A Platform for Wide Scale Integration and Visual Representation of Medical Intelligence in Cardiology: The Decision Support Framework

TP Exarchos[1], MG Tsipouras[1], D Nanou[1], C Bazios[1],
Y Antoniou[1], DI Fotiadis[1,2,3]

[1]Unit of Medical Technology and Intelligent Information Systems, Department of Computer Science,
University of Ioannina, Ioannina, Greece
[2]Biomedical Research Institute, Ioannina, Greece
[3]Michaelideion Cardiology Center, Ioannina, Greece

Abstract

In this paper we present the Decision Support Framework (DSF) of the NOESIS platform. NOESIS addresses wide scale integration and visual representation of medical intelligence in cardiology and aims at the development of a web-based personalized system with enhanced intelligence that supports health professionals in taking the best possible decision for prevention, diagnosis, and treatment. The core of the NOESIS project is a set of Fuzzy Expert Systems (FES), one for each cardiovascular sub-domain, automatically generated from the DSF. An initial set of crisp rules, generated using data mining techniques, is employed to define a fuzzy model, using the sigmoid function and fuzzy equivalents of the binary operators. The parameters used to define the models are tuned using a global optimization algorithm.

1. Introduction

Medical knowledge is inherently complex and uncertain. Medical experts may provide different interpretations for symptoms since all of them depend also on a given context and most of them are established by statistical utilization. In addition, there is a large number of medical concepts and a large amount of associations between symptoms and diagnoses. Therefore, it is crucial for health professionals to be able to acquire useful information from a variety of resources and most importantly within a specific context. The NOESIS system objective is to reduce uncertainty, in the complex domain of cardiovascular diseases, as much as possible. The main part of the NOESIS system is the Decision Support System (DSS) which supports health professionals in taking the best possible decision for prevention, diagnosis, and treatment.

Several cardiological DSSs have been proposed in the literature, dealing mainly with the diagnosis of myocardial ischemia or arrhythmia. Myocardial ischemia diagnosis using the ECG signal can be described as a sequence of two tasks: ischemic beat detection and ischemic episode definition. The first is related to the classification of beats as normal or ischemic. Several techniques have been proposed for ischemic beat classification, which evaluate the ST segment changes and the T-wave alterations, using different methodological approaches. More specifically, they use parametric modelling, wavelet theory, set of rules, artificial neural networks [1], multicriteria decision analysis and genetic algorithms [2]. Arrhythmic beat classification is another field of interest, where each beat is classified into several different rhythm types. Proposed approaches in the literature employ artificial neural networks, fuzzy neural networks, "mixture of experts approach", hermite functions combined with self-organizing maps, time-frequency analysis combined with knowledge-based systems and rule-based systems, for beat classification combined with a knowledge-based determinist automaton, for arrhythmic episode detection and classification [3].

In the current study, an innovative framework is proposed for generating FESs. This framework can be used to generate a FES, regardless the domain of application. In order to use this methodology in a specific domain, an initial annotated dataset is required. This dataset is obtained from several resources, distributed over the web. The quality of the diagnosis, produced by the FES, is analogous to the quality of this dataset.

This framework is applied in the cardiovascular diseases domain, which cover a wide range of pathologies and affect almost every age. For reaching to the diagnosis, the following elements are required: i) the symptoms which are described by the patient, ii) the clinical findings, described by the physician after his physical examination and iii) the test results.

Based on the above, the DSF focuses in three different

functionalities: (a) Propose Diagnosis, which includes early diagnosis of coronary artery disease (CAD), myocardial ischemia diagnosis and arrhythmia diagnosis. (b) Propose Therapy which includes prediction of clinical restenosis in patients undergoing angioplasty with stenting and treatment in patients with recent valve prosthesis.

2.   Methods

The proposed framework consists of 5 stages (Fig. 1): (1) Continuous features discretization, (2) Association rule mining, (3) Association rule pruning, (4) Development of a fuzzy model, (5) Threshold optimization. After the application of these stages to an initial annotated dataset, obtained from distributed over the web databases, a FES is generated.
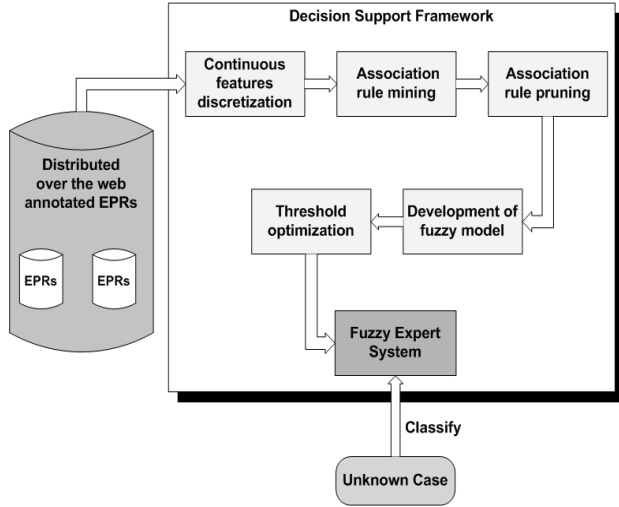

Figure 1: The architecture of the DSF

2.1.   Continuous feature discretization

The continuous valued features are transformed to categorical ones in order to be used for association rule mining (2nd Stage). This process is called discretization. The discretization approach we use is based on entropy minimization and the minimum description length principle [4]. This algorithm performs slightly better than others as reported in the literature. The algorithm uses the class entropy of candidate partitions to select a cut point for discretization. The method is then applied recursively to the two intervals of the previous split, thus creating multiple intervals for the feature. For a set of instances S, a feature A, and a cut point P, the class information entropy of the partition induced by P is given as:

$$E(A,P;S) = \frac{|S_1|}{|S|} Ent(S_1) + \frac{|S_2|}{|S|} Ent(S_2) \qquad (1)$$

where S_1 and S_2 are two intervals of S bounded by cut point P, and $Ent(S)$ is the class entropy of a subset S. For a feature A, the cut point P which minimizes Eq. (1) is selected as a binary discretization boundary. The procedure is repeated until the stopping criterion, which is based on the minimum description length principle, is reached.

2.2.   Association rule mining

One of the well known data mining techniques is association rule mining [5] and is perhaps the most common form of local-pattern discovery in unsupervised learning systems. Association rule mining has been extensively investigated in the data mining literature. It typically aims at discovering associations between items in a transactional database. Given a set of transactions (instances) $S=\{T_1,…,T_n\}$ and a set of items $I=\{i_1,…,i_g\}$ such that any transaction T in S is a set of items in I, an association rule is an implication $X \Rightarrow Y$ where the antecedent X and the consequent Y are subsets of a transaction T in S, and X and Y have no common items. The most widely used framework for association rule evaluation is the support-confidence framework. The *support* of an association rule $X \Rightarrow Y$ is the ratio of the transactions which contain the itemsets X and Y to the total number of transactions in the database. The confidence of an association rule $X \Rightarrow Y$ is the ratio of the transactions which contain the itemsets X and Y to the transactions which contain the itemset X.

In the proposed framework, we extract association rules that have as consequent only the class attribute. These rules are called Class Association Rules (CARs) and are extracted with the Apriori algorithm [5], using user specified minimum support and minimum confidence thresholds.

2.3.   Association rule pruning

Usually, the number of rules generated from association rule mining algorithms is huge. For this reason pruning is applied in order to cut down the number of generated rules. The proposed framework uses two pruning mechanisms: First using the pessimistic error based method. More specifically if rule r's pessimistic error rate is higher than the pessimistic error rate of rule r^- (the latter is obtained by deleting one condition from the conditions of r), then rule r is pruned. The second pruning technique is the one followed by the CBA algorithm [6]. This technique divides the rules into "*correct rules*" and "*wrong rules*" and keeps only the *correct rules* as the final set of rules.

2.4.   Development of fuzzy model

The generated rules are crisp and are transformed to fuzzy rules. Fuzzy logic has been proposed as a way to represent and manage the vagueness that arises in the data or the expert knowledge concerning a certain problem.

Membership functions are a fundamental aspect in set theory, measuring the certainty of an object x belonging to a set S. Sigmoid function, which can be used as fuzzy membership function, is defined as:

$$g_f(x,a,b) = \frac{1}{1+e^{a(b-x)}} .$$ (2)

A fuzzy model is comprised from R_i fuzzy rules ($i = 1,...,R$). Each rule R_i is formed as a combination of $r_{i,j}$ simple rules ($j = 1,...,r$), as follows:

$$R_i(d^l,\theta_i) = G_{OR}\left(\begin{array}{l} G_{AND}\left(\begin{array}{l} r_{i,1}(d^l,\theta_{i,1}^a,\theta_{i,1}^b),..., \\ r_{i,a_1}(d^l,\theta_{i,a_1}^a,\theta_{i,a_1}^b) \end{array}\right), \\ G_{AND}\left(\begin{array}{l} r_{i,a_1+1}(d^l,\theta_{i,a_1+1}^a,\theta_{i,a_1+1}^b),..., \\ r_{i,a_2}(d^l,\theta_{i,a_2}^a,\theta_{i,a_2}^b) \end{array}\right), \\ ... \\ G_{AND}\left(\begin{array}{l} r_{i,a_l+1}(d^l,\theta_{i,a_l+1}^a,\theta_{i,a_l+1}^b),..., \\ r_{i,m}(d^l,\theta_{i,r}^a,\theta_{i,r}^b) \end{array}\right) \end{array}\right),$$ (3)

where $\theta_i = \{\theta_{i,j}^a,\theta_{i,j}^b\}, j = 1,...,m$. Simple rules $r_{i,j}$ (the jth simple rule in the ith rule) are defined as:

$$r_{i,j}(d^l,\theta_{i,1}^a,\theta_{i,1}^b) = g_f(f_{i,j}(d^l),\theta_{i,j}^a,\theta_{i,j}^b),$$ (4)

where $f_{i,j}(\cdot)$ is a function of the data d^l and G_{AND} and G_{OR} functions are the minimum ($\vee$) and maximum ($\wedge$) operators, which are fuzzy equivalent for the binary AND and OR operators. Using this approach the G_{AND} and G_{OR} functions are defined as:

$$G_{AND}(a_1,a_2,...,a_k) = \bigwedge_{i=1}^{k} a_i \text{ and}$$ (5)

$$G_{OR}(a_1,a_2,...,a_k) = \bigvee_{i=1}^{k} a_i .$$ (6)

The final decision of the fuzzy model M_f is made using the results from all rules:

$$M_f(d^l,\Theta) = F_f(R_1,R_2,...,R_i),$$ (7)

where $\Theta = \{\theta_1,\theta_2,...,\theta_i\}, i = 1,...,M$ and $F_f(\cdot)$ is a function combining the outcomes of all $R_i(d^l,\theta_i)$ fuzzy rules. The data are represented as $D = \{d^l,c^l\}$, $l = 1,...,K$, where $d^l \in \mathbb{R}^n$ is a single pattern with n features, $c^l \in \{1,2,...,m\}$ is its class with m different classes, and K is the number of patterns.

2.5. Threshold 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 $M_f(d^l,\Theta)$ is optimized using a training dataset (D_{train}), which is a randomly selected subset of D, containing equal number of patterns from each class. Then a global optimization technique is used to minimize the mean square error (MSE) cost function:

$$F_{MSE}(\Theta) = \frac{1}{nD}\sum_{l=1}^{D}\sum_{j=1}^{n}\left(M_f(d^l,\Theta)-c_j\right)^2 .$$ (8)

After the five above stages are applied, a FES is created.

3. Results

The above presented framework has been evaluated for generating fuzzy expert systems for two medical applications: ischemic beat detection and arrhythmic beat classification. In the case of ischemic beat detection, 11 hours of two channel ECG recordings were extracted from the ESC ST-T Database [7]. The above recordings were preprocessed in order to remove noise like baseline wandering, A/C interference and EMG contamination. Then, five features were extracted from every cardiac beat (fig. 2): ST Segment Deviation, ST Segment Slope, ST Segment Area, T Wave amplitude and T Wave normal amplitude.
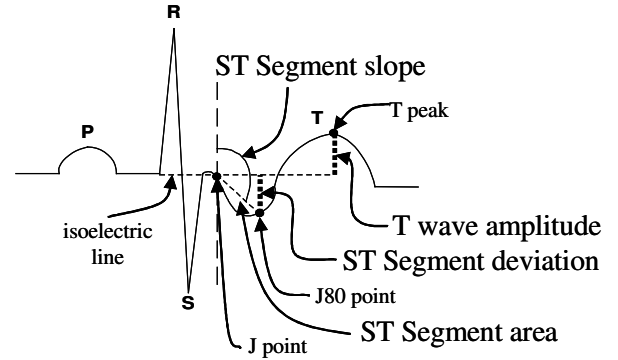

Figure 2: The features extracted from every cardiac beat

In addition, a sixth feature, the patient's age was also employed. The recordings resulted in 76,989 cardiac beats, from which 1,936 were used for the creation of the ischemia FES and 75,053 for testing it. The training set was constructed by selecting iteratively one out of a sequence of 40 beats. The results, in terms of sensitivity and specificity, from the application of the framework employing only the first 3 stages and then using all five stages, are presented in Table 1.

Table 1: Cardiac beat classification results using the DSF

	DSF (3 stages)	DSF (5 stages)
Sensitivity	86.37%	87.80%
Specificity	90.43%	92.46%

For the classification of cardiac arrhythmias, only the tachogram, extracted from ECG recordings with QRS detection, was used. A three RR interval sliding window $[RR_1, RR_2, RR_3]$ was 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° heart block (BII). The dataset is $D = \{d^l, c^l\}$ with: $d^l = [RR_1, RR_2, RR_3]^l$, the lth 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 the d^l belongs to class i, then $c^l = e_i$. All beats from all records from the MIT-BIH [8] arrhythmia database were used to create and evaluate the arrhythmia FES. The HTMLSL algorithm [9], which is a stochastic global optimization technique, is used for the optimization stage. The training dataset (D_{train}) is a randomly selected subset of D, containing equal number of patterns from each class. This number was set to 250, and thus the size of the training dataset is 1000. The test dataset (D_{test}) consists 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 mean values for sensitivity and specificity are presented in Table 2.

Table 2: Sensitivity and specificity results (%) for the arrhythmic FES.

	AF	PVC	N	BII
Sensitivity	99.05	81.79	95.62	98.98
Specificity	99.5	98.49	93.68	99.94

4. Discussion and conclusions

In the current study, we presented the Decision Support Framework of the NOESIS project. The DSF is employed in order to create fuzzy expert systems and includes five basic stages: discretization of the continuous valued features, rule mining, rule pruning, development of fuzzy model and optimization of the thresholds and generation of the fuzzy expert system. A major advantage of the proposed framework is its fully automated, since the whole procedure is data driven. In addition, the application of the DSF is web-based.

The proposed framework has been evaluated in the detection of ischemic cardiac beats in long duration ECG recordings. The data has been provided by the ESC ST-T Database and high scores are obtained (sensitivity 88%, specificity 92%). Also, it has been evaluated in the classification of cardiac beats in ECG recordings. The data has been provided by the MIT BIH arrhythmia database and high classification scores are obtained (average sensitivity 93.86%, average specificity 97.9%).

Further exploitation might focus on the application of the framework to other medical domains. This could be easily implemented, with the definition of other findings and appropriate diagnoses provided by the experts, or could be also implemented using data driven procedures, like feature selection methods.

The possibility of applying our framework to real clinical practice and evaluating its performance in real clinical conditions is of great interest.

Acknowledgements

References

[1] Papaloukas C, Fotiadis DI, Likas A, and Michalis LK. An ischemia detection method based on artificial neural networks. Artif Intell Med 2002;(24):167-178.

[2] Goletsis Y, Papaloukas C, Fotiadis DI, Likas A, and Michalis LK. Automatic Ischemic beat classification using genetic algorithms and multicriteria decision analysis. IEEE Trans Biom Eng 2004;(51):1717-1725.

[3] Tsipouras MG, Fotiadis DI, and Sideris D. An arrhythmia classification system based on the RR interval signal. Artif Intel Med (2005);(33):237-250.

[4] Fayyad UM and Irani KB. Multi-interval discretization of continuous-valued attributes for classification learning. in Proc of the 13th Int Joint Conf Artif Intel. 1993: 1022-1027.

[5] Agrawal R, and Srikant R. Fast algorithms for mining association rules. in Proc. 20th Int Conf Very Large Data Bases. 1994: 487-499.

[6] Liu B, Hsu W, and Ma Y. Integrating classification and association rule mining. in Proc. of the 4th Int Conf on Knowledge Discovery and Data Mining, 1998: 80-86.

[7] European Society of Cardiology, European ST-T database directory. Pisa: S.T.A.R., 1991.

[8] MIT-BIH Arrhythmia Database CD-ROM. Third Edition, 1997, Harvard MIT Division of Health Sciences and Technology

[9] Theos FV, Lagaris IE and Papageorgiou DG. PANMIN: Sequential and Parallel Global Optimization Procedures with a variety of options for the local search strategy. Comp Phys Commun (2004);159: 63-69.

Address for correspondence

Dimitrios I. Fotiadis
Unit of Medical Technology and Intelligent Information Systems, Dept. of Computer Science, University of Ioannina Campus, P.O. BOX 1186, GR 45110 Ioannina, Greece
fotiadis@cs.uoi.gr