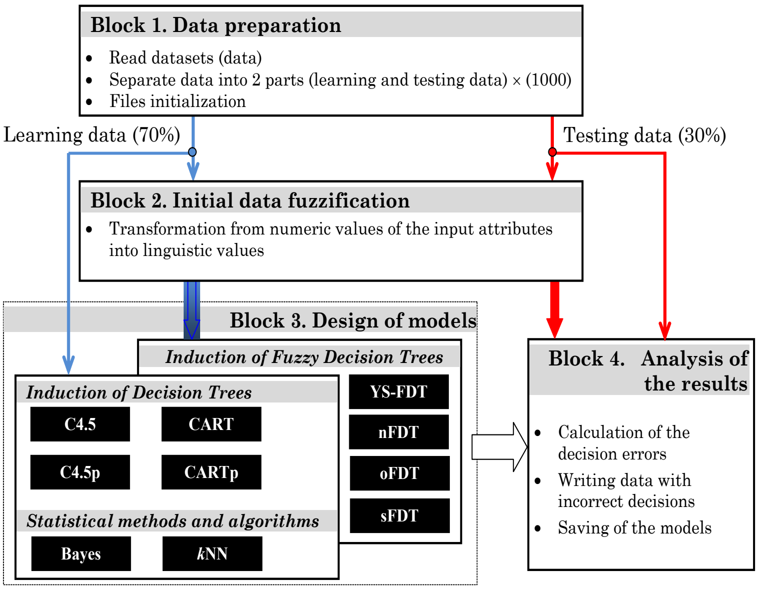

Fig. 4. The description of experiment

Table 2

Description of UCI machine learning benchmark set

Datasets	TS	NoIA	NoOI	iError
balance	625	4	3	0.539
blood	748	4	2	0.238
breast	106	9	6	0.793
bura	345	6	2	0.42
cmc	1473	9	3	0.573
diagnosis	120	6	2	0.417
ecoli	338	7	8	0.574
haberman	306	3	2	0/265
heart	270	132		0.444
ilpd	579	10	2	0.285
parkinsons	195	22	2	0.246
pima	768	8	2	0.349
thyroid	215	5	3	0.302
vertebral2	310	6	2	0.323
vertebral3	310	6	3	0.516
wdbc	569	30	2	0.373
wpbc	194	33	2	0.237

This table gives the value of error misclassification. It is calculated as the ratio of the number of misclassification combinations to the total number of combinations. The results in columns [nFDT], [oFDT], [sFDT] are according to fuzzy classification rules that have been formed based on non-ordered, ordered and stable FDT. Column [YS] describes results by implementation of well-known algorithm [13]. Columns [C4.5], [C4.5pr], [CART] and [CARTpr] consist of results of algorithms C4.5 and CART without and with application Error-Complexity Pruning method. Columns [Bayes] and [*k*NN] describe results of naïve Bayes classification and k-nearest neighbor methods.

Table 3

Results on the UCI machine learning benchmark set

Datasets	nFTD	oFTD	sFTD	YS	C4.5	C45pr	CART	CRpr	Bayes	kNN
balance	0.122	0.1221	0.1221	0.3282	0.2208	0.2213	0.2242	0.2364	**0.1029**	0.3769
blood	0.2363	0.2363	0.2363	0.2389	**0.2277**	**0.2263**	0.2328	**0.2281**	0.2493	0.3953
breast	0.4017	0.4016	0.4019	0.4454	**0.3548**	**0.359**	0.3704	**0.3567**	0.361	**0.3586**
bura	0.4105	0.4108	0.41	0.4224	0.3649	**0.3517**	0.3823	**0.3527**	0.4448	0.4339
cmc	0.5209	0.5374	0.5263	0.5392	0.5205	**0.47**	0.5366	0.4959	0.5063	0.6186
diagnosis	**0.0002**	**0**	0	0.1578	0.0071	0.0072	0.0072	0.0074	0.0476	**0**
ecoli	0.1816	0.1825	0.183	0.2408	0.4457	0.2918	0.1988	0.2036	**0.152**	0.279
haberman	0.2652	0.2652	0.2652	0.2462	0.2908	0.2749	0.2888	0.2625	**0.2537**	0.4077
heart	0.1852	0.165	0.1723	0.2575	0.2612	0.2443	0.2551	0.221	**0.1635**	0.3126
ilpd	**0.285**	**0.2851**	**0.2846**	0.2854	0.3232	**0.2837**	0.3462	**0.2828**	0.44`9	0.3991
parkinsons	0.0959	0.0959	0.0959	0.1469	0.1522	0.1498	0.14	0.1289	0.2993	**0.0839**
pima	0.2417	**0.2401**	0.2435	0.2527	0.2708	0.2547	0.3064	0.2544	0.2462	0.3585
thyroid	0.0739	0.0738	0.0738	0.1059	0.0736	0.078	0.9272	0.0955	**0.0339**	0.0568
vertebral2	**0.1714**	**0.1716**	**0.1723**	0.1818	0.1932	0.1916	0.2083	0.1892	0.2202	0.2846
vertebral3	0.1993	0.1995	0.1994	0.3027	0.195	0.1915	0.2056	0.1903	**0.1718**	0.3412
wdbc	**0.0368**	0.0368	0.0368	0.0705	0.0665	0.0641	0.0781	0.0704	0.0674	0.0719
wpbc	**0.221**	0.221	0.221	0.249	0.2888	0.24	0.3187	0.2334	0.3378	0.4269
Average	**0.2146**	**0.21439**	**0.2144**	0.2624	0.2504	**0.2294**	0.2469	**0.2241**	0.2416	0.3062

The best solution has minimal error of misclassification. This value is marked in bold. If the difference between the values is not more than 5%, we can be considered several values as the best.

Average error misclassification values of nFDT, oFDT and sFDT indicate the superiority of the proposed approach to the FDT induction. Note, that implementation of Error-Complexity Pruning method with algorithms C4.5 and CART give us good results also.

Conclusion

In many applications, black-box prediction is not satisfactory, and understanding and handling the data is of critical importance. Typically, approaches useful for understanding of data involve logical rules, evaluate similarity to prototypes, or are based on visualization or graphical methods.

There are several methods proposed for logical rule generation combining different data types (machine learning, fuzzy decision trees, association rules, Bayesian networks, neural networks, pattern recognition). We have selected the more powerful of these algorithms that have been proved from the literature that give better rules and keep the level of interpretability and accuracy in the classification task].

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 its semantics. We have proposed the technique to induction of new type of FDT, which is simple to understand and apply. The use of cumulative information estimations allows precisely estimating mutual influence of attributes. These evaluations are used to analyze group of training instances.

Acknowledgement

This publication has been produced with the support of the Integrated Infrastructure Operational Program for the project: *Systemic Public Research Infrastructure - Biobank for Cancer and Rare diseases*, ITMS: 313011AFG5, co-financed by the European Regional Development Fund and with the support of the Ministry of Education, Science, Research, and Sport of the Slovak Republic under the project “*Creation of methodological and study materials for Biomedical Informatics - a new Master program at the University of Žilina*” (reg.no. KEGA 009ŽU-4/2020).

References

1. Moumtzoglou A., Kastania A.: E-Health Systems Quality and Reliability: Models and Standards. IGI Global, 2011, 376 p.

2. Gomula J., Pancerz K., Szkola J.: Copernicus - an expert system supporting differential diagnosis of patients examined using the MMPI test: an index-rule approach, *Proc. of the Int. Conf. on Health Informatics* (HEALTHINF 2011), Rome, Italy, 2011, pp. 323–328.

3. Rodriguez-Gonzalez A., Torres-Nino J., Mayer M., Alor-Hernandez G., Wilkinson M.: Analysis of a multilevel diagnosis decision support system and its implications: a case study, *Comput. Math. Methods Med*., vol. 2012, 2012.

4. Podgorelec V., Kokol P., Stiglic B., Rozman I.: Decision trees: an overview and their use in medicine. *Journal of Medical Systems*, 26, 2002, pp. 445–463.

5. Marsala C.: A Fuzzy Decision Tree Based Approach to Characterize Medical Data. *Proc. of IEEE Int. Conf. on Fuzzy Systems*, Jeju Island, Korea, 2009, pp. 1332–1337.

6. Chalmers E., Pedrycz W., Lou E.: Predicting the outcome of brace treatment for scoliosis using conditional fuzzy clustering, *Proc. of the IFSA World Congress and NAFIPS Annual Meeting* (IFSA/NAFIPS), 2013, pp. 837–842.

7. Papageorgiou E.I., Papandrianos N.I., Apostolopoulos1 D., Vassilakos P.: Complementary use of Fuzzy Decision Trees and Augmented Fuzzy Cognitive Maps for Decision Making in Medical Informatics. Proc. of Int. Conf. on BioMedical Engineering and Informatics, Sanya, Hainan, China, 2008, pp.888-892.

8. Levashenko V., Zaitseva E.: Fuzzy decision trees in medical decision making support system, Proc. of the IEEE Federated Conf. on Computer Science & Information Systems, (IEEE FedCSIS‘12), September 9-12, Wrocław, Poland, 2012, pp. 213–219.

9. Rabcan J., Zaitseva E., Levashenko V., etc. Fuzzy decision tree based method in decision-making of COVID-19 patients’ treatment, *Mathematics*, 2021, 9, 24, 3282.

10 *IEEE Trans. on Industrial Informatics*, 2022, 18, 2, pp. 757–766.

11. Kaufmann A., Gupta. M.: Induction to fuzzy arithmetic: theory and applications, New York: Van Nostrand Reinold Co., 1985.

12. 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, 31(3), 2001, pp. 426–432.

13. Yuan Y., Shaw M.J.: Induction of Fuzzy Decision Trees. *Fuzzy Sets and Systems*, 69, 1995, pp. 125–139.

14. Olaru C., Whenkel L.: A Complete Fuzzy Decision Tree Technique. *Fuzzy Sets and Systems*, 2003, pp. 221–254.

15 rockett K., Bandar Z., O'Shea J.: Fuzzification of Discrete Attributes from Financial Data in Fuzzy Classification Trees. *Proc. of IEEE Int. Conf. on Fuzzy Systems*, Jeju Island, Korea, 2009, pp. 1320–1325.