Adeptness Associative Learning Method for Real-Time Cardiac Arrhythmia Detection

Mohamed Ezzeldin A. Bashir[1], Dong Gyu Lee[2], Ibrahim Musa Ishag[2], Makki Okasha[2], Ho Sun Shon[2], Keun Ho Ryu[2]

[1]*University of Medical Science and Technology, Sudan*
{mohamed.izz}@umst-edu.net
[2]*Database/Bioinformatics Laboratory, Department of Computer Science, Cheungbuk National University, South Korea*
{dglee, ibrahim, makki,shon0621, khryu}@dblab.chungbuk.ac.kr

Abstract

It is vital for the automated system to accurately detect and classify ECG signals very fast to provide a useful means for tracing the heart's health in the real time. Making a random training set can lead or cause negative results. Simply, training set must designed carefully to consider all possible classes of the overall arrhythmia, so as to train the algorithm with the right group. Not only, considering all possible arrhythmia, but also with the same ratio, which means giving the algorithm a richness group to be trained in the right way with effective training set. Therefore, a means is needed for determining which record in the main file to be selected to replace the removed one from the active training data set. Our proposed methodology automatically trains the classifier model, using efficient set. Accordingly the experimental works show an improvement in the performance of some classification models to detect cardiac arrhythmia.

Keywords: Electrocardiogram (ECG); Arrhythmia; Classification; and Training dataset

1. Introduction

ECG signals are a very important medical instrument that can be utilized by clinicians to extract very useful information about the functional status of the heart. So to detect heart arrhythmia, which is the anomalous heartbeat map with a different shape in the ECG signal notice by deflection on the P, QRS, and T waves, some parameters are acquired and an enormous finding is produced [1].

The ECG gives two major kinds of information. First, by measuring time intervals on the ECG, the duration of the electrical wave crossing the heart can be determined, and consequently we can determine whether the electrical activity is normal or slow, fast or irregular. Second, by measuring the amount of electrical activity passing through the heart muscle, a pediatric cardiologist may be able to find out if parts of the heart are too large, or overworked. Electrocardiography has evolved over time and is becoming more accurate as it is being automated by making use of several software techniques available.

There has been a great deal of interest in systems that provide real time ECG classification through an intermediary local computer between the sensor and control center [2]. It is vital for the automated system to accurately detect and classify ECG signals very fast to provide a useful means for tracing the heart's health in the right time. The effectiveness of such systems is affected by several factors, including the ECG signals, estimated ECG's features and descriptors, the dataset used for learning purpose, and the classification model applied [3].

This paper, concerned with the challenges for training the classifiers model with updated data to facilitate the process of developing real time cardiac health monitoring systems. It presents a method that propos solution to solve this problem. In the rest of this paper, we will provide a brief description of related work, associative learning method, and then we present the experimental works and finally the conclusion.

2. Arrhythmia Detection Training Methods

Confirming local and global dataset are the main two approaches of training dataset used to learn the classifier model. Global is built from a large database that most automatic ECG analysis research works refer to this technique such as [3], [4]. Simply, there are training and testing datasets with different percentages through which the classifier is trained using the training dataset, and later predict the unseen group of data through the testing dataset. (Rodriguez, J. et) attempted to derive approach that can build the most accurate model for classifying cardiac arrhythmia based on feature extraction [5]. He divided the dataset into random groups one for training (66%) and another for validation (33%). He used the “weka” and “answertree” tool in his experiment. Sixteen methods were used in the experiments. One main challenge faced by this technique is the morphologies of the ECG waveforms that widely vary from patient to patient. Accordingly, the classifier learned by specific data related to an identical patient will perform very well when tested with unseen data of that patient often fail when presented with ECG waveforms for other patients. To overcome this problem, the literature shows that there is a trend to learn the classifier via training dataset as much as possible. This was the commercial trend introduced by the ECG device vendors. However, such an approach criticized in different aspects. First, when using a huge amount of ECG records to build a classifier, development, maintaining, and updating will become very complicated. Second, it is difficult to learn the classifier by abnormality ECG during the monitoring process. Therefore, there is the possibility to be unable to detect specific arrhythmia when applying that model to patient records. Moreover, it is impossible to introduce all ECG waveforms from all expected patients [6].

In previous works, we suggested a nested ensemble technique to solve the problem of training dataset by manipulating the training dataset for learning the classifier through up-to-date data, and manipulating the ECG features to select the proper adequate set (morphological features) to enrich accuracy [7]. Although the results are favorable, synchronizing the two components is expensive, which negatively affects the detection of the arrhythmia in real time. Moreover, it is quite static to some extent. Then Trigger Learning Method [8] and Active learning method [9] have been suggested to detect cardiac arrhythmia on line in very sufficient manner; simply introduced to learn the classifier model by up-to-date training data.

The local learning set is a customized to a specific patient; in other words, it is a technique focused on developing a private learning dataset corresponding to each patient [10]. Its intention is to familiarize the classification model with the unique characteristics of each patient. Although this technique looks to alleviate the problem of the learning process, it suffers from a clear problem related to the difficulties to distribute an ECG database because it is time consuming and labor intensive. Moreover, few patients are accepted to be involved in the development of the ECG processing method. Thus, there are limitations to the advantages provided by such technique among the expected audience, even if it is permissible.

3. Associative Learning Method

The associative learning method comes to solve the problem of feeding the arrhythmia detection algorithm with updated training sets. The associative technique has four steps as shown in Figure 1. First, an initial learning stage is introduced to learn the classifier by a random set of data without any further consideration. The classifier performance is evaluated (check) and updated (improve) for consistency, and applied the re-movement stage to avoid a combat situation. In the initial learning step, the learning process starts by utilizing a random group of records (categories), which represent (50%) of the overall dataset. In the check stage, the overall trusted mark, which is calculated using the local trusted mark that can be measured using a label assigned to the specific category with a specific vector of features.

$$L^M(x) = \sum_{f \in features} \beta_f(F,i).C^S(x) \tag{1}$$

where f is the feature number, F is the contribution of the feature, and CS(x) represents the category score when labeled as arrhythmia (i), which calculated as follows:

$$C^S(x) = \sum_{f \in features} \beta_f(F,i) \tag{2}$$

The function βf (F, i) checks the set of features (F) in specific category labeled as arrhythmia (i). It returns "+1” if the label (i) is assigned to category (x), otherwise it returns “−1.”

The local trust index LM(x) is considered in determining the overall trust index TrustM(X), which is defined using a sigmoid function Sigmoid(X) (0.5 < TrustM(X) < 1).

The checking step ends by two judgments; either the current training set is reliable or not, depending on different classes of arrhythmia assigned to categories. Accordingly, the unreliable set needs to be modified by a new group of data. This process has two parts, first, specifying the useless category or categories; and second, replacing it or them with newly selected one(s). In the first step, category (x) in the active training set (X) is removed if the category score CS(x) is less than a threshold δremove. Second, for the record selection the method introduce in-between set of data file to avoid delay, which is called cache the cache is used to substitute the partial or complete modification of the current active training data set. The objective is to minimize hitting the main database as much as possible, so as to save time and increasing the accuracy by double filtering. Starting from the second modification stage, the substitutions take place from the cache not from the main database. The improvement step is active when there is a limited number of bad labeling using the current group, while it is useless when there are multiple defects among the categories the thing that requires an inherited improvement process consequence. It is very expensive in terms of time, which negatively affects the performance of the classifier model to label different types of arrhythmia. Therefore, the re-movement step is introduced to deal with this problem.

The selection of the substituted category depends on record reference by permitting each record belongs to specific arrhythmia class to be selected when replacement is required. In this case, the cache control point interprets a recode reference simply as a tag and a feature set with arrhythmia class. The tag field uniquely identifies a recorded to determine whether it is in the cache or in the main file, the cache control point must simultaneously examine every line's tag for a match.

The removed category or categories will be sent to the cache not to the main database. Thus the removed categories will be sent to the cache with their record references. The cache size is fixed so as not to exceed 20% of the total dataset size.

Category will be removed from the cache if selected twice and removed from the active learning set. In this case, it will be replaced with a new category from the main database to avoid selecting the same one.

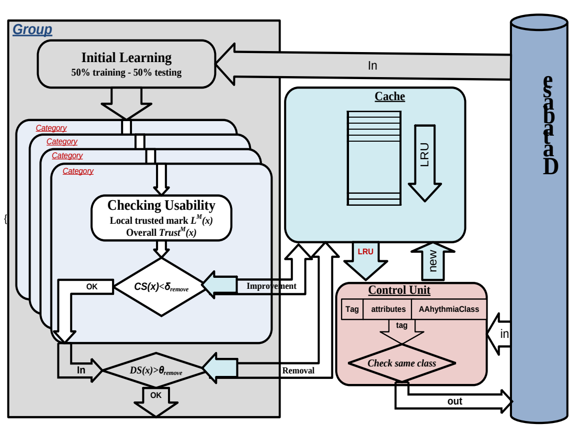

Figure 1. Associative learning process flow

4. Experimental Results

4.1. Environment

We used a database generated at the University of California, Irvine [11]. It was obtained from Waikato Environment for Knowledge Analysis (WEKA), containing 279 attributes and 452 instances [12]. The classes from 01 to 15 were distributed to describe normal rhythm, Ischemic changes (Coronary Artery Disease), Old Anterior Myocardial Infarction, Old Inferior Myocardial Infarction, Sinus tachycardia, Sinus bradycardia, Ventricular Premature Contraction (PVC), Supraventricular Premature Contraction, Left bundle branch block, Right bundle branch block, degree Atrio ventricular block, degree AV block, degree AV block, Left ventricle hypertrophy, Atrial Fibrillation or Flutter, and Others types of arrhythmia Respectively. Some instances related to specific arrhythmia classes are duplicated generating overall 573 instances. The experiments were conducted in WEKA 3.6.1 environment. Our experiment was carried out by a PC with an Intel Core processor (T M) 2 DUO, speed to 2.40 GHz. And RAM 2.00 GB.

4.2. Necessity of Including All ECG Parameters

First, we proved the necessity of including the P and T waves in conjunction with the QRS complex to evaluate arrhythmia the right way. We measure the performance of five different algorithms including OneR, J48, Naïve Bayes, Dagging, and Bagging according to the parameter(s) used to classify the arrhythmia. Table I summarizes the results obtained by each algorithm.

Table 1. The accuracy according to specific ECG parameter

Features	OneR	J48	Naïve Bayes	Dagging	Bagging
QRS only	60.4	91.2	76.5	63.5	81.0
QRS + P	60.4	91.4	77	62.4	81.6
QRS + T	61.3	91.2	76.7	63.0	82.3
QRS + P +T	61.1	92.3	77.7	64.2	83.0

4.3. Arrhythmia Detection

Figure 2 compares the accuracies achieved by the OneR, J48, naïve Bayes, dagging, and bagging methods when using the assciative learning. We also show their original performance without the proposed method for comparison.

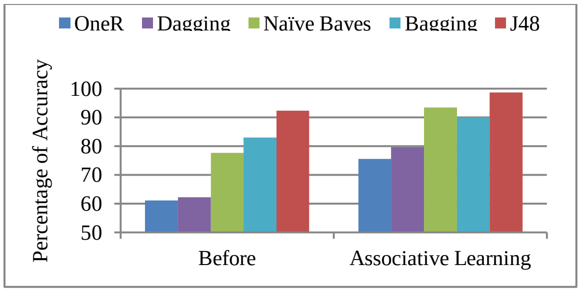

Figure 2. Accuracy achieved by different methods when using associative technique

Figure 3 illustrates the improvements due to the proposed associative learning in all algorithms tested here. We specifically compare the best-case accuracies when including all features related to the P, QRS, and T waves with that obtained after using the associative learning .

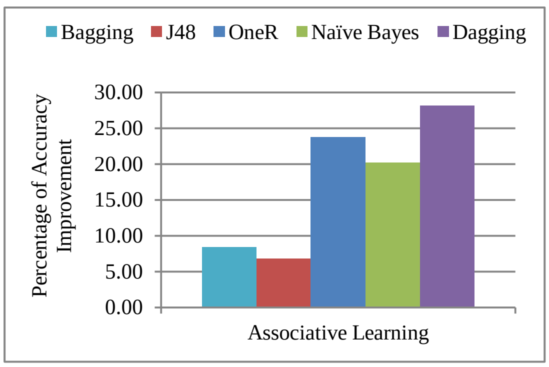

Figure 3. Accuracy improvement achieved by associative technique

Figure 3 clearly show that the associative learning improve the detection accuracy for the different types of arrhythmia. The improvement is noticeable for all algorithms with different weights due to their mechanisms. Specifically, improvements of 23.7, 6.8, 20.2, 28.1, and 8.4 percentages were achieved in performance for OneR, J48, naïve Bayes, dagging, and bagging, respectively, when applying the associative technique. In general, these are significant improvements.

It is also interesting to compare the accuracy of associative technique using the J48 algorithm with that of other methods presented in the literature. Methods from two representative studies were chosen for this comparison, which including trigger learning method [8] and active learning method [9]. Table II summarizes the comparative results of these methods, in which the last row lists the results of associative method. Among the two methods, the proposed method outperforms the other methods with an impressive accuracy of 98.6% in discriminating 15 ECG beat types.

Table 2. Accuracy comparison with other methods

Method	Accuracy %
Trigger learning method	96.1
Active learning method	97.6
Associative learning method	98.1

5. Conclusion

Cardiac health monitoring is a challenging problem in the field of data mining and knowledge extraction, and has received considerable attention over the past few years because of its importance in saving lives and reducing health risks. Today, cardiac health monitoring has reached a level of maturity when operating directly on or off-line. However, current methods are far from adequate for automated, remote cardiac health monitoring by detecting arrhythmia in real time. This is partly because of inter- and intra-patient variabilities. Thus, developing one classifier model to satisfy all patients in different situations using static training datasets is not practical. Furthermore, analyzing the QRS, P-wave, and other elements of ECG, and measuring the time interval between these elements, is necessary for real-time cardiac monitoring. This is technically infeasible with current systems because of computational limitations.

In this paper, we presented a associative technique as a proposed solution to solve these problems. The performance of our method was evaluated using various approaches, which demonstrate their effectiveness. In future, we plan to perform more experiments to account for interrelated ECG features.

Acknowledgement

This work was supported by the grant of the Korean Ministry of Education, Science and Technology (The Regional Core Research Program / Chungbuk BIT Research-Oriented University Consortium), and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF No. 2010-0001732).

6. References

[1] Dale Dubin, MD, “rapid interpretation of EKG”, *6th edition, cover publishing co.*, 2000

[2] U. Rajendra, M. Sankaranarayanan, J. Nayak, C. Xiang, and T. Tamura, “Automatic Identification of cardiac health using modeling techniques: a comparative study”, *Inf. Science. 178*, 2008, pp. 457–4582

[3] M. Ezzeldin A. Bashir , M. Akasha, D. G. Lee, Min Yi, K. H. Ryu , E. J. Bae, M. Cho, and C. Yoo, “Highlighting the Current Issues with Pride Suggestions for Improving the Performance of Real Time Cardiac Health Monitoring,” *DEXA Bilbao, Spain, LNCS 2010.J*, 2010.

[4] G. Bortolan, I. Jekova and I. Christov, “ Comparison of four methods for premature ventricular contractions and normal beats clustering”, *Comp. Card.* , 2005, pp. 921–924

[5] G. Clifford, F. Azuaje, P. McSharrg, “Advanced methods and tools for ECG data analysis”, *Artech house*, 2006.

[6] H. Palreddy and W. Tompkins, “A patient-adaptable ECG beat classifier using a mixture of experts approach”, *B. med. Eng. 44*, 199, pp. 891–900

[7] M. Ezzeldin A. Bashir , M. Akasha, D. G. Lee, Min Yi, K. H. Ryu , E. J. Bae, M. Cho, and C. Yoo, “Nested Ensemble Technique for Excellence Real Time Cardiac Health Monitoring”, *BioComp lasvegas USA 2010,* 2010

[8] M. Ezzeldin A. Bashir, D. G. Lee, M. Li, J. W. Bae, H. S. Shon, M. C. Cho, and K. H. Ryu, “Trigger Learning and ECG Parameter Tuning for Real-Time Cardiac Clinical Information System”, *IEEE Transactions on Information Technology In Biomedicine 2012,* Vol.16 No.4, 2012, pp.561-571

[9] M. Ezzeldin A. Bashir, H. S. Shon, D. G. Lee, H. Kim, and K. H. Ryu, ”Real-Time Automated Cardiac Health Monitoring by Combination of Active Learning and Adaptive Feature Selection”, *KSII Transactions on Internet and Information Systems*., Vol. 7, No.1, 2013, pp.99-118

[10] Y. H. Hu, S. Palreddy, and W. J. Tompkins, Eds., “Patient adaptable ECG beat classification using mixture of experts”, *in Neural Network for Signal Processing V. Piscataway, NJ: IEEE Press*, 1995, pp. 495–463

[11] UCI Machine Learning Repository, “http://www.ics.uci.edu/~mlearn/MLRepository.html”.

[12] WEKA web site, “http://www.cs.waikato.ac.nz/~ml/weka/index.html”