FUZZY IF-THEN RULE INDUCTION WITH CUMULATIVE INFORMATION ESTIMATIONS APPLIED TO REAL-WORLD DATA

Ján BOHÁČIK*, Martti JUHOLA**, Vitaly LEVASHENKO*
*Department of Informatics, Faculty of Management Science and Informatics,
University of Žilina, Univerzitná 8215/1, 010 26 Žilina, tel. 041/513 4182, E-mail: Jan.Bohacik@fri.uniza.sk
**Department of Computer Sciences, Faculty of Information Sciences,
University of Tampere, Kanslerinrinne 1, 330 14 Tampere, tel. 03/35517972, E-mail: Martti.Juhola@cs.uta.fi

ABSTRACT

Real-world data containing instances corresponding to patients with otoneurological diseases were explored with fuzzy IF-THEN rule induction. It was based on transformation of a fuzzy decision tree made with using cumulative information estimations as the locally optimal criterion at its nodes. This method uses linguistic variables that allow us to naturally model various situations appearing in this data. It also gives classification knowledge in a form of IF-THEN rules that are easily readable and understandable by an expert. Our study shows in comparison to multilayer perception neural networks that classification with the induced fuzzy IF-THEN rules is a useful technique for diagnostics of otoneurological diseases.

Keywords: *fuzzy rules, cumulative information estimations, perception neural networks, classification, otoneurology.*

1. INTRODUCTION

In the real world it is needed to work with real-world data. Such data is not often accurate and contains some uncertainties. These uncertainties are caused by various reasons such as vagueness and ambiguity described in [6]. Vagueness is associated with the difficulty of making sharp or precise distinctions in the world. On the other hand, ambiguity is associated with two or more alternatives such that the choice between them is left unspecified. These uncertainties have successfully been solved thanks to using fuzzy sets for several years. When we consider a fuzzy set, an element x belongs to the set with a given membership degree interpretable as possibility or truthfulness. It is usually denoted by membership function $y = \mu(x)$ whose value is in the interval $\langle 0, 1 \rangle$ [13]. If y equals 1, the x belongs to the set completely. If y equals 0, the x does not belong to the set at all. Otherwise, the x belongs to the set with a possibility between values 0 and 1.

By far fuzzy IF-THEN rules are the most visible among all techniques developed using fuzzy sets due to their wide range of successful industrial applications ranging from customer products, automotive control, medical imaging, to financial trading [12]. A group of these fuzzy rules contains rules in the following form: IF *Condition* THEN *Conclusion*. Both *Conditions* and *Conclusions* contain one or several expressions in the form '*Linguistic variable* is *linguistic term*'. Among the expressions there are operators such as AND. In Data Mining, IF-THEN rules are often classified into classification rules and association rules. This paper is focused on classification rules and this kind of rules is labeled rules hereafter. They are used for determining the classes of instances in new conditions on the basis of the classes of instances in known conditions. All rules in a group have in their conclusions only the same class variable – the variable whose terms are needed to be determined.

Real-world data relating to the field of otoneurology were explored. In otoneurology, vertigo or dizziness and other balance disorders are investigated [3]. They are a common nuisance and can be stemmed from a serious disease such as a tumour involving the acoustic nerve. They can encounter people of all ages. The investigations in otonerology themselves are focused on searching for the causes of the diseases and disorders, finding treatments and hindering accidents originated from such harms. For this purpose, computational classification methods can be used to identify a patient's disease and differentiate diseases from each other. Most of these methods but neural networks require a transformation of the quantitative part of data in the process of discretization [1]. But discretization of quantitative variables into crisp intervals does not always correspond to real situations in the data. For example, if commonly accepted threshold of severe rotating feeling of an patient is 75.5000, then 75.4999 does not mean strong and 75.5001 severe rotating feeling definitely. Thus, it seems more realistic to use fuzzy intervals for strong as well as severe rotating feelings with respect to the state of the patient.

The paper is organized as follows. In Section 2, we describe the used otoneurological data and its transformed fuzzy form. Section 3 contains details of the fuzzy IF-THEN rule induction approach applied to that data. Its true positive rates and accuracy were evaluated. The results of evaluations are discussed in Section 4. Finally, we conclude this paper with discussion in Section 5.

2. DATASET

As a dataset, a group of 815 instances classified into six possible diseases and described by 38 variables as replies to queries about patients' symptoms, medical history, clinical findings and the results of physiological measurements was used. In our earlier report we evaluated the 38 variables to be the most important ones from among a larger set [5] and, among them, we also identified 5 most important variables *Time from Symptoms*, *Frequency of spells*, *Duration of attack*, *Duration of hearing symptoms*, *Head trauma*. All of the 38 variables and their types are shown in Tab. 1. For the types, the following symbols are employed: B = binary, N = nominal, O = ordinal and Q = quantitative. Category numbers are used after the ordinal and nominal ones. There were 11 % missing values in the dataset. We imputed them with modes for 11 binary variables and one nominal variable and with medians for 10 ordinal and 16 quantitative variables. The only, four-valued nominal variable *Hearing loss type* was still substituted by three binary variables to justify the use of Euclidean distance. Thus, there were 40 variables altogether. Euclidean measure cannot be applied to nominal variables, except binary. This is not essential in the current work, but we desired to use the data similarly to the approach in our earlier article [4] to justify a comparison. Imputation was carried out disease-wise since it is naturally crucial that there are differences between diseases. Use of modes and medians in imputation is a simple and common way, but sufficient here, since the number of the missing values was small and we earlier found [7] that such sophisticated imputation methods as Expectation Maximization (EM) and linear regression did not get better classification results when discriminant analysis was used.

Table 1 Variables and their types

No.	Variable	Type
1	*Patient's age*	Q
2	*Time from Symptoms*	O7
3	*Frequency of spells*	O6
4	*Duration of attack*	O6
5	*Severity of attack*	O5
6	*Rotational vertigo*	Q
7	*Floating vertigo*	Q
8	*Tumarkin-type drop attacks*	O4
9	*Positional Vertigo*	Q
10	*Unsteadiness outside attacks*	O4
11	*Duration of hearing symptoms*	O7
12	*Hearing loss of right ear between attacks*	B
13	*Hearing loss of left ear between attacks*	B
14	*Hearing loss type*	N4
15	*Severity of tinnitus*	O4
16	*Time of first tinnitus*	O7
17	*Ear infection*	B
18	*Ear operation*	B
19	*Head or ear trauma with noise injury*	B
20	*Chronic noise exposure*	B
21	*Head trauma*	B
22	*Ear trauma*	B
23	*Spontaneous nystagmus*	B
24	*Swaying velocity of posturography eyes open* (cm/s)	Q
25	*Swaying velocity of posturography eyes closed* (cm/s)	Q
26	*Spontaneous nystagmus (eye movement) velocity* (°/s)	Q
27	*Caloric asymmetry* (%)	Q
28	*Nystagmus to right*	Q
29	*Nystagmus to left*	Q
30	*Pursuit eye movement amplitude gain* (%)	Q
31	And its *latency* (ms)	Q
32	*Audiometry 500 Hz right ear* (dB)	Q
33	*Audiometry 500 Hz left ear* (dB)	Q
34	*Audiometry 500 2 kHz right* (dB)	Q
35	And *left ear* (dB)	Q
36	*Nausea or vomiting*	O4
37	*Fluctuation of hearing*	B
38	*Lightheadedness*	B

The dataset itself comes from Helsinki University Central Hospital in Finland where it has been collected for several years. The six diseases appearing in it are vestibular schwannoma, benign positional vertigo, Meniere's disease, sudden deafness, traumatic vertigo and vestibular neuritis. Their absolute frequencies are 130, 146, 313, 41, 65 and 120 respectively. The corresponding relative frequencies are 16 %, 18 %, 38 %, 5 %, 8 % and 15 % respectively. It is clear from the frequencies that the subset of Meniere's disease is far larger than two small subsets of sudden deafness and traumatic vertigo. These are not frequent disorders or diseases in general. Notwithstanding this, a few hundred thousand people in Finland suffer from vertigo and balance problems. These are diagnostically difficult diseases. In order to incorporate fuzzy sets into the dataset, we transformed it with the help of the fuzzification algorithm introduced in [8]. The transformed dataset is called fuzzified otoneurological dataset here and has its format as follows. It contains known instances corresponding to the 815 collected instances of the dataset. The set of these instances is marked $\boldsymbol{V}$. Each instance $\boldsymbol{e} \in \boldsymbol{V}$ is described by 40 linguistic variables $\mathbf{A} = \{A_1, A_2, \ldots, A_{40}\}$. Each linguistic variable A_k, $k = 1, 2, \ldots, 40$, measures some important feature and is represented by a group of primary linguistic terms $a_{k,1}, a_{k,2}, \ldots, a_{k,l}, \ldots, a_{k,n_k}$, which is denoted by $A_k = \{a_{k,1}, a_{k,2}, .., a_{k,l}, \ldots, a_{k,n_k}\}$. Disease linguistic variable C classifies all possible instances $\boldsymbol{e}$ into six disease primary linguistic terms $c_1, c_2, c_3, c_4, c_5, c_6$, which correspond to the above-mentioned diseases. The fact C classifies instances into those diseases is denoted by $C = \{c_1, c_2, c_3, c_4, c_5, c_6\}$. The possibility for a particular *linguistic term* and instance $\boldsymbol{e}$ is denoted by membership function $\mu_{linguistic\ term}(\boldsymbol{e})$.

3. FUZZY IF-THEN RULE INDUCTION

A fuzzy rule induction method based on transformation of a fuzzy decision tree made with cumulative information estimations and the process of classification with the method are presented here. The following form of fuzzy rules is considered: r_i = IF $\boldsymbol{E}_i$ THEN C is c_j (ECR$_i$), $j = 1, 2, \ldots, 6$, where

$$\boldsymbol{E}_i = A_{i_1} \text{ is } a_{i_1,j_1} \text{ AND } A_{i_2} \text{ is } a_{i_2,j_2} \text{ AND } \ldots$$
$$\ldots \text{ AND } A_{i_{m_i}} \text{ is } a_{i_{m_i},j_{m_i}}$$
$$(\text{or equally } \boldsymbol{E}_i = a_{i_1,j_1} \text{ AND } a_{i_2,j_2} \text{ AND } \ldots$$
$$\ldots \text{ AND } a_{i_{m_i},j_{m_i}})$$

is a sentence linguistic term containing at least one linguistic variable $A_k \in \mathbf{A}$ and any of $A_k \in \mathbf{A}$ is its part once, or is not its part at all. ECR$_i$ means *E*xtra *C*riteria for

*R*ule r_i, $i = 1, 2, \ldots, p$. ECR_i is an abstraction used because of generality and the criteria themselves can be different when another algorithm is applied for making the rules. In our case, $\mathrm{ECR}_i = \{F_i^1, F_i^2, F_i^3, F_i^4, F_i^5, F_i^6\}$ where F_i^j is a value interpreted as the certainty degree of disease primary linguistic term c_j in fuzzy rule r_i. If the possibility of sentence linguistic term $\boldsymbol{E}_i$ for instance $\boldsymbol{e}$ $\mu_{\mathbf{E}_i}(\boldsymbol{e})$ is required, it is computed as multiplication of the possibilities of its linguistic terms, i.e. $\mu_{\mathbf{E}_i}(\boldsymbol{e}) = \mu_{a_{i_1,j_1}}(\boldsymbol{e}) \cdot \mu_{a_{i_2,j_2}}(\boldsymbol{e}) \cdot \ldots \cdot \mu_{a_{i_{m_i},j_{m_i}}}(\boldsymbol{e})$.

The applied form of the fuzzy rule induction method can be described as it is in Tab. 2 and Tab. 3. In Tab. 2 the process of fuzzy decision tree induction based on cumulative information estimations is presented and in Tab. 3 the process of transformation of such a tree into the fuzzy rules is recounted. For the fuzzy decision tree induction, five input parameters α, β, $\mathbf{A}$, C, $\boldsymbol{V}$ are required. Parameters $\alpha, \beta \in \langle 0, 1 \rangle$ are used for controlling the height of the fuzzy decision tree, i.e. the numbers of linguistic variables in the paths from the root to the leaves. Increasing α and decreasing β leads to decreasing these numbers and vice versa. It may increase classification accuracy for instances outside $\boldsymbol{e} \in \boldsymbol{V}$ as well as decrease classification accuracy for instances $\boldsymbol{e} \in \boldsymbol{V}$. In Tab. 2 there are some other symbols and their meanings are as follows. Function $\mathrm{argmax}\{\mathrm{f}(x) \mid x \in X\}$ returns the $x \in X$ which $\mathrm{f}(x)$ has the maximal value for. $\mathbf{M}(\boldsymbol{V})$ is the cardinality of $\boldsymbol{V}$. Expression $\mathbf{A} - \boldsymbol{E}$ is a set containing all linguistic variables in $A_k \in \mathbf{A}$ with the exception of those variables that appear in the sentence linguistic term $\boldsymbol{E}$. Symbols $\mathbf{II}(\boldsymbol{E})$ and $\mathbf{I}(c_j/\boldsymbol{E})$ express cumulative information and conditional information belonging to cumulative information estimations whose detail explanation is in our previous study [9]. Briefly, cumulative information $\mathbf{II}(\boldsymbol{E})$ is defined as it is in Formula 1. In it, expression $\boldsymbol{E} = \theta$ means that sentence linguistic term $\boldsymbol{E}$ does not contain any linguistic variable. Otherwise, it contains at least one linguistic variable $A_k \in \mathbf{A}$. It describes vagueness of $\boldsymbol{E}$. The greater its value is, the greater the vagueness of $\boldsymbol{E}$ is. Conditional information $\mathbf{I}(c_j/\boldsymbol{E})$ of $c_j \in C$ assuming that $\boldsymbol{E}$ is known is defined in Formula 3 and describes the vagueness of c_j if $\boldsymbol{E}$ is known. Its definition contains $\mathbf{II}(\boldsymbol{E}$ AND $c_j)$ defined in Formula 2. It is cumulative information for a sentence linguistic term containing also a disease linguistic term $c_j \in C$. Some other symbols such as $\mathbf{F}(c_j/\boldsymbol{E})$ and $\mathbf{CRIT}(A_k; \boldsymbol{E})$ appear in Tab. 3. Symbol $\mathbf{F}(c_j/\boldsymbol{E})$ is defined in Formula 4 and is interpreted as frekvency of disease linguistic term $c_j \in C$ for a rule with condition $\boldsymbol{E}$. Symbol $\mathbf{CRIT}(A_k; \boldsymbol{E})$ is defined in Formula 8 and stands for the criterion used for association of linguistic variable $A_k \in \mathbf{A}$ with the node whose path from the root to the branch it is connected to corresponds to combinations '*linguistic variable* is *primary linguistic term*' in $\boldsymbol{E}$. It is defined with Formulas 6, 7. Formula 6 is defined with Formula 5 because of lack of space. $\mathbf{I}(\boldsymbol{E};A_k)$ is mutual information expressing average amount of information that is gained about disease linguistic variable C if $\mu_{a_k}(\boldsymbol{e})$, $a_k \in A_k$, $\boldsymbol{e} \in \boldsymbol{V}$, and $\boldsymbol{E}$ are known. $\mathbf{HH}(A_k)$ is cumulative entropy. Its formula is a part of Shannon's entropy whose definition is $\mathbf{H}(A_k) = \mathbf{HH}(A_k)/\mathbf{M}(\boldsymbol{V})$.

$$\mathbf{II}(\boldsymbol{E}) = \begin{cases} -\log_2 \sum_{\boldsymbol{e} \in \boldsymbol{V}} \mu_{\boldsymbol{E}}(\boldsymbol{e}); & \text{if} \quad \boldsymbol{E} \neq \theta \\ -\log_2 \mathbf{M}(\boldsymbol{V}) \quad ; & \text{if} \quad \boldsymbol{E} = \theta \end{cases} \tag{1}$$

$$\mathbf{II}(\boldsymbol{E} \quad \text{AND} \quad c_j) = -\log_2 \sum_{\boldsymbol{e} \in \boldsymbol{V}} \mu_{\boldsymbol{E}}(\boldsymbol{e}) \cdot \mu_{c_j}(\boldsymbol{e}) \tag{2}$$

$$\mathbf{I}(c_j / \boldsymbol{E}) = \mathbf{II}(\boldsymbol{E} \quad \text{AND} \quad c_j) - \mathbf{II}(\boldsymbol{E}) \tag{3}$$

$$\mathbf{F}(c_j / \boldsymbol{E}) = 2^{-\mathbf{I}(c_j / \boldsymbol{E})} \tag{4}$$

$$\begin{aligned} \mathbf{TMP}(c_j; a_{k,i}; \boldsymbol{E}) &= \mathbf{II}(\boldsymbol{E} \quad \text{AND} \quad c_j) + \\ &+ \mathbf{II}(\boldsymbol{E} \quad \text{AND} \quad a_{k,i}) - \\ &- \mathbf{II}(\boldsymbol{E} \quad \text{AND} \quad a_{k,i} \quad \text{AND} \quad c_j) - \mathbf{II}(\boldsymbol{E}) \end{aligned} \tag{5}$$

$$\begin{aligned} \mathbf{I}(\boldsymbol{E}; A_k) &= \sum_{c_j \in C} \sum_{a_{k,i} \in A_k} \\ &\mathbf{M}(\boldsymbol{E} \quad \text{AND} \quad a_{k,i} \quad \text{AND} \quad c_j) \cdot \mathbf{TMP}(c_j; a_{k,i}; \boldsymbol{E}) \end{aligned} \tag{6}$$

$$\mathbf{HH}(A_k) = \sum_{a_{k,i} \in A_k} \mathbf{M}(a_{k,i}) \cdot \mathbf{II}(a_{k,i}) \tag{7}$$

$$\mathbf{CRIT}(A_k; \boldsymbol{E}) = \frac{\mathbf{I}(\boldsymbol{E}; A_k)}{\mathbf{HH}(A_k)} \tag{8}$$

Table 2 Algorithm for fuzzy decision tree induction

fuzzy decision tree = makeTree(α; β; $\mathbf{A}$; C; $\boldsymbol{V}$)	
Step 1	Make the root and associate fuzzy attribute $\max A_k = \mathrm{argmax}\{\mathbf{CRIT}(A_k; \theta) \mid A_k \in \mathbf{A}\}$ with it. Make a branch for each $a_{k,l} \in A_k$, 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 $\boldsymbol{E}$ for the current branch. $\boldsymbol{E}$ consists of all "*Linguistic variable associated with node* is *linguistic term associated with branch*" from the root to the current branch connected with AND operator.
Step 3	Set branchII = $\mathbf{II}(\boldsymbol{E})$ and minClassTermI = $\min\{\mathbf{I}(c_j/\boldsymbol{E}) \mid c_j \in C\}$. If [branchII $\geq -\log_2(\alpha \cdot \mathbf{M}(\boldsymbol{V}))$] $\vee$ [minClassTermI $\leq -\log_2(\beta)$] $\vee$ [$(\mathbf{A} - \boldsymbol{E}) = \theta$], 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 $F^j = \mathbf{F}(c_j/\boldsymbol{E})$ $\forall c_j \in C$ and class linguistic term $\mathrm{argmax}\{F^j \mid c_j \in C\}$ with the made leaf. Go to Step 2.
Step 5	Make a node, connect it with the current branch and associate linguistic variable $\max A_k = \mathrm{argmax}\{\mathbf{CRIT}(A_k; \boldsymbol{E}) \mid A_k \in \boldsymbol{V}, A_k \notin \boldsymbol{E}\}$. Consider the current branch processed. Make a branch for each $a_{k,l} \in \max A_k$, connect them

	with the made node, associate them with the particular $a_{k,l}$ and consider them unprocessed. Go to Step 2.

Table 3 Transformation into fuzzy rules

$\{r_i\}$ = makeRules(fuzzy decision tree)	
Step 1	For each leaf i, mark the linguistic term associated with it as c^i. For each leaf i, take the branch going to it and make linguistic term $\boldsymbol{E}$ for this branch. $\boldsymbol{E}$ consists of all "*Fuzzy attribute associated with node* is *linguistic term associated with branch*" from the root to that branch and they are connected with AND operator. Set $\mathrm{ECR}_i=\{F_i^j \mid$ all F_i^j that were associated with leaf $i\}$ for each leaf i.
Step 2	Make a rule in the form r_i = IF E_i THEN C is c^i (ECR_i) for each c^i, E_i, and ECR_i.

Table 4 Classification of an instance ***e***

$\{\mu_{c_j}(\boldsymbol{e})\}$ = classify($\{r_i\}$; $\boldsymbol{e}$; C)	
Step 1	Compute $\mu^i_{\boldsymbol{E}}(\boldsymbol{e})$ for each fuzzy rule r_i = IF $\boldsymbol{E}_i$ THEN C is c_j (ECR_i).
Step 2	Set $\mu_{c_j}(\boldsymbol{e}) = \sum_{\forall i} \mu^i_{\boldsymbol{E}}(\boldsymbol{e}) \cdot F_i^j$, where $F_i^j \in \mathrm{ECR}_i$.

Since classification is a process of determining values of $\mu_{c_j}(\boldsymbol{e})$ for all disease linguistic terms $c_j \in C$ on the basis of fuzzy rules made in accordance with Tab. 2 and Tab. 3 here, an algorithm that may perform it is needed. Such an algorithm is presented in our previous study [10] and it is described according to the used terminology of this paper in Tab. 4. This algorithm supposes that the values of all membership functions $\mu_{a_{k,l}}(\boldsymbol{e})$, $a_{k,l} \in A_k \in \mathbf{A}$, are known for a classified instance $\boldsymbol{e}$.

4. EXPERIMENTS AND RESULTS

The algorithms described in the previous section are labeled MCI - *M*inimization of *C*umulative *I*nformation hereafter and had been implemented in Java as a part of software library Fuzzy Rule Miner technically introduced in [2]. This implementation was used for testing and its results were compared with multilayer perception neural networks (MLP) implemented in Matlab as we had done comprehensive research ([4] and some other papers) in the area of MLP. These two methods also highly differ from each other in the understandability of the obtained knowledge. While MLP are frequently applied as a 'black-box technique' to the classification of data without any explanation of what the networks learnt, generated fuzzy rules are easily understandable and readable for the expert.

The neural networks were formed with five input nodes, six hidden nodes at one layer and six output nodes corresponding to the six diseases. The most important five variables stated in Section 2 were only input, since using more variables, i.e. 9, 13, 20, 30 and 40, deteriorated results. This seemingly contradictory situation was caused by the relative scarcity of training instances compared to the number of weights to be trained in a network. The significantly imbalanced distribution among the six diseases also had impairing influence, which finally resulted, with 30 and 40 attributes, in a situation where the Meniere's disease having largest number of instances was the only learnt. The number of hidden nodes was the only free parameter with respect to the network structure, and it was tested up to as small as eight to keep the size of the structure small enough. To guarantee reasonable training for a multilayer perception network, there is a rule of thumb that the number of weights to be trained should not be larger than 1/10 of the number of training instances [11]. When we used five input, six hidden and six output nodes, we obtained $5 \cdot 6 + 6 \cdot 6 = 66$ weights, which were fewer than 1/10 of $0.90 \cdot 815$ training instances in crossvalidation. Feedforward backpropagation learning algorithm was applied with the sigmoidal threshold function, adaptive learning rate and momentum coefficient. Before the learning stage, we experimented with disjoint validation sets to observe possible overlearning, which did not affect when no more than 500 training epochs were executed.

Table 5 Average true positive rates and accuracies

Disease term	True positive rate [%]		
	MCI[40]	**MLP[5]**	**MLP[40]**
Vestibular schwannoma	71	71	4
Benign positional vertigo	69	68	2
Meniere's disease	97	91	99
Sudden deafness	71	1	0
Traumatic vertigo	28	70	0
Vestibular neuritis	84	83	1
Accuracy [%]	**79**	**76**	**39**

Crossvalidation technique was employed in the performed tests themselves. On the basis of this technique, 10 % of the entire dataset was picked up as a testing set and the other instances (90%) as a learning set. Ten disjoint testing sets of 10 % were run along with the corresponding learning sets. Thus, every instance was included in some testing set. Initialization values were, of course, also varied between runs. Therefore, we performed 10 runs for each testing set. We computed true positive rate for each disease j, $j = 1, 2, 3, 4, 5, 6$, and accuracy for all the data. Let p_j be the number of instances in the disease j and tp_j true positive instances computed for disease j. The true positive rate was formed as it is defined in the following Formula 9:

$$tpr_j = \frac{tp_j}{p_j} \cdot 100 \text{ [\%]} \tag{9}$$

Let k be the number of diseases and M the number of all instances in a testing set. Formula 10 was used for calculation of the accuracy of classification for all diseases:

$$acc = \frac{\sum_{j=1}^{k} tp_j}{M} \cdot 100\ [\%] \qquad (10)$$

The above-defined true positive rate and accuracy were computed as means over all 100 runs (10 times 10 crossvalidation). The means of the results are presented in Tab. 5. In it, the true positive rates for particular diseases are in columns MCI[40], MLP[5] ad MLP[40] and the accuracies of these methods are in the last row. MLP[5] means multilayer perception neural networks described in this section and used for five most important input variables stated in Section 2. MLP[40] means multilayer perception neural networks applied for forty variables as they are explained in Section 2. In the case of MCI[40], forty fuzzy variables $\mathbf{A} = \{A_1, A_2, \ldots, A_{40}\}$ defined in Section 2 are used as the input for the algorithms from Section 4. Including MLP[40] into the comparison allows us to compare the results with the same set of initial input variables. On the other hand, MLP[5] represents the results of MLP including our previous investigations with MLP as well as the used dataset.

5. DISCUSSION

The classification accuracy of MCI[40] with 79 % is better than the results of MLP as it is in Tab. 5. The neural networks lost the instances of sudden deafness virtually entirely, since it was the disease with the smallest number of instances comprising mere 41 of all 815 instances. In addition to this fact, it is known to be medically awkward, i.e. its diagnosis can be erroneously mixed with the features of some other one such as Meniere's disease. Since the distribution between the diseases was so skewed in the present data that the disease with the largest number of instances included 38 % of the instances in the dataset and the one with least number only 5 % of the instances, this dataset was difficult especially for multilayer perception neural networks, which often seem to require a fairly uniform distribution of diseases in instances to learn features of our dataset. In addition, a sufficiently great number of training instances related to the size of a network, the number of its arc weights in the graph structure was needed which limited the number of variables to five. This was a considerable restriction in our earlier work [4]. When five most important variables were used, an average accuracy of 76 % was gained for this data in [4]. For the most important nine variables, it was 73 %, but considerably decreased down to 47 % and 39 % for twenty and forty variables respectively. However, the use of principal components of the data in [4] enabled the application of all forty attributes and produced at least equally good results. The use of principal components may be an interesting technique for improving accuracies. However, for practical applications, this idea is perhaps not the best choice because principal component analysis has to be performed for the whole data, including training instances, not only testing instances.

Sudden deafness was not detected by multilayer perception neural networks successfully and its classification is significantly better with the presented method. Besides, we believe there is a potential to improve its results. Firstly we are going to study the effects in the changes of the data preprocessing, especially of the process of fuzzification of quantitative values. Some other algorithm different from [8] or/and the assistance of medical experts may help so that we can define more accurate membership functions for primary linguistic terms. It can lead to reducing information loss during fuzzification. Secondly, for more than 10 years our otoneurological data has been collected with respect to crisp classification where only one disease is considered fully possible and all the others are considered fully impossible, which does not always correspond to reality. For example, a doctor can think that Meniere's disease is possible with 0.5, sudden deafness with 0.8 and traumatic vertigo with 0.2 for a patient. Since our method allows us to model this situation, further investigation into our data collection is being pondered.

REFERENCES

[1] Berka, P.: Knowledge Discovery in Databases, Academia, Prague, 2003, ISBN 80-200-1062-9.

[2] Boháčik, J., Matiaško, K., Levashenko, V.: Software for Making and Using Fuzzy Rules, Journal of ELECTRICAL ENGINEERING, Volume 57, 2006, pp. 85-88, ISSN 1335-3632.

[3] Furman, J.: Otoneurology, Neurologic Clinics, Volume 23, No. 2/2005, Elsevier, Netherlands.

[4] Juhola, M., Laurikkala, J.: Increasing Accuracy of Neural Network Classification with Principal Component Analysis for Otoneurological and Female Urinary Incontinence Data, 18th International Congress of the European Federation for Medical Informatics, IOS Press, Saint Malo, France, 2003, pp. 1092-1097.

[5] Kentala, E., Pyykko, I., Auramo, Y., Juhola, M.: Database for vertigo, Otolaryngology – Head and Neck Surgery, Volume 112, No. 3/1995, Elsevier, pp. 383-390.

[6] Klir, G. J.: Where Do We Stand on Measures of Uncertainty, Ambiguity, Fuzziness and the Like?, Fuzzy Sets and Systems, Volume 24, No. 2/1987, Amsterdam, pp. 141-160, ISSN 0165-0114.

[7] Laurikkala, J., Kentala, E., Juhola, M., Pyykko, I., Lammi, S.: Usefulness of Imputation for the Analysis of Incomplete Otoneurological Data, International Journal of Medical Informatics, Volume 58-59, No. 1/2001, pp. 235-242.

[8] Lee, H. M., Chen, C. M., Chen, J. M., Jou, Y. L.: An Efficient Fuzzy Classifier with Feature Selection Based on Fuzzy Entropy, IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, IEEE Systems, Man, and Cybernetics Society, Volume 31, No. 3/2001, pp. 426-432, ISSN 1083-4419.

[9] Levashenko, V., Zaitseva, E.: Usage of New Information Estimations for Induction of Fuzzy

Decision Trees, 3rd IEEE Int. Conference on Intelligent Data Engineering and Automated Learning, 2002, Kluwer Publisher, the UK, pp. 493-499, ISBN 978-3-540-44025-3.

[10] Matiaško, K., Boháčik, J., Levashenko, V., Kovalík, S.: Learning fuzzy rules from fuzzy decision trees, Journal of Information, Control and Management Systems, Volume 4, No. 2/2006, pp. 143-154, ISSN 1336-1716.

[11] Swingler, K.: Applying Neural Networks - A Practical Guide, Academic Press, London, 1996, ISBN 0-12-679170-8.

[12] Yen, J.: Fuzzy Logic – A Modern Perspective, IEEE Transactions on Knowledge and Data Engineering, Volume 11, No. 1/1999, pp. 153-165, ISSN 1041-4347.

[13] Zadeh, L.: Fuzzy Sets, Journal of Inf. and Control, Volume 8, No. 3/1965, pp. 338-353.

ACKNOWLEDGEMENTS

This work was supported by the Academy of Finland (fellowship 120264). The first author is grateful to the Centre for International Mobility (CIMO), Finland, for its 9-month financial support at the Department of Computer Sciences of the University of Tampere, Finland. The authors also wish to say thank to Docent E. Kentala, M.D., and Professor I. Pyykkö, M.D., for medical information and medical data collected at the Department of Otorhinolaryngology of the Helsinki University Central Hospital, Finland.

Received April 9, 2008, accepted July 11, 2008

BIOGRAPHIES

Ján Boháčik was born in 1983. In July 2006 he graduated in the study field of Information and Control Systems at the Faculty of Management Science and Informatics of the University of Žilina and defended Master thesis named Analysis of Algorithms for Making Fuzzy Rules. Since October 2006 he has been a full-time PhD student of Applied Informatics at the Department of Informatics of that Faculty. His scientific research is focused on making and using fuzzy rules in the process of Knowledge Discovery in Databases.

Martti Juhola received his M.Sc., Phil.Lic. and Ph.D. degrees in Computer science at the University of Turku, Finland, in the 1980s, where he was an academic assistant, lecturer, and researcher, later becoming a professor at the University of Kuopio, Finland. Since 1997 he has been a professor at the University of Tampere. His research interests include medical informatics, signal analysis, neural networks, pattern recognition, information retrieval and applications of algorithms.

Vitaly Levashenko was born in 1968. In 1991 he graduated at the Minsk Radio-Engineering Institute (BSUIR). He received his PhD degree in Computer Science from this Institute in 1997. Since 2002 he has been working as an associate professor at the Faculty of Management Science and Informatics of the University of Žilina. His main scientific and educational interests include Multiple-Valued Logics, Fuzzy Logics, Data Mining and Decision Making Support Systems.