QRS complex classification (normal, abnormal) using norms of linear algebra

Klemen Škrlj
Faculty of Computer and Information Science
Univeristy of Ljubljana
ks5379@student.uni-lj.si

Abstract

We implemented a model for QRS complex classification using different norms of linear algebra. Our system is able to differentiate between normal and abnormal heart beats that are caused by ventricular ectopic. We first filter raw ECG signal with high pass filter to remove baseline drift. Than two different strategies for initial threshold are shown. We compare QRS complexes between each other using three different norms by itself or in majority voting schema. Best model is chosen in respect to the domain that it is used in and trade-off between sensitivity and specificity.

1. Introduction

ECGs are important biomedical signals, which are reflective of an electric activity of the heart. Every ventricular depolarization of the heart reflects in a new QRS complex in the signal. This is a combination of three graphical deflections (Q, R and S points). When such complexes are detected, a classification has to be made for every one. There are mainly 2 classes: normal and abnormal heart beat.

Since no human is like another, there are some physiological differences in the signal and various types of noises which pose different challenges for accurate classification. Noise is typically a consequence of muscle noise, artifacts due to electrode motion, power line interference and baseline wander. We want a system that automatically distinguishes between before mentioned classes and thus demands minimal human interaction.

2. Methods

In our exercise, we decided to implement a system that compares and classifies QRS complexes using different norms of linear algebra. Here we have a binary classification problem where one class represents normal heart beat and the other one is a consequence of ventricular ectopic heat beat (abnormal).

Because our method uses signal sections in their raw form, we need to make them comparable between each other. Here a problem of baseline drift occurs. If we don't remove it, we compare one QRS complex that is straight and one that is slanted. Even though those two might be exactly the same, the norm wouldn't show that. We used High-pass recursive filter defined by Equation 1 to mitigate this problem.

$$
\begin{aligned}
H(z) &= \frac{c_1(1-z^{-1})}{(1-c_2 z^{-1})} \\
c_1 &= \frac{1}{1+tan(F_c \pi T)} \\
c_2 &= \frac{1-tan(F_c \pi T)}{1+tan(F_c \pi T)}
\end{aligned}
\tag{1}
$$

Here F_c is set to 2.2Hz and represents cut-off frequency and T is sampling frequency. As seen on Figure 1 the baseline wonder is negated after this filter is applied.

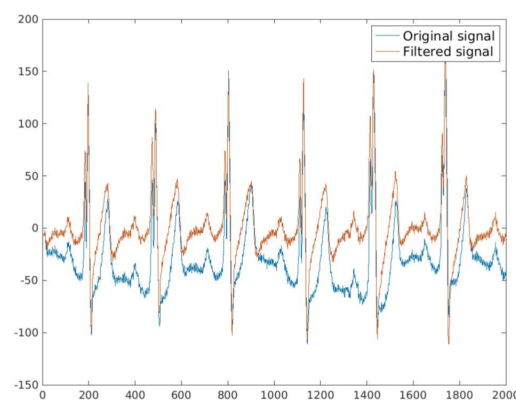

Figure 1. Comparison between original and filtered signal

After that, our training process begins. To distinguish between normal and abnormal beats we need a representation of some true positive examples. We take first five minutes of each record and average out all normal heart beats. Given a location of fiducial point FP, the beat is defined as signal in range:

$$[FP - 60ms, FP + 100ms]$$

Then two methods for initial threshold are tested. In the first one we calculate chosen norm between this average normal beat and 50% of the same one and this gives us our threshold value. In the second example, we calculate norm between average and all other normal beats in the training portion of the signal. Our initial threshold is then an average norm value multiplied with a constant C, which we chose empirically (between 2 and 2.5 depending on the norm).

After we have our initial threshold, we begin with prediction phase. Here we go over all predefined fiducial points, extract the whole beat and calculate norm of linear algebra to the average normal beat. In our experiment we use 3 different norms, each one defined in Equation 2.

$$\begin{aligned} d1 &= \frac{1}{N}(|x1 - y1| + ... + |x_N - y_N|) \\ d2 &= \sqrt{\frac{1}{N}(|x_1 - y_1|^2 + ... + |x_N - y_N|^2)} \\ d_\infty &= max(|x_1 - y_1|, ..., |x_N, y_N|) \end{aligned} \quad (2)$$

We also employ our adaptive threshold technique since static threshold value is more prone to errors when signal characteristics change over time. We got inspiration for this formula from [1] and [2] where they used similar technique but on differently pre-processed signal. When we get a new beat that is below the threshold, we correct the threshold with Equation 3, where C is a constant between 2 and 2.5 and α is a factor controlling how much influence does current value have on a threshold. Empirically we chose values between 0.005 and 0.0005.

$$T = \alpha \cdot C \cdot currentNorm + (1 - \alpha) \cdot T \quad (3)$$

Since we have three different ways of calculating if beat is normal or not, we also test majority voting scheme. Here we calculate initial threshold based on our second technique and then perform thresholding with static values and with adaptive values. For each new beat we calculate all three norms and check how many of them are higher than their thresholds. If 2 or more values go over threshold then this beat is classified as an abnormal one.

3. Results

We tested multiple versions of the algorithem on MIT/BIH database[3], where the records that don't have any normal beats during our training period were discarded (7/48 records). Results are shown in Table 1. Number at the beginning of the name represents first or second technique for initial threshold, "Adapt" represents adaptive thresholding and "MV" stands for majority voting scheme.

Name	Se	Sp	+P
1-D_1	90.9	94.0	99.4
1-D_2	90.5	94.8	99.4
1-D_∞	**91.4**	**95.5**	**99.5**
2-D_1	91.5	98.1	99.8
2-D_2	90.8	**98.7**	**99.8**
2-D_∞	90.9	95.0	99.5
2-D_1-Adapt	96.6	95.0	99.5
2-D_2-Adapt	**96.6**	95.0	99.5
2-D_∞-Adapt	93.2	92.1	99.2
2-MV	90.8	94.7	99.4
2-MV-Adapt	**90.9**	**98.1**	**99.8**

Table 1. Results

For evaluating algorithms we choose sensitivity calculated by formula

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

specificity

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

positive predictivity

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

where TP denotes true positive detections, FP false positive and FN false negative. The highest possible value of all formulas is 100%.

4. Discussion

When choosing the best constants and α values for our algortihem we had in mind the notion that specificity is a bit more important than sensitivity in this domain. In best case we wouldn't want to miss any abnormal heart beats, but we can live with some false positive detections where normal heart beat is classified as abnormal. Results show that norm d_∞ performed the best when using 1. strategy for initial threshold. When using 2. strategy static threshold with d_2 norm had the best Sp and +P, while adaptive thresholding helped with higher sensitivity and overall more balanced ratio between Se and Sp. Majority voting with adaptive threshold was better than the one with static threshold but there is again drop in Se and increase in Sp. Overall model "2-D_2-Adapt" seems to have the best trade off between Se and Sp.

References

[1] H. Chen and S. Chen, "A moving average based filtering system with its application to real-time QRS detection," in *Computers in Cardiology, 2003*, IEEE, 2003.

[2] S. Chen and H. Chen, "Development of a real-time QRS beat classifier using a nonlinear trimmed moving averaging-based system," in *Computers in Cardiology, 2003*, IEEE, 2003.

[3] G. Moody and R. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[4] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet," *Circulation*, vol. 101, June 2000.