Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

A Comparison of Methodologies for Fuzzy Expert System Creation - Application to Arrhythmic Beat Classification

Markos G. Tsipouras, Themis P. Exarchos and Dimitrios I. Fotiadis, *Member, IEEE*

Abstract—**In this work, three different methodologies for fuzzy expert systems creation are compared: a well-known neuro-fuzzy approach, a knowledge-based approach and a novel methodology, based on rule-extraction. The adaptive neuro-fuzzy information system (ANFIS) is used to automatically generate a fuzzy expert system. In the knowledge-based approach and the rule-extraction methodology, the idea is to start with a model described by crisp rules, provided by medical experts in the first case or extracted using data mining techniques in the second, and then to transform them into a set of fuzzy rules, creating a fuzzy model. In either case, the adjustment of the model's parameters is performed via a stochastic global optimization procedure. All three approaches are applied to a medical domain problem, the cardiac arrhythmic beat classification. The ability to interpret the decisions made from the created fuzzy expert systems is a major advantage compared to other "black box" approaches.**

I. Introduction

Medical expert systems are a challenging field, requiring the synergy of different scientific areas. The representation of medical knowledge and expertise, the decision making in the presence of uncertainty and imprecision, the choice and adaptation of a suitable model, are some issues which a medical expert system should take under consideration. Uncertainty is traditionally treated in a probabilistic manner; recently, however, methods based on fuzzy logic have gained ground [1]. The model's parameter adaptation amounts to optimizing a properly constructed "error" function. Expert systems are a branch of artificial intelligence, which make extensive use of specialized knowledge to solve problems at the level of a human expert. This knowledge is represented by a set of rules [2]. An expert system's review of applications can be found in [3]. An expert system can be created by defining a crisp or fuzzy model (set of rules) and then optimizing its parameters to fit a given dataset. Several approaches have been proposed in the literature for the development of fuzzy or crisp models; in most of them the model is trained using a known optimization technique, i.e. fuzzy rules with genetic algorithms [4], fuzzy rules with simulated annealing [5], multicriteria decision analysis with genetic algorithms [6]. Neuro-fuzzy algorithms have also been proposed; the fuzzy rules are modelled using an artificial neural network (ANN) structure and popular training techniques are applied [7].

In this work, three different methodologies for fuzzy expert system creation are compared: a knowledge-based approach, a rule-extraction methodology and the well-known ANFIS approach [8]. All three are applied to a medical domain problem, the cardiac arrhythmic beat classification into four types of cardiac rhythms: normal sinus rhythm and three types of arrhythmia: ventricular flutter/fibrillation, ventricular tachycardia and 2^{o} heart block. The classification is performed using only the RR interval signal and features extracted from it. For the knowledge-based approach and the rule-extraction methodology the idea is similar: having an initial crisp model (i.e. a model comprised from crisp rules), transform it into a fuzzy model and then tune the parameters of this model using global optimization. The initial rules can be either defined by experts (knowledge-based approach) or automatically extracted using data mining techniques (rule-extraction methodology). In either case, the rules are initially represented using the monotonic crisp membership function in a disjunctive normal form (DNF). Then, the rules are transformed from crisp to fuzzy ones, using a fuzzy membership function and T and S norms. Using different fuzzy membership functions and alternative definitions for the T and S norms, several fuzzy models can be defined. Finally, the adjustment of the model's parameters is performed via stochastic global optimization. The ANFIS is also used to automatically generate a fuzzy expert system. The MIT-BIH arrhythmia database [9] is used for optimizing the fuzzy model's parameters and evaluating the fuzzy expert systems. Results, for all three approaches, along with qualitative and quantitative comparison, are presented.

II. Materials and Methods

A. Medical Background

Arrhythmia can be defined as any type of rhythm that deviates from the normal sinus rhythm (NSR). Automatic arrhythmic beat detection and classification, using the ECG and/or features extracted from it, is a critical task in clinical cardiology, especially when performed in real time.

Manuscript received April 3, 2006. This research is partially funded by the program "Heraklitos" of the Hellenic Ministry of Education. Also it is supported by the European Commission as part of the project NOESIS (IST-2002-507960).

M. G. Tsipouras, T. P. Exarchos and D. I. Fotiadis are with the Unit of Medical Technology and Intelligent Information Systems, Dept. of Computer Science, Univ. of Ioannina, Ioannina, Greece, GR 45110 (+30-26510-98803; markos@cs.uoi.gr, me01238@cc.uoi.gr, fotiadis@cs.uoi.gr).

Therefore several studies have been proposed in the literature, based on neural networks [10,12], autoregresive modelling [11], support vector machines [13], ECG morphology [14], rule-based systems [15]. Some of the studies are based on the analysis of the ECG signal, extracting features from it and using them for the detection and/or classification of cardiac beats. However, this is not always feasible due to: (a) the presence of noise making feature extraction difficult and in some cases impossible, and (b) the inability to perform in real time. An alternative would be to use only the RR-interval signal (tachogram) but, in this case, it is expected that only certain types of arrhythmias can be detected and classified. Also, most of the studies are "black box" approaches and, therefore, lack the ability of interpreting their decisions.

B. Dataset

The beat classification is performed using only the tachogram. Therefore, a QRS detection method, proposed by Tompkins [16] is used to detect the R waves and then the tachogram is formed measuring the time intervals between consecutive R waves. Then, a three RR-interval sliding window is used to create the dataset: $D=\{d^l,c^l\}$ with $d^l=[RR_1,RR_2,RR_3]^l$ the l^{th} three RR-interval window and c^l the class of the middle RR interval (RR_2). The class c^l is represented as $c^l\in\{0,1\}^4$, where, if d^l belongs to class i, then $c^l=e_i$. All beats from all records of the MIT-BIH arrhythmia database are used for the creation of the dataset D. Both rhythm and beat annotations from the database are used to specify the class, following the scheme: If RR_2 is annotated as ventricular flutter/fibrillation, then $c^l=[1,0,0,0]$, else if RR_2 is annotated as PVC then $c^l=[0,1,0,0]$, else if RR_2 belongs to 2° heart block episode, then $c^l=[0,0,0,1]$, else $c^l=[0,0,1,0]$.

C. The knowledge-based approach

Initial set of rules: Three rules, provided by medical experts, comprise the initial set of rules:

Rule 1: Ventricular flutter/fibrillation (VF). If one of the following conditions is true then the middle RR interval (RR_2) of the window is classified into the VF category:

$E_{1,1}:(RR_1<\theta_{1,1})\wedge(RR_2<\theta_{1,2})\wedge(RR_3<\theta_{1,3})$,

$E_{1,2}:RR_1+RR_2+RR_3<\theta_{1,4}$.

Rule 2: Premature ventricular contractions (PVC). If one of the following conditions is true then the middle RR interval of the window is classified as PVC:

$E_{2,1}:(RR_1/RR_2>\theta_{2,1})\wedge(RR_3/RR_2>\theta_{2,2})$,

$E_{2,2}:(RR_3/RR_1>\theta_{2,3})\wedge(RR_1/RR_2>\theta_{2,4})$,

$$E_{2,3}:(|RR_1-RR_2|<\theta_{2,5})\wedge(RR_2<\theta_{2,6})\wedge\left(\frac{2RR_3}{RR_1+RR_2}<\theta_{2,7}\right)$$

$$E_{2,4}:(|RR_2-RR_3|<\theta_{2,8})\wedge(RR_2<\theta_{2,9})\wedge\left(\frac{2RR_1}{RR_2+RR_3}<\theta_{2,10}\right)$$

Rule 3: 2° heart block (BII). If one of the following conditions is true then the middle RR interval of the window is classified into BII category:

$E_{3,1}:(RR_2\in[\theta_{3,1},\theta_{3,2}])\wedge(|RR_1-RR_2|<\theta_{3,3})$,

$E_{3,2}:(RR_2\in[\theta_{3,1},\theta_{3,2}])\wedge(|RR_2-RR_3|<\theta_{3,3})$.

The crisp rules $(R_1(d^l,\theta_1)$, $R_2(d^l,\theta_2)$ and $R_3(d^l,\theta_3))$, which comprise the crisp model, in DNF representation are: $R_1(d^l,\theta_1)=E_{1,1}\vee E_{1,2}$, $R_2(d^l,\theta_2)=E_{2,1}\vee E_{2,2}\vee E_{2,3}\vee E_{2,4}$ and $R_3(d^l,\theta_3)=E_{3,1}\vee E_{3,2}$, with $\theta_1=\{\theta_{1,j}\}, j=1,..,4$, $\theta_2=\{\theta_{2,j}\}, j=1,..,10$ and $\theta_3=\{\theta_{3,j}\}, j=1,2,3$. If all rules are false, then the beat is classified as normal. If more than one rules are true, then the beat is considered as unclassified.

Fuzzyfication of the set of rules: To transform the crisp rules into fuzzy, the sigmoid function, defined as: $g(x,a,b)=\left(1+e^{a(b-x)}\right)^{-1}$ is used as the fuzzy membership function, the T_{norm} is the *algebraic product*: $T_{norm}(a_1,a_2)=a_1a_2$, and the S_{norm} is the *probabilistic OR*: $S_{norm}(a_1,a_2)=a_1+a_2-a_1a_2$. The fuzzy rules $(R_1^f(d^l,\theta_1)$, $R_2^f(d^l,\theta_2)$ and $R_3^f(d^l,\theta_3))$ for the arrhythmia classification problem, are defined as:

$$R_1^f(d^l,\theta_1)=S_{norm}\left(\begin{array}{c}T_{norm}\left(\begin{array}{l}g(RR_1,\theta_{1,1}^a,\theta_{1,1}^b),\\ g(RR_2,\theta_{1,2}^a,\theta_{1,2}^b),\\ g(RR_3,\theta_{1,3}^a,\theta_{1,3}^b)\end{array}\right),\\ \left(g(RR_1+RR_2+RR_3,\theta_{1,4}^a,\theta_{1,4}^b)\right)\end{array}\right),$$

$$R_2^f(d^l,\theta_2)=$$

$$S_{norm}\left(\begin{array}{l}T_{norm}\left(g(RR_1/RR_2,\theta_{2,1}^a,\theta_{2,1}^b),g(RR_3/RR_2,\theta_{2,2}^a,\theta_{2,2}^b)\right),\\ T_{norm}\left(g(RR_3/RR_1,\theta_{2,3}^a,\theta_{2,3}^b),g(RR_1/RR_2,\theta_{2,4}^a,\theta_{2,4}^b)\right),\\ T_{norm}\left(\begin{array}{l}g(|RR_1-RR_2|,\theta_{2,5}^a,\theta_{2,5}^b),g(RR_2,\theta_{2,6}^a,\theta_{2,6}^b),\\ g(2RR_3/(RR_1+RR_2),\theta_{2,7}^a,\theta_{2,7}^b)\end{array}\right),\\ T_{norm}\left(\begin{array}{l}g(|RR_2-RR_3|,\theta_{2,8}^a,\theta_{2,8}^b),g(RR_2,\theta_{2,9}^a,\theta_{2,9}^b),\\ g(2RR_1/(RR_2+RR_3),\theta_{2,10}^a,\theta_{2,10}^b)\end{array}\right)\end{array}\right),$$

$$R_3^f(d^l,\theta_3)=S_{norm}\left(\begin{array}{l}T_{norm}\left(\begin{array}{l}g(RR_2,\theta_{3,1}^a,\theta_{3,1}^b),g(RR_2,\theta_{3,2}^a,\theta_{3,2}^b),\\ g(|RR_1-RR_2|,\theta_{3,3}^a,\theta_{3,3}^b)\end{array}\right),\\ T_{norm}\left(\begin{array}{l}g(RR_2,\theta_{3,1}^a,\theta_{3,1}^b),g(RR_2,\theta_{3,2}^a,\theta_{3,2}^b),\\ g(|RR_2-RR_3|,\theta_{3,3}^a,\theta_{3,3}^b)\end{array}\right)\end{array}\right),$$

with $\theta_1 = \left\{\theta^a_{1,j}, \theta^b_{1,j}\right\}, j = 1,..,4$, $\theta_2 = \left\{\theta^a_{2,j}, \theta^b_{2,j}\right\}, j = 1,...,10$ $j = 1,...,7$ and $\theta_3 = \left\{\theta^a_{3,j}, \theta^b_{3,j}\right\}$, $j = 1,...,3$.

The final decision is made as follows: if the maximum value of the results of the three rules $\left(m = max\left(R^f_1\left(d^l, \theta_1\right), R^f_2\left(d^l, \theta_2\right), R^f_3\left(d^l, \theta_3\right)\right)\right)$ is less or equal to θ_4 then d^l is classified as normal sinus rhythm. If m is more than θ_4 then the d^l is classified in the category of the rule that has the maximum result:

$$M\left(d^l, \Theta\right) = \begin{cases} [0,0,0,1] & \text{if } m = R_1\left(d^l, \theta_1\right) > \theta_4 \\ [0,0,1,0] & \text{if } m = R_2\left(d^l, \theta_2\right) > \theta_4 \\ [0,1,0,0] & \text{if } m \le \theta_4 \\ [1,0,0,0] & \text{if } m = R_3\left(d^l, \theta_3\right) > \theta_4 \end{cases},$$

where $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$.

Optimization: The fuzzy model $M\left(d^l, \Theta\right)$ is optimized with respect to Θ using a training dataset (D_{train}), which is a randomly selected subset of D, containing 250 patterns from each class, and the HTMLSL [17] optimization method. The mean square error cost function is used:

$$f\left(D_{train}, \Theta\right) = \frac{1}{\left|D_{train}\right|} \sum_{l=1}^{1000} \left\| M\left(d^l, \Theta\right) - c^l \right\|^2 .$$

D. The rule-extraction methodology

In the rule-extraction methodology, an initial set of rules is automatically extracted from a decision tree, induced from the D_{train}. Functions of the three RR intervals (used in the knowledge-based model) are included in the feature vector $\left(RR_1/RR_2\right.$, RR_3/RR_1, RR_3/RR_2, $\left|RR_1 - RR_2\right|$, $\left|RR_2 - RR_3\right|$, $2RR_3/\left(RR_1 + RR_2\right)$, $2RR_1/\left(RR_2 + RR_3\right)$, $\left.RR_1 + RR_2 + RR_3\right)$ with a total of 11 features. The construction of the decision tree is implemented using the C4.5 inductive algorithm [18]. Each internal node of the tree corresponds to a feature, while each outgoing branch corresponds to the range of values of that component. The leaf nodes represent the class to be assigned to a sample. The C4.5 algorithm applies to a set of data and generates a decision-tree, which minimizes the expected value of the number of tests for the classification of the data. The most important factor in the C4.5 algorithm is its ability to automatically select the feature which is appropriate at each node. The feature of each node is selected in order to divide input samples effectively. The *information gain* is used as a measure of effectiveness.

After the induction of the decision tree, a pruning method is applied to reduce the tree's size and complexity. There are two common methods for pruning: prepruning and post-pruning. Post-pruning tends to give better results than prepruning since it makes pruning decisions based on a fully grown tree, unlike prepruning, which can suffer from early termination of the tree growing process [18]. Post-pruning is performed by replacing a subtree with a new leaf node whose class label is determined from the majority class of records associated with the subtree (subtree replacement). The subtree replacement was performed by calculating the pessimistic error. The confidence value was set to 25% and the minimum number of records per leaf was set to 2.

The crisp set of rules is created from the final decision tree as follows: (a) One condition is created for every leaf of the tree, by parsing the tree from the root node to a leaf. The tests encountered along the path form the conjuncts of the condition while the class label at the leaf node is assigned to the rule consequent: $E_i\left(A, \Theta\right) = c_{root}\left(a_{root}, \theta_{root}\right) \wedge c_{n_j}\left(a_{n_j}, \theta_{n_j}\right) \wedge ... \wedge c_{n_k}\left(a_{n_k}, \theta_{n_k}\right)$, where $Cond_i$ is a condition, $A = \left\{a_1, a_2, ..., a_{n_f}\right\}$ is the feature vector, $\Theta = \left\{\theta_1, \theta_2, ..., \theta_{n_f}\right\}$ is a vector containing all thresholds, n_f is the number of features characterizing a record, a_j and θ_j are the feature and the threshold used in the conjunct j and $c_j\left(a_j, \theta_j\right)$ is the respective conjunct. (b) A general rule is created for each class, using all the conditions $Cond_i(A,\Theta)$ having as consequent this class: $R_y\left(A, \Theta\right) = Cond_{j_1}\left(A, \Theta\right) \vee Cond_{j_2}\left(A, \Theta\right) \vee ... \vee Cond_{j_n}\left(A, \Theta\right)$, where y is the class. These general rules comprise the crisp set of rules, which is exhaustive and mutually exclusive. Therefore, for each feature vector A, one of the general rules is true, thus defining its class. After the creation of the initial crisp set of rules, the fuzzyfication and optimization steps are followed.

E. ANFIS

The neuro-fuzzy techniques have emerged from the fusion of artificial neural networks and fuzzy inference systems and form a popular framework for solving real world problems. ANFIS is based on a Sugeno type fuzzy system. The feature vector, described above, is used as input and the learning algorithm is a hybrid method, proposed in [19].

III. Results

The dataset includes 484 beats annotated as VF, 6,183 beats annotated as PVC, 102,793 beats annotated as N and 420 beats annotated as BII, totally. For each methodology, the normalized confusion matrix and the corresponding accuracy are shown in Table I. The accuracy of the three produced fuzzy expert systems is: 90.24% for the knowledge-based approach, 94.13% for the ANFIS and 95.98% for the rule- extraction methodology.

IV. Discussion

Three methodologies are compared, to create fuzzy expert systems for the arrhythmic beat classification problem: a knowledge-based approach, the ANFIS and a novel one

TABLE I
NORMALIZED CONFUSION MATRIX AND ACCURACY FOR EACH METHODOLOGY(%).

Methodology		Database VF	PVC	N	BII	Accuracy
Knowledge-based approach	VF	89.87	1.08	0.43	0	90.24
	PVC	0.36	81.54	6.79	0.24	
	N	0.5	17.08	92.57	2.78	
	BII	0	0	0.16	96.98	
	Uncl	9.27	0.3	0.05	0	
Rule-extraction methodology	VF	98.9	1.44	0.65	0	95.98
	PVC	1.1	92.4	5.32	0.48	
	N	0	6.16	93.61	0.51	
	BII	0	0	0.42	99.01	
ANFIS	VF	98.2	1.82	0.7	0	94.13
	PVC	1.8	90.7	9.83	1.67	
	N	0	7.48	89.31	0	
	BII	0	0	0.16	98.33	

based on rule-extraction. The later, initially extracts several rules from a dataset, creating thus, a set of rules. The set of rules is then fuzzyfied and, finally, the parameters entering the model (Θ) are tuned using global optimization. Compared to the ANFIS methodology, the proposed method exhibits several advantages: (a) optimization is used in both the data mining stage and in the fuzzy model, therefore, the final expert system is better tuned and thus having better results; (b) the fuzzy model is generated fuzzyfing an initial crisp set of rules, thus it is simpler than the respective model generated from ANFIS, and its optimization is faster. On the other hand, the initial data mining stage, not existing in the ANFIS, increases the computational cost. Compared to the knowledge-based approach, both the ANFIS and the proposed methodology, are data-driven and, thus, they do not require a-priori knowledge but the quality of the produced expert system depends on the quality of the initial dataset. Also the results presented in Table I indicate that the proposed method has the best results, with respect to the arrhythmic beat classification problem.

Table II presents several methods proposed in the along with the reported accuracy. All methods have accuracy in the range of 90% to 98%. The methods of [10,13] are based on "black box" approaches, such as neural networks and support vector machines. Therefore, there is no interpretation for their decisions. Some of the proposed methods have been tested on small subsets of the MIT-BIH arrhythmia database [10,11,13]. In [14] ECG morphology is utilized, which is an approach not feasible in cases of high noise. In the proposed methodology, only QRS detection was performed, on the ECG signal. The results were obtained using the MIT-BIH arrhythmia database for evaluation, while for each decision made, interpretation is available.

TABLE II
SUMMARY OF METHODS FOR ARRHYTHMIC BEAT CLASSIFICATION

Authors	Method	Dataset (beats)	Accuracy (%)
Osowski & Linh [10]	cumulants of the second, third and fourth order and fuzzy hybrid neural network	7,185	96.06
Ge et al. [11]	autoregressive modelling	856	96.84
Osowski et al. [13]	support vector machines	12,785	95.91
Chazal et al. [14]	ECG morphology and linear discriminates	100,000	97.50
Tsipouras et al. [15]	knowledge-based system	109,880	94.26
this work	Knowledge-based approach	109,880	90.24
	Rule-extraction methodology		95.98
	ANFIS		94.13

REFERENCES

[1] L.H. Tsoukalas and R.E. Uhrig, *Fuzzy and neural approaches in engineering*, John Wiley & Sons, 1997.
[2] J. Giartano and G. Riley. *Expert systems, Principles and programming*, 3rd eds., PWS Publishing Company, 1998.
[3] S.H. Liao, "Expert system methodologies and applications-a decade review from 1995 to 2004", *Expert Systems with Applications*, vol. 28, pp. 93-103, 2005.
[4] W. Pedrycz and J.V. de Oliveira, "An algorithmic framework for development and optimization of fuzzy models", *Fuzzy sets and systems,* vol. 80, pp. 37-55, 1996.
[5] J.M. Garibalti and E.I. Ifeachor, "Application of simulated annealing fuzzy model tuning to umbilical cord acid-base interpretation", *IEEE Trans. Fuzzy Systems,* vol. 7, pp. 72-84, 1999.
[6] Y. Goletsis, C. Papaloucas, D.I. Fotiadis, A. Likas and L.K. Michalis, "Automated ischemic beat classification using genetic algorithms and multicriteria decision analysis", *IEEE Trans. Biom. Eng.,* vol. 51, pp. 1717-1725, 2004.
[7] R. Babuska and H. Verbruggen, "Neuro-fuzzy methods for nonlinear system identification", *Ann. Rev. Control*, vol. 27, pp.73-85 2003.
[8] J. S. Jang, C. T. Sun and E. Mizutani, *Neuro fuzzy and soft computing*, Matlab curriculum series, Prentice Hall, Upper Saddle River, USA, 1997.
[9] Harvard-MIT Division of Health Sciences and Technology, *MIT-BIH Arrhythmia Database CD-ROM*, 3rd eds., 1997.
[10] S. Osowski and T.H. Linh, "ECG beat recognition using fuzzy hybrid neural network", *IEEE Trans Biom Eng*, vol. 48 pp.1265-71, 2001.
[11] D. Ge, N. Srinivasan and S.M. Krishnan, "Cardiac arrhythmia classification using autoregressive modering", *Biomed. Eng. OnLine*, 1:5, 2002.
[12] U.R. Acharya, P.S. Bhat, S.S. Iyengar, A. Rao and S. Dua, "Classification of heart rate data using artificial neural network and fuzzy equivalence relation", *Pat. Recognition*, vol. 36, pp. 61-8 2003.
[13] S. Osowski T.H. Linh and T. Markiewicz, "Support vector machine-based expert system for reliable heartbeat recognition", *IEEE Trans. Biom. Eng.*, vol. 51, pp. 582-589, 2004.
[14] P. de Chazal, M. O'Dwyer and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features", *IEEE Trans. Biom. Eng.*, vol. 51, pp. 1196-1206, 2004.
[15] M.G. Tsipouras, D.I. Fotiadis and D. Sideris, "An arrhythmia classification system based on the RR interval signal", *Artificial Intelligence in Medicine*, vol. 33, pp. 237-250, 2005.
[16] P.S. Hamilton and W.J. Tompkins, "Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database", *IEEE Trans. Biom. Eng.*, vol. 33, pp. 1157-1165, 1986.
[17] F.V. Theos, I.E. Lagaris, D.G. Papageorgiou, "PANMIN: sequential and parallel global optimization procedures with a variety of options for the local search strategy", *CPC*, vol. 159, pp. 63–69, 2004.
[18] J.R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kauffman, USA, 1993.
[19] I. Guler, E.D. Ubeyli, "Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelets coefficients", *J Neuroscience Methods*, vol. 148, pp.113-121, 2005.