Proceedings of the 29th Annual International
Conference of the IEEE EMBS
Cité Internationale, Lyon, France
August 23-26, 2007.

Integration of global and local knowledge for fuzzy expert system creation - Application to Arrhythmic Beat Classification

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

Abstract—**In this work, we propose a method for the automated expert system creation. The method is based on the integration of global knowledge (i.e. knowledge from the field experts) and local knowledge (i.e. knowledge derived from the available data) in a single inference engine. Starting from an initial set of rules (expert's knowledge) and an annotated dataset, data mining is performed to the dataset and a second set of rules is acquired. Both of them are integrated into a single set of rules. Fuzzy modeling is then applied to the rules, transforming them into a fuzzy model, and finally, an optimization technique is used to tune the fuzzy model's parameters. The method is applied to a medical domain problem, the cardiac arrhythmic beat classification and satisfactory results have been obtained. The method experiences several advantages compared to approaches based solely on expert's knowledge or mined knowledge while the ability to interpret the decisions made from the created fuzzy expert system is a major advantage compared to "black box" approaches.**

I. Introduction

Expert systems (ESs) are a branch of artificial intelligence that makes extensive use of specialized knowledge to solve problems at the level of a human expert. This knowledge is represented by a set of crisp or fuzzy rules [1], which formulate a crisp or fuzzy inference engine; this engine is the core of the ES. An alternative methodology for creating an ES is the use of crisp or fuzzy rules extracted directly from the data [2,3].

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]. Also, a great effort has been made in the induction of decision trees using fuzzy partitions (fuzzy decision trees) and optimization of the parameters entering these trees [8-11]. In all the above research attempts, it has been shown that fuzzy decision trees and fuzzy rules, after the optimization of the parameters used improve the accuracy of the respective crisp models significantly.

An area where expert systems are widely employed is the medical domain. Several parameters must be taken into consideration in order to create a medical expert system; the representation of medical knowledge and expertise, the decision making, and the choice and adaptation of a suitable model, are some of them. Also, uncertainty and imprecision, inherited in medical problems, can be treated by incorporating fuzzy logic [12].

The generation of the inference engine is based either on experts' knowledge or on knowledge derived from data mining techniques; both these approaches suffer from several disadvantages. An ES based solely on experts' knowledge does not take under consideration specific properties of the dataset. In addition, experts' knowledge is available in limited research domains. On the other hand, ESs based on knowledge discovery approaches do not take advantage of existing knowledge and tend to identify spurious relationships in the dataset. Merging these two sources of knowledge may eliminate their disadvantages resulting to a more accurate ES.

In this work we propose a method for the creation of a fuzzy expert system (FES), by integrating global and local knowledge. The experts' knowledge, represented by a set of crisp rules, is combined with knowledge derived from an annotated dataset, using a data mining technique, which is also represented by a set of crisp rules. The first set of rules reflects the global knowledge accumulated for a domain of application while the second is more adjusted to the specific dataset. The integrated set of rules is fuzzyfied and thus a fuzzy model is generated. This fuzzy model contains several parameters, which are adjusted using optimization techniques to fit a given dataset. The methodology is applied to the cardiac arrhythmic beat classification into four types of cardiac rhythms: normal sinus rhythm and three types of arrhythmia: ventricular flutter/fibrillation (VF), premature ventricular contractions (PVC) and 2^o heart block (BII). The classification is performed using only the *RR* interval signal and features extracted from it. The MIT-BIH arrhythmia database [13] is used for the evaluation of the created FES.

Manuscript received March 28, 2007. This research is funded by the program "Heraklitos" of the Operational Program for Education and Initial Vocational Training of the Hellenic Ministry of Education under the 3rd Community Support Framework and the European Social Fund.

M.G. Tsipouras and D.I. Fotiadis are with Unit of Medical Technology and Intelligent Information Systems, Dept. of Computer Science, University of Ioannina, Ioannina, Greece, GR 45110 (+30-26510-98803; fax: +30-26510-98889; e-mail: markos@cs.uoi.gr, fotiadis@cs.uoi.gr).

T.P. Exarchos is with Dept. of Medical Physics, Medical School, University of Ioannina, Ioannina, Greece, GR 45110 (e-mail: me01238@cc.uoi.gr).

II. Materials and Methods

A. Medical Background

Arrhythmia can be defined as any type of rhythm that deviates from the normal sinus rhythm. 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. Therefore several studies have been proposed in the literature, based on neural networks [14,15], autoregressive modelling [16], support vector machines [17], ECG morphology [18] and rule-based systems [19]. 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 (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, several studies are "black box" approaches and, therefore, lack the ability of interpreting their decisions.

B. Dataset

The MIT-BIH arrhythmia database [13] is used for the creation of the dataset. QRS detection is used to detect the *R* waves and then the tachogram is formed measuring the time intervals between consecutive *R* waves. A three *RR* interval sliding window, is used (RR_1, RR_2, RR_3) as well as functions of those intervals. These functions provide useful information of the non-linear relations between the three consecutive *RR* intervals, related to specific cardiac rhythm patterns. The functions have been proposed by expert cardiologists [19]. Thus, the dataset is defined as $D=\{d^l, c^l\}$ with $d^l = (RR_1,\ RR_2,\ RR_3,\ RR_1+RR_2+RR_3,\ RR_1/RR_2,\ RR_3/RR_1,\ RR_3/RR_2,\ |RR_1-RR_2|,\ |RR_2-RR_3|,\ 2RR_3/(RR_1+RR_2),\ 2RR_1/(RR_2+RR_3))$, the l^{th} feature vector and c^l the class of the RR_2 interval, represented as $c^l \in \{1,2,3,4\}$. Both rhythm and beat annotations from the database are used to specify the class: if RR_2 is annotated as VF, then $c^l = 1$, else if RR_2 is annotated as PVC then $c^l = 2$, else if RR_2 belongs to BII episode, then $c^l = 4$, else $c^l = 3$. Table I presents the dataset and the number of beats belonging to each class.

TABLE I
Number of Beats in Used for Training and Testing the Method in Arrhythmic Beat Classification.

Classes	D	D_{train}	D_{test}
VF	484	300	184
PVC	6183	300	5883
N	102793	300	102493
BII	420	300	120
Overall	109880	1200	108680

C. Global knowledge

Three rules, provided by medical experts, comprise the first 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:

$$E^G_{1,1}: (RR_1 < \theta^G_{1,1}) \wedge (RR_2 < \theta^G_{1,2}) \wedge (RR_3 < \theta^G_{1,3}),$$

$$E^G_{1,2}: RR_1 + RR_2 + RR_3 < \theta^G_{1,4}.$$

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:

$$E^G_{2,1}: (RR_1/RR_2 > \theta^G_{2,1}) \wedge (RR_3/RR_2 > \theta^G_{2,2}),$$

$$E^G_{2,2}: (RR_3/RR_1 > \theta^G_{2,3}) \wedge (RR_1/RR_2 > \theta^G_{2,4}),$$

$$E^G_{2,3}: (|RR_1 - RR_2| < \theta^G_{2,5}) \wedge (RR_1 < \theta^G_{2,6}) \wedge (RR_2 < \theta^G_{2,7}) \wedge (RR_1 + RR_2)/2RR_3 < \theta^G_{2,8}$$

$$E^G_{2,4}: (|RR_2 - RR_3| < \theta^G_{2,9}) \wedge (RR_2 < \theta^G_{2,10}) \wedge (RR_3 < \theta^G_{2,11}) \wedge (RR_2 + RR_3)/2RR_1 < \theta^G_{2,12}$$

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:

$$E^G_{4,1}: (RR_2 \in [\theta^G_{4,1}, \theta^G_{4,2}]) \wedge (|RR_1 - RR_2| < \theta^G_{4,3}),$$

$$E^G_{4,2}: (RR_2 \in [\theta^G_{4,4}, \theta^G_{4,5}]) \wedge (|RR_2 - RR_3| < \theta^G_{4,6}).$$

All parameters θ^G (thresholds) were defined by experts. If none of the above rules fires for a record then it is considered as normal sinus rhythm (N).

D. Local knowledge

Data mining is applied in order to extract a set of crisp rules from the annotated data. In our approach we used decision trees (DTs), however any rule mining technique could be employed. The construction of the DT is implemented using the C4.5 inductive algorithm [20]. We followed the pessimistic error based post-pruning method to reduce the size of the DT and its complexity. The minimum number of records in a leaf was set to 5% of the total records included in the training set (D_{train}, presented in Table I). This ensures that the generated DT will include rules with high coverage and thus will be simple, reducing the complexity to the following stages (fuzzyfication, optimization). The final tree is parsed from the root to each leaf in order to construct the second set of rules:

$$E^L_{1,1}: RR_2 \le \theta^L_{1,1} \wedge (RR_1 + RR_2 + RR_3) \le \theta^L_{1,2},$$

$$E^L_{2,1}: RR_2 \le \theta^L_{2,1} \wedge (RR_1 + RR_2 + RR_3) > \theta^L_{2,2} \wedge RR_3/RR_1 > \theta^L_{2,3},$$

$$E^L_{3,1}: RR_2 \le \theta^L_{3,1} \wedge (RR_1 + RR_2 + RR_3) > \theta^L_{3,2} \wedge RR_3/RR_1 \le \theta^L_{3,3},$$

$$E^L_{4,1}: RR_2 > \theta^L_{4,1}.$$

All parameters θ^L (thresholds) correspond to the cut points of the features in the DT.

E. Fuzzyfication and integration of the rules

Both sets of rules are fuzzyfied, using the sigmoid function: $g(x,a,b)=\left(1+e^{a(b-x)}\right)^{-1}$ as fuzzy membership function, where x is the input feature and a,b are parameters, and T and S norms, defined as the *min* and the *max* operators [21], respectively. One fuzzy rule is created for each class, combining all fuzzy rules (either from the experts or from the DT) concerning this class:

$R_1^f\left(d^l,\theta_1\right)=S_{norm}\left(FE_{1,1}^G,FE_{1,2}^G,FE_{1,1}^L\right),$

$R_2^f\left(d^l,\theta_2\right)=S_{norm}\left(FE_{2,1}^G,FE_{2,2}^G,FE_{2,3}^G,FE_{2,4}^G,FE_{2,1}^L\right),$

$R_3^f\left(d^l,\theta_3\right)=S_{norm}\left(\theta_{3,1}^G,FE_{3,1}^L\right),$

$R_4^f\left(d^l,\theta_4\right)=S_{norm}\left(FE_{4,1}^G,FE_{4,2}^G,FE_{4,1}^L\right),$

where $FE_{i,j}^X$ is the fuzzy expression resulted from the $E_{i,j}^X$ crisp expression, using the fuzzy membership function and the T and S norms definitions. θ_i is a vector containing all parameters entering the i^{th} fuzzy rule. The R_3^f rule is based on $FE_{3,1}^L$ fuzzy expression and on $\theta_{3,1}^G$ parameter (which is also included in θ_3). This parameter is used due to the fact that there is no expert's rule for N category. Based on the fuzzy rules, the fuzzy model is defined as: $M\left(d^l,\Theta\right)=arg\max_{i=1,4}\left(R_i^f\left(d^l,\theta_i\right)\right)$, where $\Theta=\{\theta_1,\theta_2,\theta_3,\theta_4\}$ is a vector containing all parameters used in the model.

F. Optimization of the fuzzy model

The fuzzy model $M\left(d^l,\Theta\right)$ is optimized with respect to Θ using the D_{train} dataset (Table I). The cost function, used for the optimization, is defined as: $F\left(\Theta,D_{train}\right)=1-trace\left(X_{i,i}\right)/\left|D_{train}\right|$, where X is the normalized confusion matrix, calculated for D_{train} as:

$$X_{M\left(d^l,\Theta\right),c^l}=\frac{\#\ of\ patterns\ in\ c^l\ classified\ to\ M\left(d^l,\Theta\right)}{total\ \#\ of\ patterns\ in\ c^l}.$$

For the optimization the Direct [22] global optimization method was employed.

III. Results

In order to evaluate our method, D_{test} dataset is used (Table I), which includes the remaining records of the original dataset after the selection of D_{train}. Results are obtained for four different settings: (i) crisp global knowledge, (ii) crisp local knowledge, (iii) crisp integrated global and local knowledge and (iv) fuzzy and optimized integrated global and local knowledge. The normalized confusion matrix along with the accuracy, for all different settings, are shown in Table II. The best results are obtained when fuzzy and optimized integrated global and local knowledge is used (93.24%).

TABLE II
NORMALIZED CONFUSION MATRIX AND ACCURACY (%).

Methodology		Database				Accuracy
		VF	PVC	N	BII	
Global knowledge (expert's rules)	VF	77.17	0.49	0.01	0	88.70
	PVC	22.28	93.66	14.27	0	
	N	0.54	5.85	85.64	1.67	
	BII	0	0	0.08	98.33	
Local knowledge (mined rules)	VF	96.2	1.97	1.09	0	92.74
	PVC	2.72	83.32	5.41	0	
	N	1.09	14.7	93.11	1.67	
	BII	0	0	0.64	98.33	
Integration (global and local knowledge)	VF	94.02	1.36	0.42	0	92.48
	PVC	4.89	84.77	6.14	0	
	N	1.09	13.87	92.8	1.67	
	BII	0	0	0.64	98.33	
Fuzzy, optimized integration	VF	95.65	1.46	0.46	0	93.24
	PVC	4.35	91.48	11.7	0	
	N	0	7.05	87.47	1.67	
	BII	0	0	0.37	98.33	

Further evaluation was performed to the fourth setting (fuzzy and optimized integrated global and local knowledge) in order to find out which set of fuzzy rules (the ones generated from global or local knowledge) fired, regarding the class of the beat. Table III presents the respective results.

TABLE III
NUMBER OF BEATS CLASSIFIED AND CORRECT CLASSIFICATIONS.

Set of fuzzy rules		Class			
		VF	PVC	N	BII
Based on global knowledge	classified	180	5310	95192	3
	correct	176	4817	89072	1
Based on local knowledge	classified	4	573	7301	117
	correct	0	565	583	117

IV. Discussion

A novel method for FES creation has been presented. Global and local knowledge, which are represented in the form of rules, are fuzzyfied and then are integrated into a single fuzzy model. The parameters of the fuzzy model are optimized, using global optimization. The proposed method has been evaluated in the classification of cardiac beats and satisfactory results have been reported.

The proposed method offers several advantages; the integration of both global and local knowledge ensures that the final FES combines the knowledge accumulated for a domain of application as long as knowledge on the dataset. In addition the incorporation of global knowledge makes the final FES not greatly affected from the quality of the given annotated dataset, which is a major limitation of all data-driven methodologies [23]. Furthermore, medical experts tend to rely more on decision support systems that are able to provide interpretation and contain a substantial amount of expert's knowledge. Finally, the method can be used in research areas where limited knowledge and limited data are available, combining both of these into a single FES.

The best accuracy was reported using the optimized fuzzy model with the integration of global and local knowledge. This setting is advantageous since it combines the predictions of both sets of rules and increases their accuracy. Also, as it is shown in Table II, the employment of the fuzzyfication and the optimization steps increases the accuracy of the classifier that employs the integration of the crisp sets of rules. The results presented in Table III, demonstrate that each set of fuzzy rules more accurately classifies specific classes. The set of fuzzy rules based on the global knowledge has very accurate results in the VF category (97.78%) and very good on the N (93.57%) and PVC (90.72%) categories, while the set of rules based on local knowledge is very accurate in BII (100%) and on PVC (98.6%) categories. This clearly demonstrates that specific instances of the dataset are covered by the global knowledge on the problem, while for several others the use of local knowledge is mandatory.

Table IV presents several methods proposed in the literature for arrhythmic beat classification. All methods have accuracy in the range of 90% to 97.5%. The methods of [14,17] 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 [14,16,17]. In [18] the ECG morphology is utilized, which is an approach not feasible in cases of noisy signals. In the proposed methodology, only QRS detection was performed, on the ECG signal, while for each decision made, interpretation is available.

A drawback of the proposed method is that a substantial amount of both expert's knowledge and annotated data should be available in order to be applied. Therefore, the method is not applicable in domains lacking one of the above prerequisites. Further improvement might focus on the employment of more sophisticated integration strategies, since in the current work a straightforward approach was used. Also, the fuzzyfication step was performed using a simple approach for the fuzzy membership function and the definitions of the T and S norms, since the main goal of this work is the integration of global and local knowledge in a single FES; the employment of more advanced methods for the fuzzyfication step can further improve the reported results.

TABLE IV
SUMMARY OF METHODS FOR ARRHYTHMIC BEAT CLASSIFICATION

Authors	Method	Dataset (beats)	Accuracy (%)
Osowski & Linh [14]	cumulants of the second, third and fourth order and fuzzy hybrid neural network	7,185	96.06
Ge et al. [16]	autoregressive modelling	856	96.84
Osowski et al. [17]	support vector machines	12,785	95.91
Chazal et al. [18]	ECG morphology and linear discriminates	100,000	97.50
Tsipouras et al. [19]	knowledge-based system	109,880	94.26
this work	Integrated knowledge	109,880	92.48
	Fuzzy, optimized integrated knowledge		93.24

REFERENCES

[1] J. Giartano and G. Riley. *Expert systems, Principles and programming*, 3rd eds., PWS Publishing Company, 1998.
[2] J. Abonyi and J.A. Roubos, "Structure identification of fuzzy classifiers", *5th online world conference on soft computing in industrial applications*, Internet, September 2000.
[3] J.A. Abonyi, J.A. Roubos, F. Szeifert, "Data-driven generation of compact, accurate, and linguistically-sound fuzzy classifiers based on a decision tree initialization, *International journal of approximate reasoning*, pp. 1-21, Jan. 2003.
[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. Papaloukas, 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] C.Z.Janikow, "Fuzzy decision trees: issues and methods", *IEEE Trans. Systems. Man and Cybernet.,* vol. 28 (1), pp. 1–14, 1998.
[9] C. Olaru and L. Wehenkel, "A complete fuzzy decision tree technique", *Fuzzy sets and systems*, vol. 138, pp. 221-254, 2003.
[10] W. Pedrycz and Z.A Sosnowski, "Genetically Optimized Fuzzy Decision Trees", *IEEE Transactions on systems, man and cybernetics-part B, cybernetics*, vol. 35, pp. 633-641, 2005.
[11] K. Crockett, Z. Bandar, J. O'Shea and D. Mclean, "On Constructing a Fuzzy Inference Framework using Crisp Decision Trees", *Fuzzy Sets and Systems*, vol. 157, pp. 2809– 2832, 2006.
[12] M. Akay, *Nonlinear biomedical signal processing vol 1: Fuzzy logic, Neural networks and New algorithm*, IEEE Press Series on Biomedical Engineering, New York, USA, 2000.
[13] Harvard-MIT Division of Health Sciences and Technology, *MIT-BIH Arrhythmia Database CD-ROM*, 3rd eds., 1997.
[14] S. Osowski and T.H. Linh, "ECG beat recognition using fuzzy hybrid neural network", *IEEE Trans Biom Eng*, vol. 48 pp.1265-71, 2001.
[15] 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.
[16] D. Ge, N. Srinivasan and S.M. Krishnan, "Cardiac arrhythmia classification using autoregressive modering", *Biomed. Eng. OnLine*, 1:5, 2002.
[17] 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.
[18] 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.
[19] 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.
[20] J.R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kauffman, USA, 1993.
[21] Wang LX. A course in fuzzy systems and control, Prentice-Hall, 1986.
[22] J.M. Gablonsky and C.T. Kelley, "A locally-biased form of the direct algorithm", *Journal of Global Optimization*, vol. 21, pp. 27-37, 2001.
[23] M.G. Tsipouras, T.P. Exarchos, D.I. Fotiadis, A. Kotsia, K.K. Naka, and L.K. Michalis, "A decision support system for the diagnosis of coronary artery disease, in *proc. 19th IEEE Computer Based Medical Systems*, Utah, USA, 2006, pp. 279-284.