A FRAMEWORK FOR FUZZY EXPERT SYSTEM CREATION

MARKOS G. TSIPOURAS, CONSTANTINOS A. VOGLIS,
ISAAK A. LAGARIS, DIMITRIOS I. FOTIADIS

*Department of Computer Science, University of Ioannina,
GR 45110 Ioannina, Greece*

In this work a framework for the development and tuning of a fuzzy expert system is proposed. Given an initial set of crisp rules, the methodology consists of three steps: (a) disjunctive normal form representation of the crisp rules, (b) creation of a fuzzy model and (c) tuning of the model using a global optimization method. The proposed methodology is evaluated in the arrhythmia classification problem, using only the RR interval signal. Expert cardiologists determined an initial set of rules, which is used for the creation of a fuzzy model. Four types of cardiac rhythms are classified: normal sinus rhythm, ventricular flutter/fibrillation, premature ventricular contractions and 2o heart block. The results indicate sensitivity (average for all categories) 94%, specificity 98% and positive predictive value 94%.

1. Introduction

Expert systems are a branch of artificial intelligence that makes extensive use of specialized knowledge, usually represented in a rule-based manner, to solve problems at the level of a human expert [1]. An expert system is created defining a crisp model (set of rules) and then tuning its parameters. Mathematical optimization techniques are a powerful tool for the tuning of the parameters of a model. Fuzzy logic can be used for the management of the uncertainty and fuzziness of the rules [2].

Several approaches have been proposed in the literature for the development of fuzzy models. Most of them combine a classification model with a known optimization technique: fuzzy rules with genetic algorithms [3], fuzzy rules with simulated annealing [4] and multicriteria decision analysis with genetic algorithms [5]. Most of the proposed methods use a “derivative-free” optimization method (genetic algorithms, simulated annealing), due to the fact that analytical formulas of the derivatives are not available.

This research is part 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.

In this work a three-step framework for the creation of a fuzzy expert system (FES) is proposed. The system is based on several rules, provided by experts. In the first step the crisp rules are represented in a disjunctive normal form (DNF). In step 2, a fuzzy model is created by transforming the crisp rules into fuzzy, ones, using the sigmoid function. The binary AND and OR operators are transformed into fuzzy operators, using algebraic product and probabilistic OR functions. The model is tuned (step 3), finding the optimal parameters for the sigmoid function, forming the fuzzy expert system. The proposed methodology is evaluated using a medical problem, the arrhythmic beat classification [6,7,8] of ECG recordings.

2. Materials and Methods

2.1. *DNF Representation*

Each crisp rule R_i, $i=1,...,M$, is comprised from a set of simple rules $r_{i,j}$, $j=1,...,m$. A simple crisp rule is defined as:

$$r_{i,j}\left(d^l,\theta_{i,j}\right)=g_c\left(f_{i,j}(d^l),\theta_{i,j}\right), \tag{1}$$

where $f_{i,j}(\cdot)$ is a function of the data d^l entering the model, $\theta_{i,j}$ is a threshold and $g_c(\cdot)$ is the crisp membership function, defined as:

$$g_c^{inc}\left(x,\theta\right)=\begin{cases}0 & x\le\theta\\ 1 & x>\theta\end{cases} \text{ (increasing) or } g_c^{dec}\left(x,\theta\right)=\begin{cases}1 & x\le\theta\\ 0 & x>\theta\end{cases} \text{ (decreasing).} \tag{2}$$

A crisp rule is defined as:

$$R_i\left(d^l,\theta_i\right)=$$
$$\begin{cases}\left(r_{i,1}\left(d^l,\theta_{i,1}\right) \; AND \; r_{i,2}\left(d^l,\theta_{i,2}\right) \; AND \; ... \; AND \; r_{i,a_1}\left(d^l,\theta_{i,a_1}\right)\right) \quad OR\\ \left(r_{i,a_1+1}\left(d^l,\theta_{i,a_1+1}\right) \; AND \; r_{i,a_1+2}\left(d^l,\theta_{i,a_1+2}\right) \; AND \; ... \; AND \; r_{i,a_2}\left(d^l,\theta_{i,a_2}\right)\right) \quad OR\\ \quad ...\\ \left(r_{i,a_l+1}\left(d^l,\theta_{i,a_l+1}\right) \; AND \; r_{i,a_l+2}\left(d^l,\theta_{i,a_l+2}\right) \; AND \; ... \; AND \; r_{i,\tau}\left(d^l,\theta_{i,\tau}\right)\right)\end{cases}, \tag{3}$$

where θ_i is a vector containing all thresholds used in the i^{th} rule $\left(\theta_i=\left\{\theta_{i,j}\right\},\ j=1,...,\tau\right)$. Each rule $R_i\left(d^l,\theta_i\right)$ is formed in a DNF representation, i.e. as a disjunction (sequence of OR) of conjunctions (sequence of AND). This form is chosen because every logical expression, consisting of a combination of AND, OR and NOT binary operators, can be transformed to a DNF logical expression, therefore, every set of rules, defined for a specific problem, can be written in this form.

The classification of cardiac arrhythmias is performed using only the RR interval duration signal, which is extracted from ECG recordings with QRS detection [9,10]. A three RR interval sliding window $\left[RR_1, RR_2, RR_3\right]$ is used to classify the middle RR interval (RR_2) into one of the four categories: (1) ventricular flutter/fibrillation (VF), (2) premature ventricular contraction (PVC), (3) normal sinus rhythm (N) and (4) 2^{o} heart block (BII).

The dataset is $D = \left\{d^l, c^l\right\}$ with $d^l = \left[RR_1, RR_2, RR_3\right]^l$, 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 $c^l \in \left\{0,1\right\}^4$, where, if d^l belongs to class i, then $c^l = e_i$. The MIT-BIH arrhythmia database [11] is used to create windows of three consecutive RR intervals and define their class c^l. All beats from all records from the MIT-BIH arrhythmia database are used for the creation of the dataset D, excluding only 2 beats at the start and 2 beats at the end of each record due to the window length, which is used in the method. Both rhythm and beat annotations from the database are used to specify the class of a window, following the scheme:

- If RR_2 is annotated as ventricular flutter/fibrillation, then $c^l = \left[1,0,0,0\right]$.
- Else if RR_2 is annotated as PVC then $c^l = \left[0,1,0,0\right]$.
- Else if RR_2 belongs to 2^{o} heart block episode, then $c^l = \left[0,0,0,1\right]$.
- Else $c^l = \left[0,0,1,0\right]$.

Three rules are used for the classification:

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

1. $\left(RR_1 < \theta_{1,1}\right)$ *AND* $\left(RR_2 < \theta_{1,1}\right)$ *AND* $\left(RR_3 < \theta_{1,1}\right)$
2. $RR_1 + RR_2 + RR_3 < \theta_{1,2}$

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

1. $\left(RR_1 / RR_2 > \theta_{2,1}\right)$ *AND* $\left(RR_3 / RR_2 > \theta_{2,2}\right)$

2. $\left(RR_3 / RR_1 > \theta_{2,3}\right)\ AND\ \left(RR_1 / RR_2 > \theta_{2,4}\right)$
3. $\left(\left|RR_1 - RR_2\right| < \theta_{2,5}\right) AND \left(RR_2 < \theta_{2,6}\right) AND \left(2RR_3 / \left(RR_1 + RR_2\right) < \theta_{2,7}\right)$
4. $\left(\left|RR_1 - RR_2\right| < \theta_{2,5}\right) AND \left(RR_2 < \theta_{2,6}\right) AND \left(2RR_3 / \left(RR_1 + RR_2\right) < \theta_{2,7}\right)$

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

1. $\left(RR_2 \in \left[\theta_{3,1}, \theta_{3,2}\right]\right) AND \left(\left|RR_1 - RR_2\right| < \theta_{3,3}\right)$
2. $\left(RR_2 \in \left[\theta_{3,1}, \theta_{3,2}\right]\right) AND \left(\left|RR_2 - RR_3\right| < \theta_{3,3}\right)$

The $R_1\left(d^l, \theta_1\right)$, $R_2\left(d^l, \theta_2\right)$ and $R_3\left(d^l, \theta_3\right)$ crisp rules, which comprise the crisp model, in DNF representation are:

$$R_1\left(d^l, \theta_1\right) = \begin{cases} \left(\left(RR_1 < \theta_{1,1}\right)\ AND\ \left(RR_2 < \theta_{1,1}\right)\ AND\ \left(RR_3 < \theta_{1,1}\right)\right)\ OR \\ \left(RR_1 + RR_2 + RR_3 < \theta_{1,2}\right) \end{cases}$$

$$R_2\left(d^l, \theta_2\right) = \begin{cases} \left(\left(RR_1 / RR_2 > \theta_{2,1}\right)\ AND\ \left(RR_3 / RR_2 > \theta_{2,2}\right)\right)\ OR \\ \left(\left(RR_3 / RR_1 > \theta_{2,3}\right)\ AND\ \left(RR_1 / RR_2 > \theta_{2,4}\right)\right)\ OR \\ \begin{pmatrix} \left(\left|RR_1 - RR_2\right| < \theta_{2,5}\right)\ AND\ \left(RR_2 < \theta_{2,6}\right) \\ AND\ \left(2RR_3 / \left(RR_1 + RR_2\right) > \theta_{2,7}\right) \end{pmatrix}\ OR \\ \begin{pmatrix} \left(\left|RR_2 - RR_3\right| < \theta_{2,5}\right)\ AND\ \left(RR_2 < \theta_{2,6}\right) \\ AND\ \left(2RR_1 / \left(RR_2 + RR_3\right) > \theta_{2,7}\right) \end{pmatrix} \end{cases}, \quad (4)$$

$$R_3\left(d^l, \theta_3\right) = \begin{cases} \left(\left(RR_2 > \theta_{3,1}\right)\ AND\ \left(RR_2 < \theta_{3,2}\right)\ AND\ \left(\left|RR_1 - RR_2\right| < \theta_{3,3}\right)\right)\ OR \\ \left(\left(RR_2 > \theta_{3,1}\right)\ AND\ \left(RR_2 < \theta_{3,2}\right)\ AND\ \left(\left|RR_3 - RR_2\right| < \theta_{3,3}\right)\right) \end{cases}$$

with $\theta_1 = \left\{\theta_{1,1}, \theta_{1,2}\right\}$, $\theta_2 = \left\{\theta_{2,j}\right\}, j = 1,...,7$ and $\theta_3 = \left\{\theta_{3,1}, \theta_{3,2}, \theta_{3,3}\right\}$.

2.2. *Fuzzy Model*

To transform the crisp rules into fuzzy, the sigmoid function is used as fuzzy membership function and the *algebraic product* and *probabilistic* OR functions, which are continuous equivalents of the binary AND and OR operators:

$$g_f\left(x, a, b\right) = \frac{1}{1 + e^{a(b-x)}} \text{ or } g_f\left(x, a, b\right) = \frac{1}{1 + e^{a(x-b)}} \text{ (sigmoid function)}, \quad (5)$$

$$G_{AND}\left(a_1,a_2,...,a_k\right)=\prod_{i=1}^{k} a_i \quad (\textit{algebraic product}), \tag{6}$$

$$\begin{matrix} G_{OR}\left(a_1,a_2\right)=a_1+a_2-a_1a_2 \\ G_{OR}\left(a_1,a_2,...,a_k\right)=G_{OR}\left(a_1,G_{OR}\left(a_2,...G_{OR}\left(a_{k-1},a_k\right)\right)\right) \end{matrix} \quad (\textit{probabilistic OR}). \tag{7}$$

The $R_1\left(d^l,\theta_1\right)$, $R_2\left(d^l,\theta_2\right)$ and $R_3\left(d^l,\theta_3\right)$ fuzzy rules, for the arrhythmia classification problem, are defined as:

$$R_1\left(d^l,\theta_1\right)=G_{OR}\left(\begin{matrix} G_{AND}\left(\begin{matrix} g_f\left(RR_1,\theta^a_{1,1},\theta^b_{1,1}\right), g_f\left(RR_2,\theta^a_{1,2},\theta^b_{1,2}\right), \\ g_f\left(RR_3,\theta^a_{1,3},\theta^b_{1,3}\right) \end{matrix}\right), \\ \left(g_f\left(RR_1+RR_2+RR_3,\theta^a_{1,4},\theta^b_{1,4}\right)\right) \end{matrix}\right)$$

$$R_2\left(d^l,\theta_2\right)=G_{OR}\left(\begin{matrix} G_{AND}\left(g_f^{inc}\left(RR_1/RR_2,\theta^a_{2,1},\theta^b_{2,1}\right), g_f^{inc}\left(RR_3/RR_2,\theta^a_{2,2},\theta^b_{2,2}\right)\right), \\ G_{AND}\left(g_f^{inc}\left(RR_3/RR_1,\theta^a_{2,3},\theta^b_{2,3}\right), g_f^{inc}\left(RR_1/RR_2,\theta^a_{2,4},\theta^b_{2,4}\right)\right), \\ G_{AND}\left(\begin{matrix} g_f^{dec}\left(\left|RR_1-RR_2\right|,\theta^a_{2,5},\theta^b_{2,5}\right), g_f^{dec}\left(RR_2,\theta^a_{2,6},\theta^b_{2,6}\right), \\ g_f^{inc}\left(2\theta RR_3\ \theta\left(RR_1+RR_2\right),\ {}^{a}_{2,7},\ {}^{b}_{2,7}\right) \end{matrix}\right), \\ G_{AND}\left(\begin{matrix} g_f^{dec}\left(\left|RR_2-RR_3\right|,\theta^a_{2,5},\theta^b_{2,5}\right), g_f^{dec}\left(RR_2,\theta^a_{2,6},\theta^b_{2,6}\right), \\ g_f^{inc}\left(2\theta RR_1\ \theta\left(RR_2+RR_3\right),\ {}^{a}_{2,7},\ {}^{b}_{2,7}\right) \end{matrix}\right) \end{matrix}\right) \tag{8}$$

$$R_2\left(d^l,\theta_2\right)=G_{OR}\left(\begin{matrix} G_{AND}\left(g_f^{inc}\left(RR_1/RR_2,\theta^a_{2,1},\theta^b_{2,1}\right), g_f^{inc}\left(RR_3/RR_2,\theta^a_{2,2},\theta^b_{2,2}\right)\right), \\ G_{AND}\left(g_f^{inc}\left(RR_3/RR_1,\theta^a_{2,3},\theta^b_{2,3}\right), g_f^{inc}\left(RR_1/RR_2,\theta^a_{2,4},\theta^b_{2,4}\right)\right), \\ G_{AND}\left(\begin{matrix} g_f^{dec}\left(\left|RR_1-RR_2\right|,\theta^a_{2,5},\theta^b_{2,5}\right), g_f^{dec}\left(RR_2,\theta^a_{2,6},\theta^b_{2,6}\right), \\ g_f^{inc}\left(2\theta RR_3\ \theta\left(RR_1+RR_2\right),\ {}^{a}_{2,7},\ {}^{b}_{2,7}\right) \end{matrix}\right), \\ G_{AND}\left(\begin{matrix} g_f^{dec}\left(\left|RR_2-RR_3\right|,\theta^a_{2,5},\theta^b_{2,5}\right), g_f^{dec}\left(RR_2,\theta^a_{2,6},\theta^b_{2,6}\right), \\ g_f^{inc}\left(2\theta RR_1\ \theta\left(RR_2+RR_3\right),\ {}^{a}_{2,7},\ {}^{b}_{2,7}\right) \end{matrix}\right) \end{matrix}\right)$$

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

The final decision for each $R_i\left(d^l,\theta_i\right)$ rule is made as follows: if the maximum value of the results of the three rules is less or equal to θ_4 then d^l is classified as normal sinus rhythm (category 3). If the maximum value of the results of the three rules is more than θ_4 then the d^l is classified in the category of the rule that has the maximum result.

$$M_f\left(d^l,\Theta\right)=\begin{cases}[0,0,0,1] & if\ \max\left(R_1\left(d^l,\theta_1\right),R_2\left(d^l,\theta_2\right),R_3\left(d^l,\theta_3\right)\right)=R_1\left(d^l,\theta_1\right)>\theta_4\\ [0,0,1,0] & if\ \max\left(R_1\left(d^l,\theta_1\right),R_2\left(d^l,\theta_2\right),R_3\left(d^l,\theta_3\right)\right)=R_2\left(d^l,\theta_2\right)>\theta_4\\ [0,1,0,0] & if\ \max\left(R_1\left(d^l,\theta_1\right),R_2\left(d^l,\theta_2\right),R_3\left(d^l,\theta_3\right)\right)\le\theta_4\\ [1,0,0,0] & if\ \max\left(R_1\left(d^l,\theta_1\right),R_2\left(d^l,\theta_2\right),R_3\left(d^l,\theta_3\right)\right)=R_3\left(d^l,\theta_3\right)>\theta_4\end{cases},\qquad(9)$$

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

2.3. ***Optimization***

Formulating the training process of a model as an optimization problem is a common practice in order to construct efficient classification expert systems. The fuzzy model $M_f\left(d^l,\Theta\right)$ is optimized using a training dataset (D_{train}), which is a randomly selected subset of D, containing 250 patterns from each class. Thus, the size of the training dataset is 1000. The mean square error (MSE) cost function is used:

$$F_{MSE}\left(D_{train},\Theta\right)=\frac{1}{1000}\sum_{l=1}^{1000}\left\|M_f\left(d^l,\Theta\right)-c^l\right\|.\qquad(10)$$

The optimization method used is the HTMLSL [12], a stochastic algorithm based on MLSL. The algorithm attempts to find all local minima of an objective function $F(x)$ inside a bounded set $S\subset\mathbb{R}^n$, that are potentially global. These local minima are obtained by a local-search procedure, starting from suitably chosen points in a properly maintained sample. Stochastic algorithms in the framework of multistart suffer from the problem of recovering the same local minima repeatedly, a fact that diminishes their efficiency. MLSL is constructed in such a way so as to avoid this undesirable repetition. At the k^{th} iteration:

1. Construct a sample selecting at random N points from S and evaluate the objective function for each point.
2. Choose from the sample a subset of points to be used as start points for local searches.

3. Perform a local search from each start point. If a new minimum is discovered store it.
4. Determine whether to stop or not. If not, repeat from step 1.

From the stored local minima the one with the lowest value might be the global minimum. The HTMLSL algorithm utilizes the information from the first derivatives on the local optimization step.

3. Results

The fuzzy expert system is evaluated with the test dataset (D_{test}) consisting of the remaining patterns of D after selecting D_{train} ($D_{test} = D - D_{train}$). 20 different pairs of D_{train} and D_{test} are created. The HTMLSL algorithm is set to perform a maximum of 10000 iterations. The optimization process has a total runtime less than 10 seconds on a Pentium 4 processor. The mean confusion matrix (the mean value of the 20 confusion matrices) is then computed and it is presented in Table 1. Sensitivity, specificity and positive predictive value, for the mean confusion matrix is also calculated and presented in Table 2.

Table 1. Confusion matrix for the arrhythmia classification problem (%).

	AF	PVC	N	BII
AF	89.87	1.08	0.43	0.00
PVC	0.36	81.54	3.79	0.24
N	0.50	17.08	95.57	0.78
BII	0.00	0.00	0.16	98.98

Table 2. Sensitivity, specificity and positive predictive value (%).

	AF	PVC	N	BII
Sensitivity	99.05	81.79	95.62	98.98
Specificity	99.50	98.49	93.68	99.94
Positive Predictive Value	98.35	94.89	83.88	99.84

4. Discussion and conclusions

The proposed methodology describes a three-step procedure for the creation of a fuzzy expert system: (a) DNF representation of the crisp rules, (b) creation of the fuzzy model and (c) tuning of the model using global optimization. The initial set of crisp rules could be provided by experts or automatically extracted using data mining. Given the initial set of rules, the proposed methodology can produce an expert system, for any given problem. This is a major advantage, comparing to other similar methodologies proposed in the literature, which, in most of the cases, deal with a certain problem and do not provide a general framework. Also, the formulation of the fuzzy model provides with analytical forms of the derivatives of the fuzzy model, which is a very helpful for the optimization step.

The proposed methodology is tested in order to produce an expert system for the arrhythmia classification problem with high efficiency. This expert system uses only the RR interval signal, which is easy to extract from ECG recordings, on the contrary to other time-consuming ECG processing techniques used.

References

1. J. Giartano, G. Riley, Expert systems, Principles and programming, *PWS Publishing Company*, 3rd edition (1998).
2. L.H. Tsoukalas, R.E. Uhrig, Fuzzy and neural approaches in engineering, *John Wiley & Sons*, (1997).
3. W. Pedrycz, J.V. de Oliveira, An algorithmic framework for development and optimization of fuzzy models, *Fuzzy sets and systems*, **80**, 37 (1996).
4. J.M. Garibalti, E.I. Ifeachor, Application of simulated annealing fuzzy model tuning to umbilical cord acid-base interpretation, *IEEE Trans Fuzzy Systems*, **7**, 72 (1999).
5. Y. Goletsis, C. Papaloucas, D.I. Fotiadis, A. Likas, L.K. Michalis, Automated ischemic beat classification using genetic algorithms and multicriteria decision analysis, *IEEE Trans Biom Eng*, **51**, 1717 (2004).
6. P. de Chazal, M. O’Dwyer, R.B. Reilly, Automatic classification of heartbeats using ECG morphology and heartbeat interval features, *IEEE Trans Biom Eng*, **51**, 1196-1206 (2004).
7. M.G. Tsipouras, D.I. Fotiadis, D. Sideris, Arrhythmia classification using the RR-interval duration signal, *CinC 2002*, 485 (2002).
8. M.G. Tsipouras, D.I. Fotiadis, D. Sideris, An arrhythmia classification system based on the RR interval signal, *Artif Intel Med*, **33** 237 (2005).
9. J. Pan, W.J. Thompkins, A Real-Time QRS Detection Algorithm, *IEEE Trans Biom Eng*, **32**, 230 (1985).
10. P.S. Hamilton and W.J. Tompkins, Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database, *IEEE Trans Biom Eng*, **33**, 1157 (1986).
11. MIT-BIH Arrhythmia Database CD-ROM, *Harvard-MIT Division of Health Sciences and Technology*, 3rd Edition (1997).
12. 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*, **159**, 63 (2004).