Data Driven Approach to ECG Signal Quality Assessment using Multistep SVM Classification

Jakub Kužílek[1], Michal Huptych[1], Václav Chudáček[1], Jiří Spilka[1], Lenka Lhotská[1]

[1]Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University, Prague, Czech Republic

Abstract

In response to the PhysioNet/CinC Challenge 2011: Improving the quality of ECGs collected using mobile phones we have developed an algorithm based on a decision support system. It combines couple of simple rules – in order to discard recordings of obviously low quality (i.e. high-amplitude noise, detached electrodes) with more sophisticated support vector machine (SVM) classification that deals with more difficult cases where simple rules are inefficient.

It turns out that complicatedly computed features provide only small information gain and therefore we used for SVM classifier only time-lagged covariance matrix elements, which provide useful information about signal structure in time.

Our results are 0.836.

1. Introduction

Cardiovascular diseases are the number one killer in the world [1]. The necessity of heart state evaluation is increasing and basic robust method for its evaluation is electrocardiography (ECG). The lack of specialists in many countries increases the need for easy and efficient measuring device, which could send measured data to specialist. Efficient measuring system can be obtained using smart phones. In order to inform inexperienced user about quality of measured ECG, the artificial intelligence based (AI-based) system was created. This should reduce the overall quantity of low quality ECG send to a specialist and by doing so spare his valuable time.

Our research aims at easily implementable and accurate method, which combines understandable rules and more sophisticated classification technique. It enables us to fine tune of sensitivity and specificity of detection. Our approach for evaluation of ECG quality is based on quality scoring of each record. Scoring and decision process is divided into three steps:

1. Application of simple rules
2. Classification of ECG using Support Vector Machines
3. Combination of scores from steps 1 and 2 into final decision

The first step detects most common errors encountered during measurement such as fell off electrode or presence of high amplitude bursts. In this step we are using simple rules only, which are based on easily computed characteristic of ECG signals. Each rule adds one point to the quality score.

The second step contains a feature based SVM classifier [2, 3], which deals with more difficult ECG cases. Feature vector used by the SVM classifier contains after several clustering experiments only covariance and time lagged covariance matrix elements. Other features did not prove their value for classification purposes. The SVM classifier deals with border cases and in order to make its contribution to final quality score, it scores a record with plus/minus one point.

The third and final step combines partial scores of the first two steps of our method and compares the final score with a defined threshold. Based on comparison with the threshold it decides whether the record should be accepted or not.

The paper is further organized as follows. First we will speak about the methodology. Then the results will be presented and finally their discussion is included.

2. Method

Our method contains three steps - the first one detects easily recognizable noisy ECG measurements using simple rules, the second one makes more precise classification using SVM classifier and the third one makes decision about record rejection. Schematic representation of the proposed method can be seen on Figure 1.

The process of data classification enables us to tune up sensitivity or specificity by modifying either one of the steps. First step makes classification sensitive to many common cases. It also detects many false positives cases,

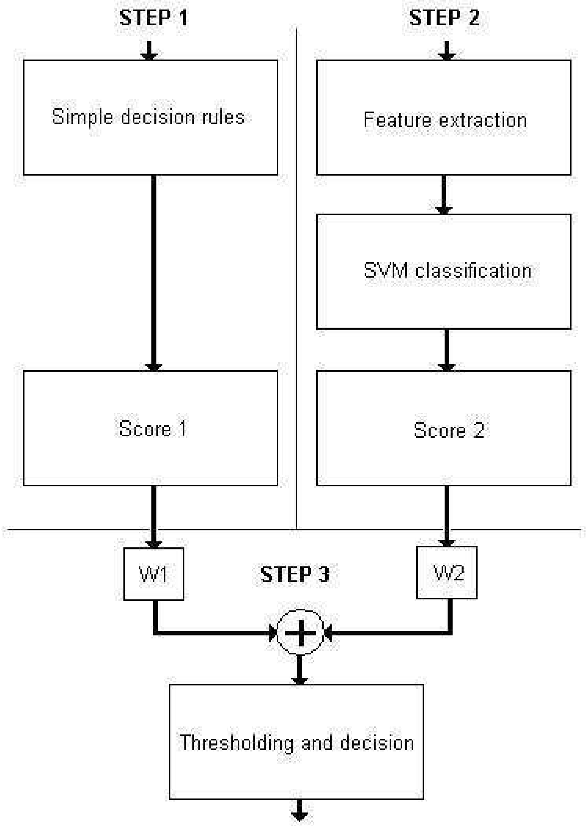

Figure 1. Data classification process.

which contain high amplitude peaks, but data is not destroyed by this artefact. The second step applied on data, which were classified as non-acceptable, gives us opportunity to improve specificity of our algorithm. First part of second step is the feature calculation. Next the SVM classifier takes linear combination of features and compares them with threshold to decide whether the record is acceptable or not.

Last step combines scores obtained by steps one and two. It can weight the score values using certain real number coefficients, which corresponds to value of each score to final decision. Finally combined score is compared to user-defined threshold and the decision is made. This step enables final tuning of algorithm using different weights for partial scores and different thresholds.

This section is further organized as follows - first we describe the rules applied to the data and then we will speak about SVM in general and specify its deployment in our algorithm.

2.1. Simple rules

The first step applies the set of simple rules on a ECG record. Each rule adds one point to the score 1.

Table 1 summarizes simple rules used in our method. We can see that these rules are based only on easily computed characteristics such as variance of signal and could be easily implemented in mobile devices. For construction of rules we used apriori knowledge about data and structure of ECG signals. We also used some basic features to estimate efficiently decision threshold for several rules.

Rules serve as fast and efficient classifier for selection of the most common cases, where noise is presented and records are not acceptable. These are sensitive and capture large portion of border cases, which are still acceptable.

In order to lower the portion of false positives (record is acceptable and rules classified it as non-acceptable) cases we employed non-linear SVM classifier, which corrects score for each record.

2.2. SVM classification

Support vector machines (SVM) is a method for estimation of parameters of linear classifier by minimizing of structural risk. This is done by setting of "restricted area" around decision border, in which no or in case of non-separable data only few training examples are presented. Estimation of parameters can be done by solving quadratic programming problem. In our case we used non-linear SVM, which is a simple extension of classic linear SVM done by introducing the kernel function.

Linear classifier is easily implementable to mobile application. Nevertheless in our case we are using set of quite complex features in order to obtain classifier that is efficient as well as robust and thus speed of the algorithm will not be our priority.

We applied non-linear SVM to data, which are classified as non-acceptable by first step of our method in order to tune up specificity of our algorithm.

As any other classifier SVM classifier works in the feature space. Our feature vector contains elements of covariance and time lagged covariance matrices. The other features such as mean, variance, kurtosis or number of QRS complexes found in the record were skipped due to small impact on final classification.
Time lagged covariance matrix [4] is defined as:

$$C_\tau^{\mathbf{x}} = E\{\mathbf{x}(\mathbf{t})\mathbf{x}^{\mathbf{T}}(\mathbf{t}-\tau)\},$$

where **x** is the signal vector (in our case of size 12) and τ is time lag. Time lagged covariance matrix contains information about correlation of differently switched signal and it should have almost the same values as normal covariance matrix in case of clear ECG, where temporal correlation is presented.

Table 1. Rules used in first step and their performance on train data

Rule	Noisy ECG covered	Clear ECG covered
Variance of any chest electrode signal is less than 10	139	0
Covariance on any electrode is greater than $5 * 10^8$	121	2
Any maxima of absolute value of signal on chest electrode is greater than 1200	168	10
Maximal value of hole ECG is greater than 600	190	26
Maximal value of difference of max and min of each signal is greater than 1000	199	26
Average value of variance of ECG is greater than $5 * 10^3$	164	23

2.3. Score combination and thresholding

Scores from steps 1 and 2 are passed to the final stage of algorithm, which weights them using real valued weights W1 and W2. For weights values we applied condition:

$$W1 + W2 = 1.$$

After weighting scores are summed up and compared to the threshold, which is set currently to zero. Zero means that record is most probably noiseless. If the score is greater than zero then the record is marked as non-acceptable.

Change of weights and threshold enables final fine tuning of classifier performance.

3. Results

Our combined rule-SVM method for estimating of acceptability of 12 lead ECG record achieved score 0.836 on the testing data set. On the training data we achieved 99.9 percent of correctly classified records. Results are shown in Table 2.

Table 2. Results of the method

Dataset	Result
Training	0.999
Testing	0.836

4. Discussion and conclusions

We proposed a fast and accurate method for estimation of ECG signal quality. Our method combines sensitive fast rule-based classifier with more precise non-linear SVM classifier, which can be still implemented as linear classifier. Results of these two classifiers are combined together using weights, which gives us another option for fine tuning of the algorithm performance. The final score, which is the combination of scores obtained by rule-based and SVM classifiers, is then compared to threshold and decision of record acceptance is made.

As we described above the rules are based on easily computed characteristics of ECG signal and the SVM classifier works as a linear classifier on feature vector consisting of time-lagged covariance matrix elements, which can be also easily calculated. So the implementation in mobile phone application should be fast, accurate and easily modifiable.

Acknowledgements

Research described in the paper has been supported by the ČVUT Grant SGS10/279/OHK3/3T/13 and research program No. MSM 6840770012 ”Transdisciplinary Research in Biomedical Engineering II” of the CTU in Prague.

References

[1] Moody GB, Silva I. Physionet/computing in cardiology challenge 2011, July 2011. URL http://physionet.org/challenge/2011/.

[2] Duda RO, Hart PE, Stork DG. Pattern Classification. 2nd edition. Wiley-Interscience, 2000.

[3] Franc V, Hlaváč V, Schlesinger MI. Statistical pattern recognition toolbox. online. URL http://cmp.felk.cvut.cz/cmp/software/stprtool/index.html.

[4] Belouchrani A, Abed-Meraim K, Cardoso JF, Moulines E. A blind source separation technique using second-order statistics. Signal Processing IEEE Transactions on 1997; 45(2):434–444.

Address for correspondence:

Jakub Kužílek
Czech Technical University in Prague
Faculty of Electrical Engineering
Department of Cybernetics
Technicka 2
166 27 Prague 6
Czech Republic
kuziljak@fel.cvut.cz