Algorithm for detection the QRS complexes based on support vector machine

Van G.V.*, Podmasteryev K.V.
Scientific-Educational Center of “Biomedical engineering”, Orel State University named after I.S.Turgenev, 29 Naugorskoe sh., Orel, 302020, Russia

*gennady.van@mail.ru

Abstract. The efficiency of computer ECG analysis depends on the accurate detection of QRS-complexes. This paper presents an algorithm for QRS complex detection based of support vector machine (SVM). The proposed algorithm is evaluated on annotated standard databases such as MIT-BIH Arrhythmia database. The QRS detector obtained a sensitivity $Se = 98.32\%$ and specificity $Sp = 95.46\%$ for MIT-BIH Arrhythmia database. This algorithm can be used as the basis for the software to diagnose electrical activity of the heart.

1. Introduction

QRS complex is a dominant of electrocardiographic signal, Figure 1. Quantitative and qualitative analysis of the QRS - complex as well as the time sequence of its appearance is one of the main tools in heart disease diagnosis. The deviation of the QRS-complex form, the change of its amplitude and time characteristics and physiological rhythms disturbance are caused by various heart diseases. The problem of exact finding of a QRS complex position on the recording and defining the diagnostics of significant characteristics becomes obvious. However, the solution of this task is complicated by personal variability, wide ranges of the norm, and also by the presence of noises and artefacts of different origin in ECG signal, Figure 2.

Figure 1 – Electrocardiogram of a single heartbeat. Modified from source picture (url:http://en.wikipedia.org/wiki/Electrocardiography#/media/File:EKG_Complex_en.svg)

Digital signal processing development led to significant improvement of QRS-complexes detection algorithms. These algorithms can be classified by the mathematical apparatus taken as their bases:
algorithms based on time and frequency domain transformations; syntactic algorithms; combined algorithms; algorithms based on machine learning methods [1].

However, despite the consistent development of new algorithms, the problem of fast and accurate QRS detection is still urgent. One of the most promising approaches is the use of artificial intelligence methods and machine learning. These methods are relatively easy to use, allow real-time implementation and have high sensitivity and specificity.

The aim of this study was to develop an algorithm of QRS-complexes detection. This algorithm is based on a support vector machine method because this method possesses high efficiency and high-speed performance that allows it to be used in online time [2, 3].

Figure 2 – Noisy ECG signal. Adapted figure from [4].

2. Methodology

In this section, we describe the proposed algorithm for the detection of QRS-complex in MIT-BIH arrhythmia database using SVM classifier. The algorithm's steps are considered in Figures 3 and 4. The implementation of the algorithm and the results of the research were received in Jupiter for the Python 3 programming language with the use of the machine learning library scikit-learn.

Figure 3 – Schematic representation of intermediate steps for SVM algorithm implementation.
2.1. ECG databases

For the training and testing the signals from the MIT-BIH Arrhythmia database are used. This database consists of 48 half-hour recordings for a total of 24 h of ECG data. Each one has a duration of 30 min and include two leads – the modified limb lead-II and one of the modified leads V1, V2, V4 or V5 [5], sampled at 360 Hz with resolution of 5 µV/bit. Two cardiologists have annotated all beats. This 24 h MIT-BIH database contains more than 109,000 beats. A typical signal from MIT-BIH Arrhythmia database is shown in Figure 5 [6, 7].

2.2. ECG signal pre-processing

The received raw digital ECG signal of a record is shown in Figure 4a. For the increase of efficiency of an algorithm the procedure of preliminary processing of a signal is used as the signal contains different noise and artifacts.

For the conversion a low pass filter with a cut-off frequency of 13 Hz and high pass filter with a cut-off frequency of 9 Hz are used, as well as the function of the moving average with a 5 measurements window sizes. The filtered ECG signal is shown in Figure 4b. In addition to such informative sign as rise speed of a signal, the filtered signal is squared, Figure 4c.

2.3. Informative features

In this stage, the adaptive threshold values by means of the informative features are created. For this algorithm the combination of such informative features as rise speed of a signal (as QRS complex has the
greatest climb rate), and the correlation of QRS complexes forms were selected (as QRS complex has a specific characteristic form) are used. [8, 9].

2.3.1 Rise speed of signal
Informative feature as rise speed of a signal is chosen, because QRS complex possesses the greatest climb rate. This informative feature is implemented as follows: an array of the values of the slope of the tangent to each point of the filtered and squared ECG signal. After that SVM classification function is applied to the received selection.

2.3.2 Correlation of QRS complexes forms
Informative feature as correlation forms of QRS-complexes are chosen, because QRS-complex possesses a specific form. This informative feature is implemented as follows: Firstly, test pulse is created based on 1000 QRS-complexes with the R-peak in the middle lasting 51 counting (141.67 ms), which occurs after their averaging, for this procedure the signal №100 is used. Secondly, an array of correlation coefficients forms of QRS-complexes in a moving window is created. After that SVM classification function is applied to the received selection.

2.4. Training phase
For train recording signals №100, №104, №214, №200, №205 are used. The training set consists of 30,000 values for each of the signs. The function of classification was calculated for each informative feature separately. Using data that indicate the R-peak without pathology, as well as data that correspond to 120 counts (333.34 ms) before and after the R-peak without pathology.

2.5. Classification phase
The main condition for testing algorithm was the use of records which do not participate in training. This was the reason that, if testing is done with the pattern used in learning or training, the accuracy will be artificially high. Signals from MIT-BIH database with 30 min of duration are used for testing. After testing, a train of 1’s is obtained at the output of SVM classifier. Then this train of 1’s is picked and by using their duration, average pulse duration of 1’s is evaluated. Those trains of 1’s, whose duration turns out to be more than the average pulse duration are detected as QRS-complex and the others are discarded. The locations of the QRS-complexes, as detected by SVM, are shown in Figure 4d.

3. Testing and results
For the calculating of algorithm efficiency such indicators were used:
Algorithm efficiency can be estimated by calculating specificity $Sp$ – a probable assessment of absence of false positive result and sensitivity $Se$ – a probable assessment of correctness of a QRS complex detection:

$$Sp = \frac{TP}{TP + FP}$$

$$Se = \frac{TP}{TP + FN}$$

where TP is a correct detection of a QRS-complex; FN is when the algorithm fails to identify a QRS-complex; FP – if algorithm detects a non-QRS-complex as a QRS-complex [7]. The efficiency indicators of this algorithm are: sensitivity – 98.32%; specificity – 95.46%.

4. Conclusion
The efficiency indicators of this algorithm are: sensitivity – 98.32%; specificity – 95.46%. Besides, this algorithm possesses a high-speed performance that allows it to be used in real time.

Further increase of efficiency of this algorithm is possible due to the use of bigger quantity of informative signs; changes of training procedure and teaching selection; changes of machine training method which has been the basis for the algorithm; changes of preliminary signal processing procedure; changes of output data interpretation procedure.
References

[1] Elgendi M. et al., “Revisiting QRS detection methodologies for portable, wearable, battery-operated, and wireless ECG systems,” PLoS One, 2014. vol. 9, 2014.

[2] Van G.V., Podmasteryev K.V. Review algorithm for detection the QRS-complex based on machine learning // Symposium Proceedings of 2nd International Scientific Symposium “Sense. Enable. SPITSE.” 2015. (St.Petersburg, Russia, 22 June - 03 July 2015). - P. 61-63.

[3] Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. — Springer-Verlag, 2009.

[4] S. J. Redmond, Y. Xie, D. Chang, J. Basilakis, and N. H. Lovell, “Electrocardiogram signal quality measures for unsupervised telehealth environments,” Physiol. Meas., vol. 33, no. 9, pp. 1517–33, Sep. 2012.

[5] Massachusetts Institute of Technology. MIT-BIH ECG database, <http://www.physionet.org/cgi-bin/atm/ATM>.

[6] Moody G.B, Mark R.G. The impact of the MIT-BIH Arrhythmia Database // IEEE Eng in Med and Biol. 2001. P.45-50. Vol 20 №3.

[7] Saini I. et al. QRS detection using K-Nearest Neighbor algorithm (KNN) and evaluation on standard ECG databases // Journal of Advanced Research. 2013. Vol. 4, 331–344, №4.

[8] Van G.V., Podmasteryev K.V. The analysis of informative features for R-peak detection // XII International Scientific Conference «Physics and Radioelectronics in Medicine and Ecology» with Scientific Youth session PhREME’2016 (Vladimir - Suzdal, Russia, 5-7 July, 2016), Proceedings, Book II. - P. 126-129.

[9] Hampton J.R. The ECG made easy. London: Churchill Livingstone, 2003.