Original Article

Fuzzy knowledge-intensive case based classification for the detection of abnormal cardiac beats

Abdeldjalil KHELASSI [1,2], Mohamed Amin Chick [2]

[1.] Department of Informatics, Faculty of Sciences, Tlemcen University, Tlemcen, Algeria
[2.] Biomedical Laboratory, Faculty of Technology, Tlemcen University, Tlemcen, Algeria

khelassi.a@gmail.com

Abstract:
This paper presents a new automated diagnostic system to classification of electrocardiogram (ECG) cardiac beats. We have developed an intensive-knowledge case based reasoning classifier which uses a distributed case base enriched by partial domain knowledge (rules). An original similarity measures is proposed by combining the sigmoid similarity function with the fuzzy sets to ameliorate the system accuracy in the detection of cardiac arrhythmias. The experiments presented in this work concern the detection of Premature Ventricular Contraction PVC, normal and abnormal cardiac beats from a pattern extracted from the Electronic medical records collected and published by Beth Israel Hospital (MIT-BIH). The achieved results demonstrate the efficiency and the performance of the developed system.

Bibliographic Information of this article:
[Abdeldjalil KHELASSI, Mohamed Amin Chick. **Fuzzy knowledge-intensive case based classification for the detection of abnormal cardiac beats**. Electronic Physician, 2012;4(2):565-571. Available at: http://www.ephysician.ir/2012/565-571.pdf]. (ISSN: 2008-5842). http://www.ephysician.ir

Keywords: Classification; Intensive-knowledge case based reasoning; Fuzzy sets; similarity measures; Cardiac arrhythmia diagnosis

© 2009-2012 Electronic Physician

1. Introduction

Up till now the most important source of information used by the cardiologists for the cardiac diseases diagnosis is the Electrocardiogram (ECG). The ECG is a signal produced by an electrocardiograph, which records the electrical activity of the heart. Through its wave's duration and axes values the cardiologists recognize the abnormality of the heart beat for detecting the cardiac arrhythmias.

The classification consists of associating an object with a predefined class. There are many methods and approaches from the artificial intelligence which prove a good performance to accomplish this task as artificial neural network ANN, fuzzy systems, and similarity based classification (SBC) and other paradigms (1-5). However, each application of these approaches has positives and weaknesses that are sometime accepted and some time not accepted according to the importance and the context of the application. The medical applications including aided diagnosis and decision support systems are a critical kind of applications where the precision and the transparency are very important because it touch the human health and life.

The CBR is an intelligent approach inspired from many discipline it draws a human reasoning model. It consists of using the prior expertise to resolve new problems. This expertise is stored as a set or collection of cases called cases base. Each case represents one problem associated with its solution. The intensive-knowledge case based reasoning is a variant of CBR in which the cases is enriched by partial domain knowledge. Also the distributed case based reasoning is a variant of CBR in which the reasoning is distributed through a set of agent and the cases through a set of case bases. These variants have been developed to ameliorate the accuracy and the performance of the systems.

In this work we have developed an original classification system (IK-CBRC) for achieving the medical applications needs and for developing a flexible and accurate model. The developed system apply the intensive-knowledge case based reasoning variant and the distributed reasoning approaches via a set of agents and a set of case bases

constructed from some electronic model records EMR from the MIT-BIH database which contains a set of electrocardiogram ECG commented by BIH doctors (6). We have also combined between the fuzzy sets and the similarity measures functions to increase the accuracy of our system. To evaluate our approach we have used other EMRs from the same database. The achieved results prove the usefulness of the developed approach and the integration of the fuzzy sets in the similarity measures.

2. Materials and Methods

2.1 The Electrocardiogram ECG

The ECG is a signal produced by an electrocardiograph, which records the electrical activity of the heart over time. Through its wave's duration and axes values they recognize the abnormal heart beat and its kind which indicates the cardiac arrhythmia (the disease). There are more than forty cardiac arrhythmias each one is characterised by some rules about the measures extracted from the ECG of the patient. The elementary unit of the Electrocardiogram ECG is the beat, as shown in Fig1, it contains six waves (P, Q, R, S, T, and U) can be also considered as four parts: the P wave, the complex QRS, the T wave and the U wave. More information about the ECG is given in (7).

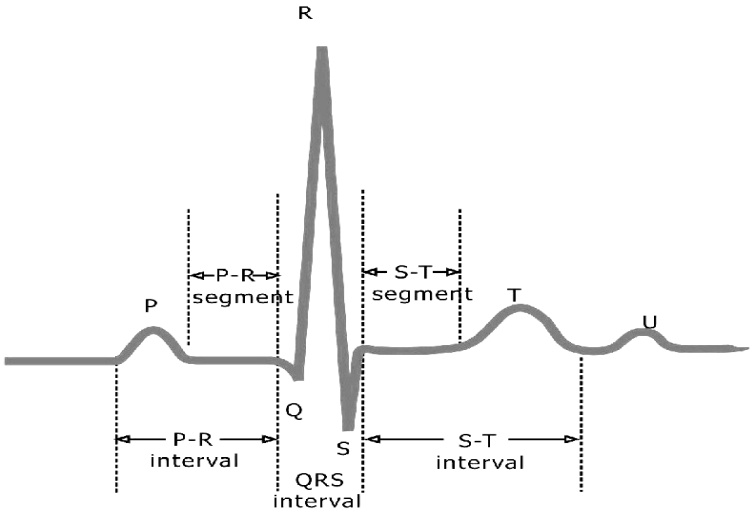

Figure1. The ECG parameters

2.2 Fuzzy sets

Fuzzy sets have been introduced by Lotfi A. Zadeh (1965) as an extension of the classical notion of sets. It consists to use a degree of membership instead of a simple membership in the classic sets. In classical set theory, the membership of elements in a set is assessed in binary terms according to a -bivalent condition- an element either belongs or does not belong to the set. By contrast, fuzzy set theory permits the gradual assessment of the membership of elements in a set; this is described with the aid of a membership function valued in the real unit interval [0, 1]. Fuzzy sets generalize classical sets, since the indicator functions of classical sets are special cases of the membership functions of fuzzy sets, if the latter only take values 0 or 1 (8).

2.3The case based reasoning CBR

Case-based reasoning is a problem solving paradigm that in many respects is fundamentally different from other major AI approaches. Instead of relying solely on general knowledge of a problem domain, or making associations along generalized relationships between problem descriptors and conclusions, CBR is able to utilize the *specific* knowledge of previously experienced, concrete problem situations (cases) (9). The case is a contextualized piece of knowledge representing an experience that teaches a lesson fundamental to achieving the goals of the reasoning system. It is composed from two parts problem part and solution part. The main idea of case based reasoning is that two similar problems have the same solutions. Aamodt and Plaza describe the CBR life cycle as a four process summarized below (Figure 2).

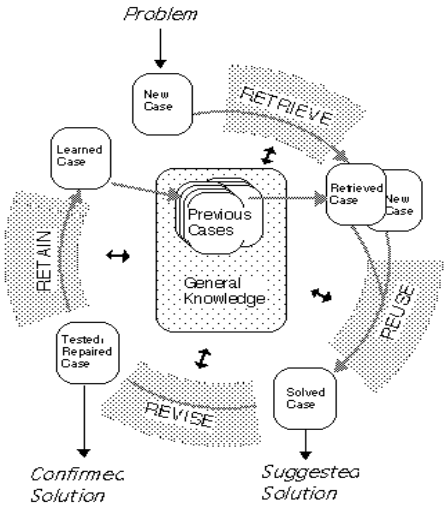

Figure2. Case-Based Reasoning cycle [A.Aamodt and E.Plaza 1994] (9)

The first process consists of retrieving from the cases base the similar case or cases which can be useful to solve the current problem. In the Second process, reuse, all solutions (cases) retrieved by the retrieve process are reused to find the potential solution. The Third one called revise process; which revises and checks the solution to fit the specifics of the current problem. Finally, the retain process, which updates the memory by adding the resolved problem as a new case to the cases base. The distributed case based reasoning consists of distributing the reasoning through a set of agents and the cases through a set of case bases. There are many works in the distribution of reasoning but each one has its proper realization. Research efforts in the area of distributed CBR concentrate on the distribution of resources with the intent of improving the performance of CBR systems. Although the phrase distributed CBR can be used in a number of different contexts (10).

A knowledge-intensive case-based reasoning method assumes that cases, in some way or another, are enriched with explicit general domain knowledge. The role of the general domain knowledge is to enable a CBR system to reason with semantic and pragmatic criteria, rather than purely syntactic ones. By making the general domain knowledge explicit, the case-based system is able to interpret a current situation in a more flexible and contextual manner than if this knowledge is compiled into predefined similarity metrics or feature relevance weights. A knowledge intensive CBR method calls for powerful knowledge acquisition and modelling techniques, as well as machine learning methods that take advantage of the general knowledge represented in the system (11, 12).

2.4 Contribution

The developed classification system called IK-CBRC contains two kinds of agents Adaptation agent and Similarity agent. Each case base contains cases from the same class. Each agent uses a predefined knowledge which contains ontologies, rules and heuristics to achieve their local goals and they collaborate for achieving the global goal. Also they interact with the users with a General User Interfaces for introducing the data and the classification parameters (Data, the similarity measures function and the adaptation rules) and for applying the machine learning algorithms. The cases are enriched by a partial domain knowledge which represents the cardiologists experience mobilised by some XML rules.

The developed system combines between many intelligent approaches. First of all, the adaptation agent infer from the domain knowledge base which contains some rules defined from the doctors experiences. If the response is unknown the adaptation edit a query associated with an ontology describing its features, this task is ensured by an interactive user interface.

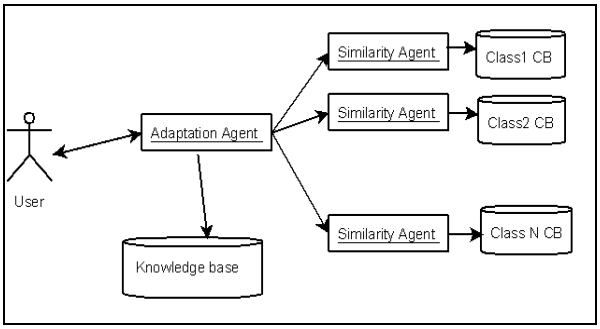

Figure 3. The IK-CBRC architecture

Following this, the Adaptation agent propagates the query to the similarity agents. Each similarity agent contains an ontology which describes its associated case base, via this ontology and the similarity knowledge, each similarity agent compute a rate of membership of the query in the associated class. Finally, the adaptation agent generates the solution by inferring form the adaptation knowledge applied in the results sent by the similarity agents. The rate of membership is computed by the following formula:

$$R = \sum_{i=1}^{n} sim(Q,C) \qquad (1)$$

Where sim (Q,C_i) the global similarity between the case C_i and the query Q, N the number of cases in the case base. The global similarity is measured with deferent functions, the traditional one as the sigmoid, and the exponential is defined by Axel (5). For example the sigmoid Similarity function is defined as:
Lets Q the query and C the case, qi ci: the attribute number i respectively of the query and the case and D denotes the space of case characterization models.

sim: DxD →[0,1]

$$sim(Q,C) = \sum_{i=1}^{n} w_i \frac{1}{1+e^{\frac{\delta(q_i,c_i)-\theta}{\alpha}}} \qquad (2)$$

Where n the number of attributes, w_i the weight of the attribute A_i defined by a machine learning algorithm (gradient descent). The parameters α and θ are defined intuitively after some experiments, and The logarithmic distance function is defined as:

σ:DxD→IR

$$\sigma(qi,ci) = \begin{cases} -\ln(c)-\ln(q) \text{ for } q,c > 0 \\ -\ln(-c)-\ln(q) \text{ for } q,c < 0 \\ \text{Undefined} \quad \text{else} \end{cases} \qquad (3)$$

We have proposed also a new similarity metric by using the fuzzy sets and the traditional global similarity function for generating three responses 1)similar 2)unknown and 3)not similar associated with the rate of membership of each response. This original proposition increases the accuracy of the system and its transparency. The rate of membership of the similarity agents' responses is computed with the following triangular membership functions:

$$\mu_s(x) = \begin{cases} 0 \; if \; x \leq a \\ \frac{x-a}{1-a} \; if \; x > a \end{cases} \qquad (4)$$

$$\mu_n(x) = \begin{cases} 0 \; if \; x \geq b \\ \frac{b-x}{b} \; if \; x < b \end{cases} \qquad (5)$$

$$\mu_u(x) = \begin{cases} 0 \text{ if } x \leq b \text{ ou } x > a \\ \frac{x-b}{0.5-b} \text{ if } x > b \text{ and } x \leq 0.5 \\ \frac{a-x}{a-0.5} \text{ if } x < a \text{ and } x > 0.5 \end{cases} \quad (6)$$

Where x is the global similarity measures.

The support of the fuzzy sets (a, b) is defined intuitively by using the agents interfaces or by using a machine learning algorithm. Also the function computing the value of x is selected by the user. The rate of membership is computed by putting Us (sim(Q,Ci)) in the place of sim (Q,Ci).

3. Results

In order to estimate the usefulness of our approach, we have performed several experiments with the developed IK-CBRC software. Also for proving the impact of the proposed approach we have realized several experiments with the same data and deferent strategies. The used data are divided in two parts one for the similarity learning (450 cardiac beats) and one for the test (400 cardiac beats). In these empirical experiments, we have applied the developed classifier for the recognition of cardiac arrhythmias via the cardiac beat measures described in table1 extracted from the Electronic Medical Records EMR recorded and collected by the laboratory of BIH (Beth Israel Hospital) in Boston in the United States, which is known as the MIT-BIH data base (13).

These EMRs contains the ECG signals of some patients recorded at a frequency of 360 Hz. Two or more cardiologists have made the diagnosis for these various records and they have annotated each cardiac cycle. The extracted dataset contains some normal beats and some cardiac arrhythmias as well as premature ventricular contraction PVC. The data is obtained and calculated using an algorithm developed and implemented in the LISI laboratory at the University of Rennes 1. This algorithm is based on the technique introduced by Pan J. and Tompkins W.J (14).

Table 1: The features of the extracted pattern from the cardiac beat ECG

Attribute	Type	Description
Pdur	REAL	The duration of the wave P.
PRseg	REAL	The PR segment.
QRS	REAL	The QRS larger.
STseg	REAL	The ST segment.
QTInterval	REAL	The QT Interval.
R_priorR	REAL	Distance between the current R and the prior one.
R_nextR	REAL	Distance between the current R and the next one.
RDI	REAL	Distance between R and R'
AmpR_S	REAL	Distance between R and S.
Beat_duration	REAL	The Beat duration.

The following tables (2, 3 and 4) contain the recognized cardiac beats of each class in the appropriate experiments. In each experiment we have used 400 classified queries (cardiac beats) from many classes including the normal the PVC classes and others. The tested queries are taken randomly.

Table2. Experiment1: In this experiment we have used just partial domain knowledge (some rules used by cardiologists*).

	Normal	PVC	Unknown
Normal	0	0	100
PVC	0	98	2
Other	0	0	200

*Rate of correct classification 74.5, Error rate 25.5%

Table3. Experiment2: In this experiment we have used the traditional similarity measures enriched by the same rules used in experiment1*.

	Normal	PVC	Unknown
Normal	100	0	0
PVC	0	98	2
Other	3	86	111

*Classification rate 77.25%, Error rate 22.75%

Table4. Experiment3: in this experiment we have used the proposed fuzzy similarity measures enriched by the same rules used in the above experiments*.

	Normal	PVC	Unknown
Normal	100	0	0
PVC	0	100	0
Other	0	0	200

*Classification rate 100%, Error rate 0%

4. Discussions

By applying the fuzzy similarity measures enriched by the partial domain knowledge we have obtained the best rate of correct classification (Experiment3) in which the classifier recognizes all queries. The first experiment demonstrates that the traditional experts systems have an important weakness (The rate of correct classification is just 74%) this big uncertainty is caused by: 1) Using just partial knowledge 2) The measures of ECG parameters was approximate. The second experiment prove the impact of the knowledge intensive case based reasoning comparing with the traditional experts systems approach but the traditional similarity metrics is less accurate than the proposed metrics in the last experiments. There are some works which combine between the fuzzy approach and the Case based reasoning as in which they incorporate the traditional case base paradigm by the Fuzzy Logic concepts in a flexible, extensible component-based architecture (15). Also which enforce the case based reasoning by a fuzzy logic system (16). Other researchers also introduce a fuzzy model for the representation of a CBR system (17). In our approach we combine the fuzzy sets with the traditional similarity measures function for generating an understandable response (Similar, not similar and unknown) which increases the system precision and the transparency.

There are also other works in the cardiac arrhythmias diagnosis in which they apply different approaches and intelligent techniques for the classification and automatic recognizing (1, 18). Others have used the Fuzzy approach, another researchers have used Support Vector Machines , or some hybrid models as (2, 3, 4, 18). The proposed approach as (4) recognizes all test data which prove that this original proposition ensure an accurate classification. The classifier can generate the unknown response which indicates the abnormality of the cardiac beat this criteria is very important which is original comparing with the cited approaches. The transparency of the response is ensured by recording the trace of the decision during the reasoning process. The use of separate and specialized cognitive agents ensures the flexibility of the classifier where we can integrate another agent for other cardiac arrhythmia.

5. Conclusion

In this work we have merged two CBR variants: The distributed CBR and the knowledge intensive CBR. The integration of fuzzy sets in the similarity measures helped to increase the accuracy of our system, improve the performance and decrease the learning complexity. Many original contributions was integrated in this work as: 1)The unknown response for inferring the abnormal cardiac beats 2) The combination of many optimization algorithms in the learning and the cases retrieving processes 3)The personalization of the reasoning criteria in deferent level of the classification which support the specialization of the system for deferent cases. 4) The flexibility and the scalability of the classifier are improved in the model by the multi-agent system approach.

Acknowledgements:

In this work we would like to thank so much every person who shares with us some useful information or ensuring for us his guide. We would like also to give special thanks for Pr Thomas Roth-Bergopher, Pr D. Aha and Pr Klaus-Dieter Althoff for their advises and help in ICCBR'10 Doctoral consortium.

Corresponding Author:
Abdeldjalil Khelassi,
Faculty of Sciences, Tlemcen University,
Tlemcen, Algeria
Tel: 021343286149
Fax: 021343286308.
E-mail: khelassi.a@gmail.com

References

1. W Jiang, SG Kong - Block-Based Neural Networks for Personalized ECG Signal Classification-Neural Networks, IEEE Trans on neural networks. 2007;18(6):1750-61.
2. Frediric, M., and Soowhan, H. Classification of cardiac Arrythmias using Fuzzy ARTMAP, IEEE Trans. Biomed. Eng. 1996; 43:4.
3. Lu H.L, and al. An Automated ECG Classification System Based on Neuro-Fuzzy System, Computers in cardiology IEEE Trans. Biomed. Eng. 2000; 48:1265- 71.
4. Yüksel Özbay. Integration of type-2 fuzzy clustering and wavelet transform in a neural network based ECG classifier. Expert Systems with Applications. 2011; 38(1):1004-10
5. Axel E. Bernal, Karen Hospevian, Tayfun Karadeniz, Jean-Louis Lassez. Similarity Based Classification. Advances in Intelligent Data Analysis. Lecture Notes in Computer Science.2003; 2810:187-97
6. MIT-BIH arrhythmia database. Available online at: http://www.physionet.org/physiobank/database/mitdb
7. J. Bensaid. Électrocardiogramme normal de l'adulte, EMC - Cardiologie-Angéiologie. 2004; 1(1):2-10.
8. Kareem S. Aggour, Marc Pavese, Piero P. Bonissone, William E. Cheetham. SOFT-CBR: A Self-Optimizing Fuzzy Tool for Case-Based Reasoning. Case-Based Reasoning Research and Development. Lecture Notes in Computer Science. 2003; 2689:5-19
9. Aamodt, E. Plaza. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. AI Communications. IOS Press. 1994; 7(1):39-59.
10. Plaza E, Gintynri M. Distributed case-based reasoning. The Knowledge engineering review. 2006;20(3):261-5.
11. Aamodt A. Knowledge-Intensive Case-Based Reasoning in CREEK. Advances in Case-Based Reasoning. Lecture Notes in Computer Science. 2004;3155:1-15.
12. Baccigalupo C, Plaza E. A Case-Based Song Scheduler for Group Customised Radio. Proceedings of the ICCBR '07 Conference. Available online at: http://www.iiia.csic.es/People/enric/papers/Baccigalupo-Plaza-2007-ICCBR.pdf
13. MIT-BIH arrhythmia database. Available online at: http://www.physionet.org/physiobank/database/mitdb
14. Pan J. Tompkins W.J. A real time QRS detection algorithme IEEE Trans. Biomed. Eng. 1985;23(4):230-6
15. Kareem S. Aggour, and al ''SOFT-CBR: a self-optimizing fuzzy tool for case-based reasoning'' 5th ICCBR. 2003:5-19
16. Meijun Yang, Qiang Shen. Reinforcing fuzzy rule-based diagnosis of turbomachines with case-based reasoning. International Journal of Knowledge-Based and Intelligent Engineering Systems. 2008;12(2):173-81
17. Voskoglou, M.G. Fuzzy Sets in Case-Based Reasoning. Fuzzy Systems and Knowledge Discovery, 2009. FSKD '09. Sixth International Conference. Tianjin. 2009;6: 252–6
18. Moavenian M, Khorrami H. A qualitative comparison of Artificial Neural Networks and Support Vector Machines in ECG arrhythmias classification. Expert Systems with Applications. 2010;37(4):3088-93