RPS/GMM Approach toward the Localization of Myocardial Infarction

MA Mneimneh, RJ Povinelli

Marquette University, WI, USA

Abstract

Due to the lack between clinical methods and applications used to diagnose ischemic heart disease, the 2007 Physionet/Computers in Cardiology challenge focuses on the ability to identify the segments, extent, and centroid of infarcts through ECG signals and body surface maps. The results from the participants are compared to a gold standard that consists of expert analysis of gadolinium-enhanced MRI data. The main hypothesis in this work is that the ordinary 12 ordinary leads contain the necessary information to identify the segment of the infarct. This hypothesis is tested using a reconstructed phase space and Gaussian Mixture Model approach in order to identify the infarcted segments. Since the challenge dataset consists of only two records for training and two for testing, the RPS/GMM approach is trained on the infarcted records from the PTB Diagnostics database and tested on the challenge data. The final score for the classification method was 1.15 out of maximum of 2.

1. Introduction

Due to the disconnections between clinical scientists and applications of the available diagnosing methods, the current methods for detecting ischemic heart disease are not as well developed as therapeutic interventions despite the knowledge regarding the underlying pathophysiology. Thus, the 2007 Physionet/Computers in Cardiology Challenge embraces the disciplines of both electrocardiography and Magnetic Resonance Imaging (MRI). However in this work we propose the use of the 12 leads ECG signals to locate, according to the standard 17 segments [1], the infarcted segments, the centroid of the infarction, and the extent of the infarct. The results of the challenge are compared to gold standards taken from MRI with gadolinium enhanced images. A Reconstructed Phase Space (RPS)/ Gaussian Mixture Model (GMM) [2] signal modeling approach is used to identify the infarcted segments.

The paper is organized in the following sections: background, data set, method, results, discussion and conclusion.

2. Background

The American Heart Association [3] shows how the 12 ordinary leads can be related back to infarction locations. The drawback in the provided rules is that they do not relate to the 17 segment locations provided in this challenge. In [1], de Luna and Wagner provided myocardial infarction localization rules that relate the Q wave, R wave, and S wave to the 17 wall segments using MRI.

Several research groups have used ECG measurements to locate infarcts. These include the work by Reddy et al. In [4], they developed an approach based on QRS measurement and neural networks to classify healthy and patients with myocardial infarction. Their accuracy was 79% with a specificity of 97%.

Lu et al's neural-fuzzy approach for classifying myocardial infarction uses ST elevation as an input parameter. Their approach identified whether a lead is infracted, and from the infarcted lead set inferred the infarct location. Their accuracy for detecting healthy patients was 89.4% and for detecting infarcted patients was 95.0% [5].

3. Datasets

The challenge dataset consists of two training and two testing patient records. One record has a moderate to large infarct, while the second has a relatively compact infarct. This data is labor-intensive to produce. The Body Surface Potential Map data, consisting of ECG data for 352 torso-surface sites, is provided for an averaged PQRST complex signal sampled at 2kHz. The 12 ordinary leads and the Frank leads are also provided [6].

However, this dataset is not sufficient for building a classification method. Therefore, we supplement the current challenge dataset with the PTB Diagnostic ECG Database, which provides a dataset of healthy and infarcted patients that is used as a training set for learner our classifier. The PTB database contains 549 records from 294 subjects. From the 549 records, 367 records taken from 148 patients had myocardial infarction [7].

4. Method

The RPS/GMM approach has been applied previously in the detection of myocardial ischemia [2]. Here, we apply the RPS/GMM approach to the 2007 Challenge.

The RPS embedding takes advantage of the effect of the myocardial infarction location in the ECG signal as represented through time. The time embedding determined from the ECG signal is used with GMMs to determine the infarcted type/location. The approach is applied on a lead by lead basis to decide which leads indicate an infraction. The set of infarcted leads generated by the algorithm are then used with the labeling method provided by [1] to localize the infarcted segments.

4.1. Preprocessing, baseline removal, and temporal filtering

The signals from the PTB diagnostics database are averaged to minimize the power line noise and temporal artifacts in the signal. Moreover, the baseline wandering is removed using a median filter of order N = 200. The temporal artifacts are removed by averaging all good beats. The beats are detected using ECGPUWAVE [7] and aligned using their cross-correlation.

4.2. Leads labeling

The PTB database provides the diagnosis for each of the patients, which can be related to the leads which might show the variation in the ECG signal. The labeling of the leads is according to Table 1 and is taken from [3]:

Table 1: Infarcted walls and respective leads

Wall affected	Infarcted leads
Septal	V_1, V_2
Anterior	V_3, V_4
Inferior	II, III, aVF
Lateral	I, aVL, V_5, V_6

For any combination of two or more infarcted locations, the leads for all cases are labeled as infarcted. The translation between the lead labeling and the 17 segment rule is determined from [1].

4.3. RPS/GMM approach

A reconstructed phase space (RPS) is a time delayed embedding of a signal, which may be topologically equivalent to the state space of the system that generated the signal if certain assumptions are met [8]. We have shown in previous work that even when these assumptions are not met RPSs contain important information for classifying a signal[9]. Here these signals are the 12 lead ECGs, and the classes are infarcted or non-infarcted. The definition of each point in an RPS is determined as follows [2]:

$$X_n = \left[x_{n-(d-1)\tau} \quad \cdots \quad x_{n-\tau} \quad x_n \right], \tag{1}$$

where N is the dimension of the space, t is the time delay, and d is the dimension.

GMMs are a set of Gaussian probability density functions used to characterize the distribution of an underlying set of data. They are widely used in engineering applications, especially speech [9]. The equation that defines a GMM is [2]:

$$p(x) = \sum_{m=1}^{M} w_m p_m(x) = \sum_{m=1}^{M} w_m \mathrm{N}(x; \mu_m, \Sigma_m), \tag{2}$$

where x is the feature vector, M is the number of mixtures, $\mathrm{N}(x; \mu_m, \mathring{a}_m)$ is a normal distribution with mean μ_m and covariance matrix Σ_m, and w_m is the mixture weight, with the constraint that the weights sum to unity. The GMM is estimated using Expectation-Maximization (EM). The GMMs are used in a Bayesian maximum likelihood classifier [10]:

$$p(X \mid c_i) = \prod_{n=1}^{N} p(x_n \mid c_i) \tag{3}$$

$$\hat{c} = \arg\max_i p(X \mid c_i) \tag{4}$$

where x_n is the nth feature vector, X is the set of all feature vectors, and c_i is the ith class.

The RPS/GMM approach uses the phase space embedding in order to create a multidimensional representation of the ECG signal that allows for the differentiation between a lead that indicates an infarction and one that does not. The GMM is used to model the embedded ECG signal and then used to compare against unknown signals to determine if an infarct is indicated. A block diagram describing the process is presented in Figure 1. In order for the RPS/GMM approach to accurately identify the infarcted leads, the signals should be of the same number of samples and the R peaks aligned at the same time sample.

The infarction size estimation is determine using the method by Selvester [11]. Selvester's method provides the relation between the infarction size and the Q wave, R wave, S wave and T wave variations, elongations, and magnitude variation. The system is based on a 50 points scoring system, where each point represent 3% of the infarct size.

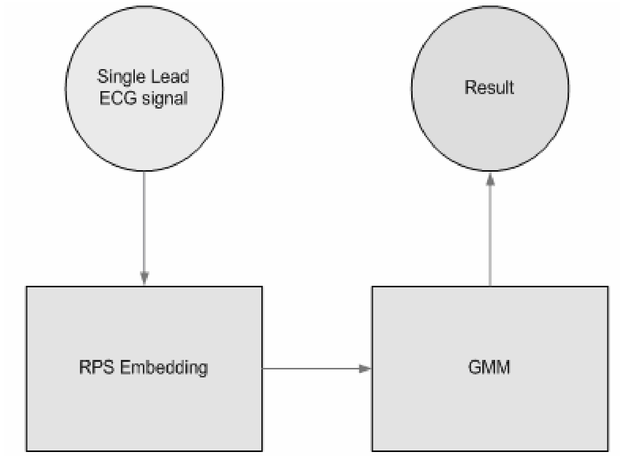

Figure 1: Block diagram describing GMM/KLT approach

4.4. Centroid location identification

Localizing the centroid is determined manually after identifying the infarcted segments. The segment at the center of the identified segments is chosen as the centroid.

5. Results

The proposed RPS/GMM approach is trained on the labeled data from the PTB diagnostics database and tested on the challenge's dataset. The approach is used to determine the location of the infarcted segments according to the 17 segmentation rules [12]. The parts of the ECG time series used in the classification process varied throughout the duration of the challenge entries as presented below.

5.1. Method for challenge scoring

The challenge includes three tasks:

1. Determine the extent of the infarct.
2. Identify the set of myocardial segments containing infarcted tissue.
3. Identify the myocardial segment containing the centroid i.e. the center of mass.

The results from each task are compared to a gold standard determined from gadolinium enhanced images of the chest. The gold standard images are given for the two training cases, but withheld for the two test cases. The infarcted locations for the test cases are hidden in the provided images.

The scoring for each of the entries is determined as follows[6]:

EPD: Determined from the percentage difference between the extent of the infarct as estimated and as determined from the gold standard.

SO: The degree of match between the sets of infarct segments as estimated and as determined from the gold standard. The degree of match is defined as the hamming distance between the gold standard segments and the estimated ones divided by the total number of either segments.

CED: The distance between the centroid of the infarct as estimated and as determined from the gold standard, where the distance is defined as the number of segment boundaries crossed along the shortest path connecting the estimates.

5.2. First pass

The first entry is the result of applying the RPS/GMM approach to the signals directly. The signals for the training and testing are used without any alignment. The resulting score for the estimation of the segment locations was 0.692 out of maximum of 2.

5.3. Second pass

In the second entry, the signals are resampled to 1KHz and the R peaks of the signals are aligned at the 500th time sample. After resampling and aligning the signals, the RPS/GMM approach is applied. The resulting score for this case was 0.555.

For this entry, an estimate for the infarction size is calculated using Aldrich's method [13]. The score for this approach is 35. Additionally, the centroid for the estimated segments is also generated and the resulting score is 2.

5.4. Third pass

For the third and final pass, the signals are processed to focus more on the QT segment. The signals are aligned as in the previous pass and resampled at 512 Hz. The RPS/GMM approach is then applied to the pre-aligned signals. The infarction percentage was calculated using Selvester’s approach [11]. The resulting score for the segment estimation is 1.15. The score for the infarct size is 27, and the CED is 1.

6. Discussion and conclusions

This paper presents an RPS/GMM approach for the identification of infarcted segments according to the generalized 17 segmentation rules. The approach is trained on the PTB diagnostics database and evaluated on the challenge testing set.

A difficulty in this challenge is the low number of training data. We address this problem by using the PTB database. However the PTB database does not provide a sufficient number of records for some of the infarction locations. In addition the exact segments are not labelled in the PTB database. The labelling is for the location of the infarction on the cardiac wall. This was then referred back to the leads that might show variations and thus the infarcted segments.

An additional difficulty faced in this challenge is in automatically assessing the quality of the averaged ECG signals determined from the PTB diagnostics database. During the averaging process, artefacts can be added to some signals due to alignment problems; especially at the QT interval. Additionally, some of the beats used in the averaging process might not show changes due to infarction, which will also affect the averaged beat.

In order to enhance the classification accuracy, the number of training and testing data should be increased. This can be performed by considering all the beats in each of the records in the PTB diagnostics database instead of using the averaged beat. The majority vote of the resulting classification of all the beats in a record may be a stronger indicator for identifying the infarcted location.

The purpose of this proposal is due to of the problems faced during the averaging of the beats. The averaged beat might contain additional artefact, especially when the beats are not identical. Thus, the majority vote of all the beats can help in neglecting the beats which do not show signs of infarction.

To conclude, a reconstructed phase space embedding with Gaussian mixture models classification approach is presented in this work for the identification of the infarcted segments from the 12 ordinary leads. The resulting score is 1.15 for the identification of the infarcted segments. The estimated infarcted segments helped in identifying the center of mass of the infarcted tissue. Additionally, the estimated infarcted leads were used to compute the infarction size using Selvester's method.

References

[1] Antoni Bayés de Luna MGW, MD; Yochai Birnbaum, MD;, Kjell Nikus MMF, MD; Anton Gorgels, MD, PhD;, Juan Cinca MPMC, MD; Olle Pahlm, MD;, Samuel Sclarovsky MSS, MD; Hein Wellens, MD;, Wojciech Zareba M. A New Terminology for Left Ventricular Walls and Location of Myocardial Infarcts That Present Q Wave Based on the Standard of Cardiac Magnetic Resonance Imaging. *Circulation* 2006, 114:1755-1760.

[2] Povinelli RJ. Towards the Prediction of Transient ST Changes**.** In *Computers in Cardiology*; *September 25-28; Lyon, France*. 2005

[3] Myocardial Ischemia, Injury and Infarction

[4] Reddy MRSE, L. Svensson, J. Haisty, W.K. Pahlm, O. Neural network versus electrocardiographer and conventional computer criteria in diagnosing anterior infarct from the ECG**.** In *Computers in Cardiology*; *11-14 Oct. 1992*. 1992: 667 - 670.

[5] HL Lu KO, P Chia. An Automated ECG Classification System Based on a Neuro-Fuzzy System**.** In *Computers in Cardiology*. 2000: 387-390.

[6] Moody G. Physionet / Computers in Cardiology 2007. 2007.

[7] Goldberger AL AL, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *AHA Journal Circulation 101* 2000, 101:e215-e220.

[8] Takens F. Detecting strange attractors in turbulence**.** In *Dynamical Systems and Turbulence; Warwick*. Edited by Dold A, Eckman B. Springer-Verlag; 1980: 366-381.

[9] Kevin M. Indrebo RJP, Michael T. Johnson. Sub-banded Reconstructed Phase Spaces for Speech Recognition. *Speech Communication* 2006, 48:760-774.

[10] Povinelli RJ JM, Lindgren AC, Ye J. Time series classification using Gaussian mixture models of reconstructed phase spaces. *IEEE Transactions on Knowledge and Data Engineering* 2004, 16:779-783.

[11] Anderson WD WN, Lee KL, White RD, Yuschak J, Behar VS, Selvester RH, Ideker RE, Wagner GS. Evaluation of a QRS scoring system for estimating myocardial infarct size. VI. Identification of screening criteria for non-acute myocardial infarcts. *Am J Cardiol* 1988, 61:729-733.

[12] M.D. Cerqueira NJW, V. Dilsizian, A.K. Jacobs, S. Kaul, et al. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart. *Circulation* 2002, 105:539-542.

[13] Aldrich HR WN, Boswick J. Use of initial ST-segment deviation for prediction of final electrocardiographic size of acute myocardial infarcts. *Am J Cardiol* 1988, 61:749-753.

Address for correspondence

Mohamed A. Mneimneh
EECE Department, Marquette University
1515 W. Wisconsin Ave.
Milwaukee, WI 53233
Mohamed.Mneimneh@marquette.edu