Structured neuro-fuzzy classifier for medical decision support

Marian B. Gorzalczany and Piotr Gradzki
Department of Electrical and Computer Engineering
Kielce University of Technology
Al. 1000-lecia P.P.7, 25-314 Kielce, Poland
Phone: +48-41-3424217, Fax: +48-41-3447758
email: ketmg@eden.tu.kielce.pl

ABSTRACT: The paper presents an approach which combines artificial neural networks and fuzzy logic, yielding a system that can be called a structured neuro-fuzzy classifier. In this approach, the fuzzy rule-based system is represented by a feed-forward network-like structure. It is able to learn and to generalize the gained knowledge (this ability comes from the network-like structure of the system) as well as it is able to explain the decisions made by providing a set of fuzzy rules describing the mechanisms of decision making. The proposed approach has been successfully applied to the design of the decision support system for the treatment of duodenal ulcer with the use of highly selective vagotomy.

KEYWORDS: Fuzzy sets, artificial neural networks, neuro-fuzzy systems, neuro-fuzzy classifiers, decision support systems

1. INTRODUCTION

Intelligent decision support systems belong to the applicational direction of research in the artificial intelligence (AI) field. AI comprises methods, tools and systems for solving problems that normally require the intelligence of humans. The main paradigm adopted in AI to achieve this goal is the symbolic paradigm based on the theory of physical symbolic systems [Newell, 1972]. The symbolic AI systems have proved effective in handling problems characterized by exact and complete representations. However, these systems have very little power in dealing with linguistic, inexact, uncertain, imprecise or ambiguous information; this kind of information significantly contributes to the description of many real-life problems. On the other hand, there are many situations where the expert knowledge is not sufficient for the design of intelligent systems (incompleteness of the existing knowledge, problems caused by different biases of human experts, difficulties in forming rules, etc. [Kasabov, 1996]). The use of the available quantitative, numerical data, e.g. coming from measurements or obtained as the results of medical tests, etc. - collected over time in databases - can solve these problems. A study of these data and the extraction of the knowledge ”encoded” in these data, significantly enhances the performance of the decision support systems.

Since the traditional AI techniques are not able to deal effectively with the qualitative, linguistic knowledge and the huge amount of quantitative, numerical data describing decision making processes, new methods and algorithms of knowledge representation and reasoning are emerging. They are labelled by the term ”computational intelligence” [Bezdek, 1994] covering three main methodologies, that is, artificial neural networks, fuzzy logic and evolutionary algorithms (genetic algorithms) and can be treated as a kind of modern extension of traditional AI techniques. The fact that all these methodologies are complementary rather than competitive [Bezdek, 1994] strongly motivates the efforts to design hybrid systems combining these techniques. The fusion of artificial neural networks and fuzzy logic within one system is, at present, the most advanced and promising direction of research in this area.

The combination of both approaches has a sound basis because they usually complement each other in the design of ”intelligent” systems. It means the reduction of dis-advantages and the enhancement of desired properties of both methodologies in the resulting system. Artificial neural networks give a computational power to process large amounts of data and the ability of learning and generalizing the gained knowledge. Fuzzy logic creates a structural framework that utilizes and interpretes - on a higher, linguistic or semantic, level - the low-level results coming from neural networks [Bezdek, 1992].

The paper presents an approach which combines artificial neural networks and fuzzy logic, yielding a system that can be called a structured neuro-fuzzy classifier. In this approach, the fuzzy rule-based system is represented by a feed-forward network-like structure. It is able to learn and to generalize the gained knowledge (this ability comes from the network-like structure of the system) as well as it is able to explain the decisions made by providing a set of fuzzy rules describing the mecha-nisms of decision making. First, its learning phase is presented during which the network ”models” the available medical knowledge from a given domain. Then, the inference phase of the network is described during which the network operates as an inference engine. In turn, the application of the proposed methodology to the design of a support system for the surgical treatment of duodenal ulcer with the use of highly selective vagotomy (HSV), is presented.

2. A GENERAL STRUCTURE OF THE NEURO-FUZZY CLASSIFIER AS A DECISION SUPPORT SYSTEM [Gorzalczany, 1998]

A neuro-fuzzy classifier has n inputs (attributes, featu-res) $x_1, x_2, ..., x_n$ and an output which has the form of a possibility distribution over the set $Y = \{y_1, y_2, \ldots, y_H\}$ of class labels. In medical field, each input x_i represents one input medical attribute (a ”symptom”) taking values from the set X_i. The input attribute may be described either by numerical values (e.g. pulse rate is equal to 80 beats per minute) or by linguistic terms (e.g. blood pressure is ‘significantly inreased’, pulse rate is ‘low’, pain level is ‘high’, etc.); the latter are represented by appropriate fuzzy sets provided by an expert. Output set Y (a set of class labels), in medical field, is a set of potential diseases or possible outcomes of operation, etc.

Let $\boldsymbol{A}' = \{A_1', A_2', \ldots, A_n'\}$, where $A_i' \in F(X_i)$, i=1,2,...,n; $F(X_i)$ denotes a family of all fuzzy sets defined on the universe X_i. Moreover, let $\boldsymbol{F_X} = \{F(X_1), F(X_2), \ldots, F(X_n)\}$. $\boldsymbol{A}' \in \boldsymbol{F_X}$ is a general representation of the set of input attributes (features - symptoms). Each x_i is represented by a corresponding fuzzy set A_i'. In particular case, when we deal with a numerical value of x_i, the fuzzy set A_i' is being reduced to fuzzy singleton. Let $B' \in F(Y) = F_Y$ be a fuzzy set representing a possibility distribution defined over the set Y of class labels. The possibility distribution assigns to each class y_j from the set Y, a number from the interval [0, 1], indicating the possibility that the object described by $\boldsymbol{A}'$ belongs to that class. In medical field, a number from the interval [0, 1] indicates how possible is that y_j (a disease, an outcome of operation, etc.) occurs, given the ”symptoms” represented by $\boldsymbol{A}'$. In particular case, when we deal with a non-fuzzy possibility distribution over Y, the fuzzy set B' is being reduced to fuzzy singleton.

The design of the neuro-fuzzy classifying system is based on K input-output sets of learning data

$$\{\boldsymbol{A}_k', B_k'\}_{k=1}^{K}, \quad \boldsymbol{A}_k' = \{A_{1k}', A_{2k}', \ldots, A_{nk}'\} \in \boldsymbol{F_X}, \quad B_k' \in F_Y. \tag{1}$$

The problem of designing the neuro-fuzzy classifier, is to find the mapping

$$M: \boldsymbol{F_X} \to F_Y \tag{2a}$$

provided its restriction on learning data (1)

$$M_{\boldsymbol{A}'B'}: \boldsymbol{F_X} \to F_Y \tag{2b}$$

is known.

It would be desirable if the system could not only learn and generalize the gained knowledge but also explain the decisions it makes.

3. STRUCTURED NEURO-FUZZY SCHEME FOR MEDICAL KNOWLEDGE MODELLING

In the proposed approach, the fuzzy rule-based system is represented by a feed-forward network-like structure. One part of this structure corresponds to premises of the fuzzy rules and the other one - to their consequences. One can distinguish three main phases in the design of the proposed system. The first phase consists in the determination of the initial rule base describing the decision making process represented by available learning data (1). In the secong phase which corresponds to the learning of the proposed neuro-fuzzy classifier, the initial rule base is being ”tuned” in order to achieve the best approximation of the desired mapping (2b). In the third phase the obtained neuro-fuzzy classifier is tested against the learning and testing data, that is, the accuracy of the obtained mapping (2a) is being verified in regard to both groups of data.

It is worth to emphasize that the proposed neuro-fuzzy classifier has the following important features: a) it is able to learn and to generalize the gained knowledge - this ability comes from the network-like structure of the system which allows to apply appropriate learning techniques, b) it is able to explain the decisions made by providing a set fuzzy rules describing the mechanisms of decision making - therefore, the system is also able to extract or to synthesize the knowledge from the representative set of data. At present state of research, the proposed system is able to process only numerical data (and not the linguistic ones represented by fuzzy sets) describing the input attributes of the system.

The structure of the proposed neuro-fuzzy classifier in learning phase, and - after removing the dark area - in the inference phase is presented in Fig. 1. Symbols x_i', *i*=1, 2,..., *n* denote the input numerical learning data from (1) (A_{ik}' are the fuzzy singletons in the present case). $m_{B'}(y_1)$, $m_{B'}(y_2)$,..., $m_{B'}(y_H)$ representing the desired possibility distribution B', denote the correspon-ding set B_k' from (1). Each input x_i is described by several adjectives represented by fuzzy sets of three types:

$$\text{"}Small\text{"}: \; m_{S_i}(x_i) = \frac{1}{1+\exp(c_{S_i} \cdot (x_i - a_{S_i}))}, \; c_{S_i} > 0, \tag{3a}$$

$$\text{"}Medium\text{"}: m_{M_i^{(k)}}(x_i) = \exp\left(-\left(\frac{x_i - c_{M_i^{(k)}}}{s_{M_i^{(k)}}}\right)^2\right), \quad s_{M_i^{(k)}} > 0, \quad k = 1,2,\ldots,K_i \tag{3b}$$

$$\text{"}Large\text{"}: \; m_{L_i}(x_i) = \frac{1}{1+\exp(-c_{L_i} \cdot (x_i - a_{L_i}))}, \; c_{L_i} > 0, \tag{3c}$$

(one *Small*-type set, one *Large*-type set and several *Medium*-type sets can be defined for each x_i ; see the nodes S, L and M, respectively, in Fig. 1). The system of Fig. 1 represents a set of R fuzzy rules of the type:

$$\text{IF } (x_1 \text{ is } A_{1r}) \text{ AND...AND } (x_n \text{ is } A_{nr}) \text{ THEN } (\textit{possibility distribution } B_r' \text{ with } CF_r), \tag{4}$$

where A_{ir} is one of the S, L, or M-type fuzzy clusters for *i*-th input and *r*-th rule ($i = 1,2,\ldots,n$, $r = 1,2,\ldots,R$). CF_r is a certainty factor associated with *r*-th rule (the "strength" of the rule) and is represented by weight w_{rj} of Fig. 1. The rules are implemented using *min*-operator (they correspond to the so-called Mamdani's implication, cf. [Halgamuge, 1994]; $\wedge$ stands for *min* in Fig. 1). The output layer of the system presented in Fig. 1 is built of sigmoid-type nodes. The learning algorithm (a gradient-descent backpropagation-like technique) adjusts the weights and w_{rj} in the output layer and parameters c_{S_i}, a_{S_i}, $c_{M_i^{(k)}}$, $s_{M_i^{(k)}}$, c_{L_i}, a_{L_i} of the input S-M-L-layer in such a way as to minimize the mean-square error between the desired possibility distribution B' and the output possibility distribution B^0 .

The important issue of determining the initial rule base for the system of Fig. 1 has been solved by applying an adaptation of the approach proposed in [Wang, 1992]. It consists of three steps. First, the degrees of a given sample ($x_1', x_2', \ldots, x_n'$) of input learning data in different fuzzy regions represented by fuzzy sets (3) for particular inputs must be determined. Second, for a given input data x_i', a fuzzy region with maximum degree, say $A_{ij\$}$ must be selected. Since there are usually lots of input data samples, and each data sample generates - in such a way - one rule, it is highly probable that there will be some conflicting rules, that is, rules that have the same IF parts but a different THEN part. A way to resolve this conflict is to assign a degree $d(R)$:

$$d(R) = m_{A_{1j\$}}(x_1') \cdot m_{A_{2j\$}}(x_2') \cdot \; \ldots \; \cdot m_{A_{nj\$}}(x_n') \tag{5}$$

to each rule generated from input data, and accept only the rule from a conflict group that has maximum degree (third step). In this way not only is the conflict problem resolved, but also the number of rules is greatly reduced.

4. STRUCTURED NEURO-FUZZY INFERENCE ENGINE

In the inference phase (Fig. 1 after completion the learning phase and after removing the dark area), symbols x_i', *i*=1, 2,..., *n* represent the numerical input attributes of a new object (in medical fields - these are the symptoms

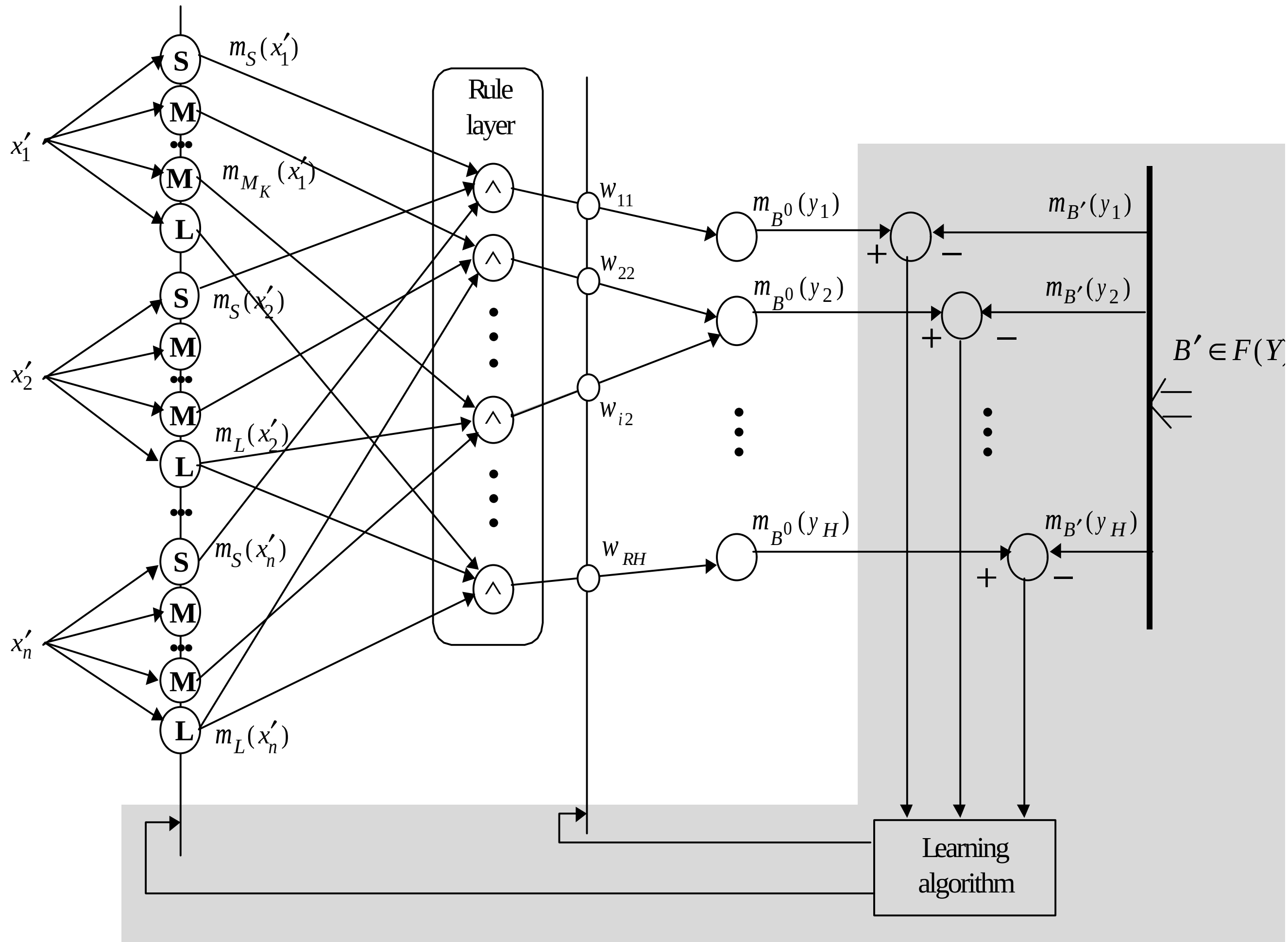


Fig. 1. Structured neuro-fuzzy classifier in the learning phase and (after removing the dark area) in the inference phase

describing a new case). The system processes the input data and generates, in a one step, the output possibility distribution (represented by fuzzy set $B^0 \in F(Y)$) over the set Y of class labels (e.g. diseases, results of operation, etc.). B^0 describes, given the input data, the possibility of occurrence of each option from the set Y. If a final, nonfuzzy decision y_{n-fd} is required, it can be derived from the B^0, by selecting the class label $y_{\hat{j}}$ which maximizes the $m_{B^0}(y_j)$, that is

$$y_{n-fd} = y_{\hat{j}} = \arg \max_{j=1,2,\ldots,H} [m_{B^0}(y_j)]. \tag{6}$$

In the inference phase both the testing of the system can be performed as well as its functioning as a decision support system can be checked. Moreover, the pruning of the neuro-fuzzy structure (removing ”weaker” rules) can be done; it will be briefly discussed in the next section.

5. APPLICATION - DECISION SUPPORT SYSTEM FOR THE TREATMENT OF DUODENAL ULCER WITH THE USE OF HIGHLY SELECTIVE VAGOTOMY

Surgical treatment of duodenal ulcer by highly selective vagotomy (HSV) is the newest and effective method of treatment of duodenal ulcer which consists in vagal denervation of the area secreting hydrochloric acid [Slowinski, 1992]. In the Department of Surgery at the F. Raszeja Mem. Hospital in Poznañ, Poland, 122 HSV patients took part in the follow-up program [Slowinski, 1992]. They were described by 11 pre-operating attributes and classified from the viewpoint of long term results of the operation into 4 classes of so-called Visick grading (see Appendix and [Slowinski, 1992]).

Decision support system uses these data to generate - for a new patient - a possibility distribution over the set of 4 mentioned classes which describes long term results of the operation by indicating the possibility of occurrence of each of these classes in a given case.

According to the general procedure for the construction of such a system, first, we have to determine its structure in terms of inputs and outputs. Overall, the are 11 input attributes. After a detailed analysis of the correlations between particular attributes, finally, a subset of 5 attributes (no. 3, 4, 6, 9, and 10 - see Appendix) has been chosen - see [Slowinski, 1992] for details. They are used as the inputs of our system. The output of the system is a set of 4 classes according to Visick grading.

In the second phase of the designing of the neuro-fuzzy classifier, each input attribute must be described by several adjectives represented by appropriate fuzzy sets. For each of the numerical-type attributes (inputs no.: 3, 6, 9, and 10) a set of 3 adjectives (*Small*, *Medium*, *Large*) - based on experts opinions and medical norms related to these attributes [Slowinski, 1992] - has been defined; see Fig. 2a for attribute no. 3 and Fig. 2b for attribute no. 6 (”initial shapes”). Membership functions of these fuzzy sets will be further ”tuned” during the learning phase of the system. For non-numerical attribute no. 4 (”complication of ulcer”), a set of 5 adjectives (singleton-type fuzzy sets: *Small*, *Medium1*, *Medium2*, *Medium3*, *Large*) representing particular verbal terms (‘none’, ‘acute haemorrhage’, ‘multiple haemorrhage’, ‘perforation in the past’, ‘pyloric stenosis’) by which this attribute is characterized by experts, has been defined.

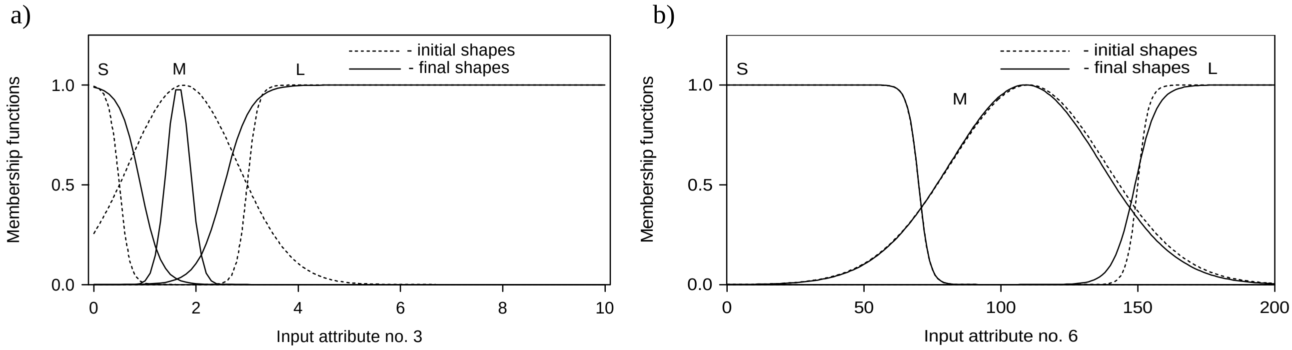


Fig. 2. Fuzzy sets describing input attribute no. 3 (a) and no. 6 (b)

The next stage of the system’s design consists in determination of the initial rule base. The available 122 samples of HSV data have been divided into two groups: 81 samples used as the learning data and 41 samples treated as the test data, preserving the proportions of occurrence of particular classes in both groups of data. Applying the procedure outlined at the end of Section III to the learnin data set, a collection of 46 rules has been obtained.

In the fourth phase, the structure of Fig. 1 representing the obtained set of fuzzy rules has been learned, that is, the tuning of the output weights w_{rj} (certatinty factors CF_r ’s characterizing particular rules) as well as the tuning of membership functions representing the adjectives for particular numerical-type input attributes have been performed (see Fig. 2 - ”final shapes” - for input attributes no. 3, and no. 6). As the learning algorithm, a simple gradient-descent backpropagation-like technique has been used. Fig. 3 presents the plot of the cost function (the mean-square error between the desired possibility distributions B' coming from the learning data and the output possibility distributions B^0 - see Fig. 1) versus the number of learning epochs (epoch means processing the whole learning data set). This plot clearly indicates that still a big ”margin” for further minimizing of the cost function remains. A study is being carried out on the application of more sophisticated learning techniques like optimization algorithms, e.g. conjugate-gradient and variable-metric methods; they should significantly improve the results of learning.

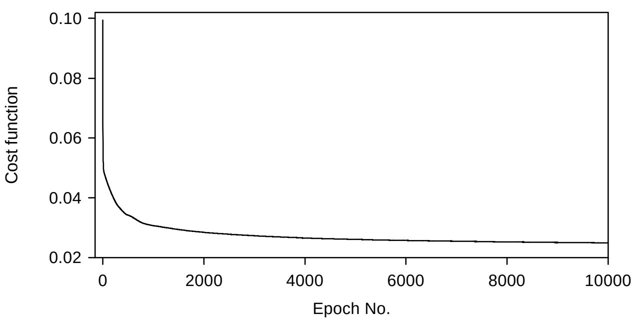
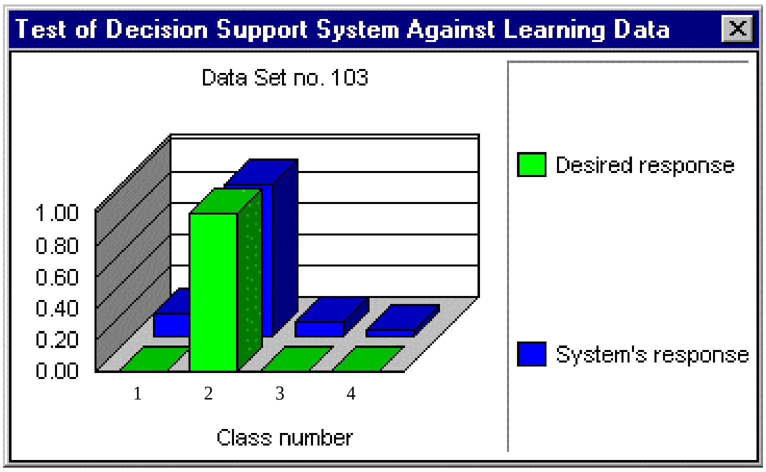


Fig. 3. Cost function versus epoch number plot

After completion the learning phase and after removing the dark area of Fig. 1, one can obtain the structured neuro-fuzzy classifier in the inference phase (the fifth phase of system's design). In this phase both the testing of the system (against learning and test data) can be performed as well as its functioning as a decision support system can be checked. Examples of the test of the system against selected samples of learning data and testing data are presented in Fig. 4.

a)



b)

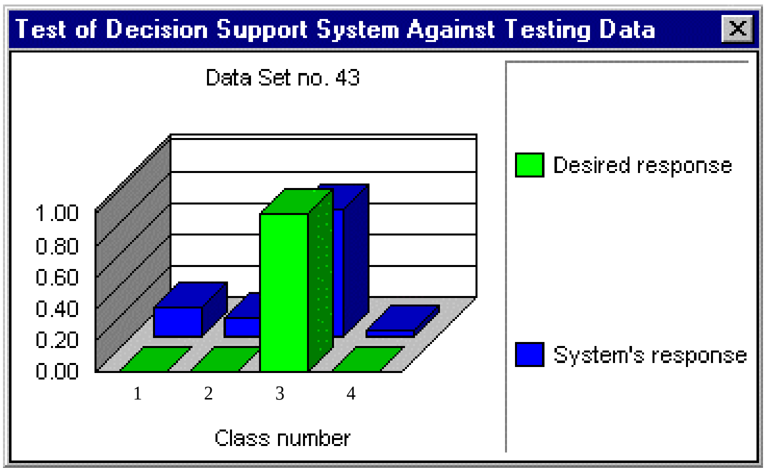


Fig. 4. Test of the system against a selected set of learning data (a) and testing data (b)

Fig. 5, in turn, shows an examplary response of the decision support system in the inference phase. This response suggests that for a given set of input data (symptoms) anticipated long term result of the operation is very good (class no. 2). There is a very small possibility that the result will be excellent (class no. 1) but one should not take into account two remaining results: satisfactory and unsatisfactory (classes 3 and 4, respectively).

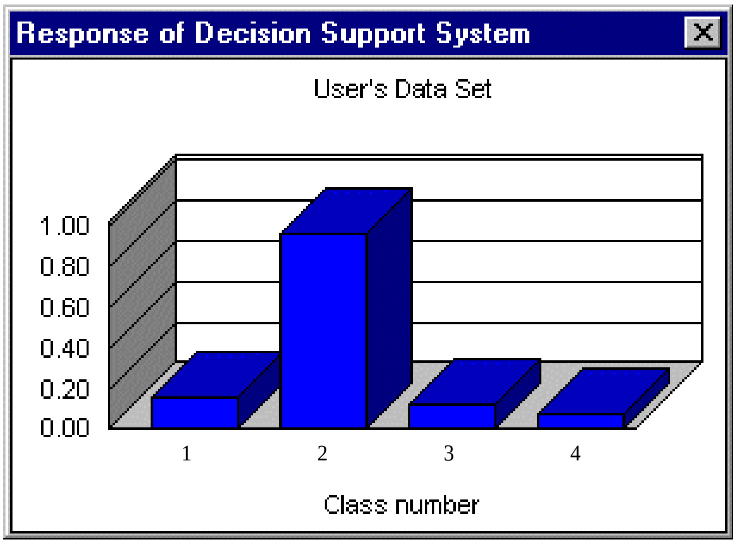


Fig. 5. An examplary response of the decision support system

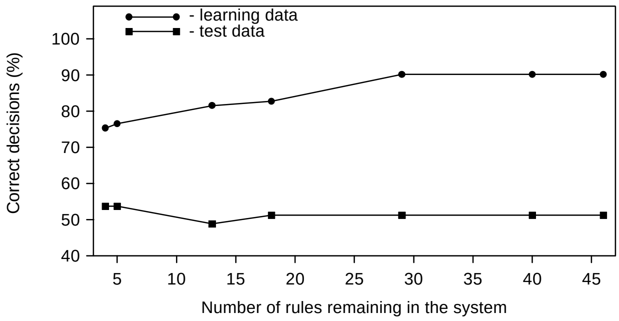


Fig. 6. Number of correct decisions versus number of rules remaining in the system

The final, sixth phase of system's design consists in pruning of the obtained structured neuro-fuzzy classifier by gradually removing "weaker" rules, that is, the rules characterized by small certainty factors CF_r 's, and observing how it affects the accuracy of the system in terms of the number of correct decisions made. Fig. 6 presents the number of correct decisions (in %) versus the number of rules remaining in the structure of Fig. 1. Removing the first 17 weakest rules does not affect the accuracy of the system neither with regard to the learning data nor to the testing ones. Removing further rules gradually decreases the accuracy of the system tested against the learning data. However, it does not affect the generalizing properties of the system, that is, its accuracy with regard to the test data. Finally, one can find that minimal number of four basic fuzzy rules (one for each of 4 classes) contains significant knowledge concerning the mechanisms of decisionmaking in the considered domain and provides relatively high level of system's accuracy. It should be stressed again that still a big "margin" for further minimization of the cost function in the learning phase remains and the application of more efficient learning techniques should benefit in increasing the accuracy of the system, and its learning and generalizing capabilities.

6. CONCLUSIONS

We have presented the structured neuro-fuzzy classifier that can be applied to the design of intelligent decision support systems, in particular, from the field of medicine. The proposed classifier has the following impor-tant features: a) it is able to learn and to generalize the gained knowledge - this ability comes from the network-like structure of the system which allows to apply appropriate learning techniques, b) it is able to explain the decisions made by providing a set fuzzy rules describing the mechanisms of decision making - therefore, the system is also able to extract or to synthesize the knowledge from the representative set of data.

The proposed methodology has been successfully applied to the design of medical decision support system in the surgical treatment of duodenal ulcer with the use of highly selective vagotomy.

7. REFERENCES

Bezdek, J.C., 1992, ”Guest Editorial”, *IEEE Trans. Neural Networks*, vol. 3, no. 5, p. 641.

Bezdek, J.C., 1994, ”What is computational intelligence”, in *Computational Intelligence Imitating Life*, J.M. Zurada, R.J. Marks II and C.J. Robinson (*Eds*.), IEEE Press.

Gorzalczany, M.B., 1998, ”Neuro-fuzzy classifier for decision-making support in medicine”, in *Proc. of IEEE ICIPS’98 (2-nd IEEE Int. Conference on Intelligent Processing Systems)*, Gold Coast, Australia, pp. 318-322.

Halgamuge, S.K.; Glesner, M., 1994, ”Neural networks in designing fuzzy systems for real world applications”. *Fuzzy Sets and Systems*, vol. 65, pp. 1-12.

Kasabov, N.K., 1996, *Foundations of neural networks, fuzzy systems, and knowledge engineering*. Cambridge, MA, MIT Press.

Newell, A.; Simon, H.A., 1972, *Human problem solving*. Prentice Hall, Englewood Cliffs, NJ.

Slowinski, K., 1992, ”Rough classification of HSV patients”, in *Intelligent Decision Support - Handbook of Applications and Advances of the Rough Set Theory*, R. Slowinski (*Ed*.), Kluwer Acad.Publ., pp. 77-93.

Wang, L.X.; Mendel, J.M., 1992, ”Generating fuzzy rules by learning from examples”. *IEEE Trans. System, Man, Cybernetics*, vol. 22, no. 6, pp.1414-1427.

APPENDIX: LIST OF MEDICAL PARAMETERS WHICH ARE THE INPUTS AND OUTPUTS OF THE DECISION SUPPORT SYSTEM OF SECTION 5

<u>A. Input (pre-operating) attributes:</u>

1. Sex. 2. Age. 3. Duration of disease. 4. Complication of ulcer.

5. HCL concentration [mmol HCL /100ml]. 6. Volume of gastric juice per 1h [ml]. 7. Volume of residual gastric juice [ml]. 8. Basic acid output (BAO) [mmol HCL / h].	basic secretion
9. HCL concentration [mmol HCL /100ml]. 10. Volume of gastric juice per 1h [ml]. 11. Maximal acid output [mmol HCL / h].	secretion stimulated by histamine

<u>B. Outputs - set of classes (so-called Visick grading [Slowinski, 1992]) according to long term results of the operation:</u>

Class 1. <u>Excellent:</u> absolutely no symptoms, perfect result.

Class 2. <u>Very good:</u> patient considers result perfect, but interrogation elicits mild occasional symptoms easily controlled by a minor adjustment of diet.

Class 3. <u>Satisfactory:</u> mild or moderate symptoms easily controlled by care, which cause some discomfort, but patient and surgeon are satisfied with result which does not interfere seriously with life or work.

Class 4. <u>Unsatisfactory:</u> moderate or severe symptoms of complications which interfere with work or normal life; patient or surgeon dissatisfied with result; includes all cases with recurrent ulcer and those submitted to further operation.