Report on ECG Data Classification

115033910044 高晨昊

1. Introduction

**Electrocardiography** (**ECG** or **EKG**[\*](https://en.wikipedia.org/wiki/Electrocardiography#Note)) is the process of recording the electrical activity of the heart over a period of time usingelectrodes placed on a patient's body. These electrodes detect the tiny electrical changes on the skin that arise from the heart muscle depolarizing during each heartbeat.

In the given data we can see that they are classified into 6 groups representing 6 different status of heart.In each group a diagram are divided into different heartbeats.A heartbeat consisits of 12 different graphs.Our job is to extract features from such graphs.

In my work I used 2 different methods to clssify such heartbeats: Support Vector Machine(SVM) and Backpropagation Neural Network(BPNN).

1.1 **Support Vector Machine**

In machine learning, **support vector machines** (**SVMs**, also **support vector networks**) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked for belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

We used the **LIBSVM** -- A Library for Support Vector Machines.

* 1. **Backpropagation Neural Network**

The back propagation (BP) neural network algorithm is a multi-layer feedforward network trained according to error back propagation algorithm and is one of the most widely applied neural network models. BP network can be used to learn and store a great deal of mapping relations of input-output model, and no need to disclose in advance the mathematical equation that describes these mapping relations. Its learning rule is to adopt the steepest descent method in which the back propagation is used to regulate the weight value and threshold value of the network to achieve the minimum error sum of square.

Fortunately,in Matlab a function of BP Neura Network has been integrated as newff(), so we can use it directly

1. Feature Extraction

Feature extraction might be the most challenging work in this task.Since we know little about the traits of ECG graphs,we had to compare the different graphs and find some useful information.

For the first 4 class we take a look at the I graph of 12 feature graphs at all the heartbeat of the first patient in the class, they are showed as follows:

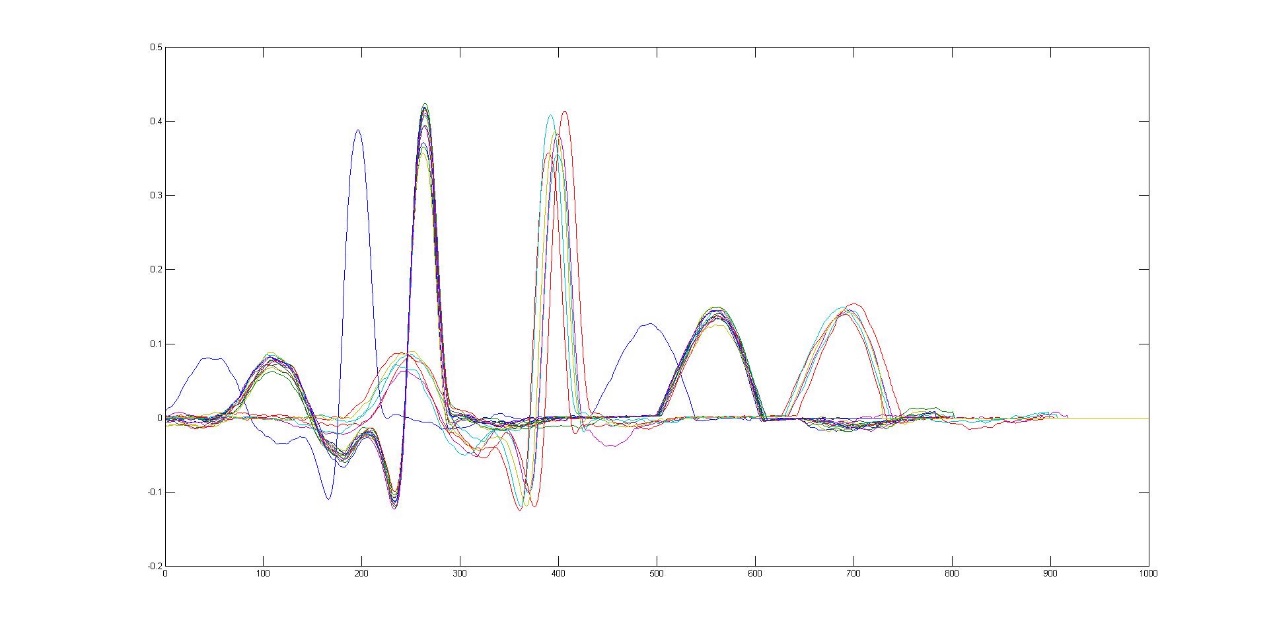


Fig1.k=1(Electrical axis left side)

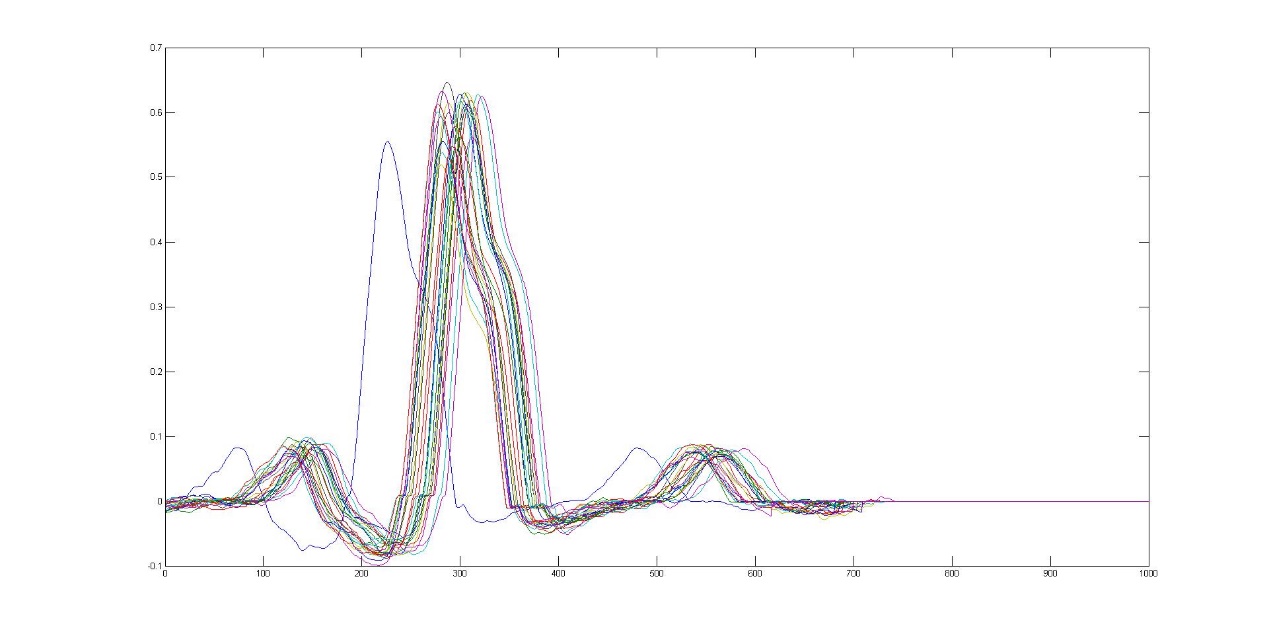


Fig2.k=2(Left bundle branch block beat)

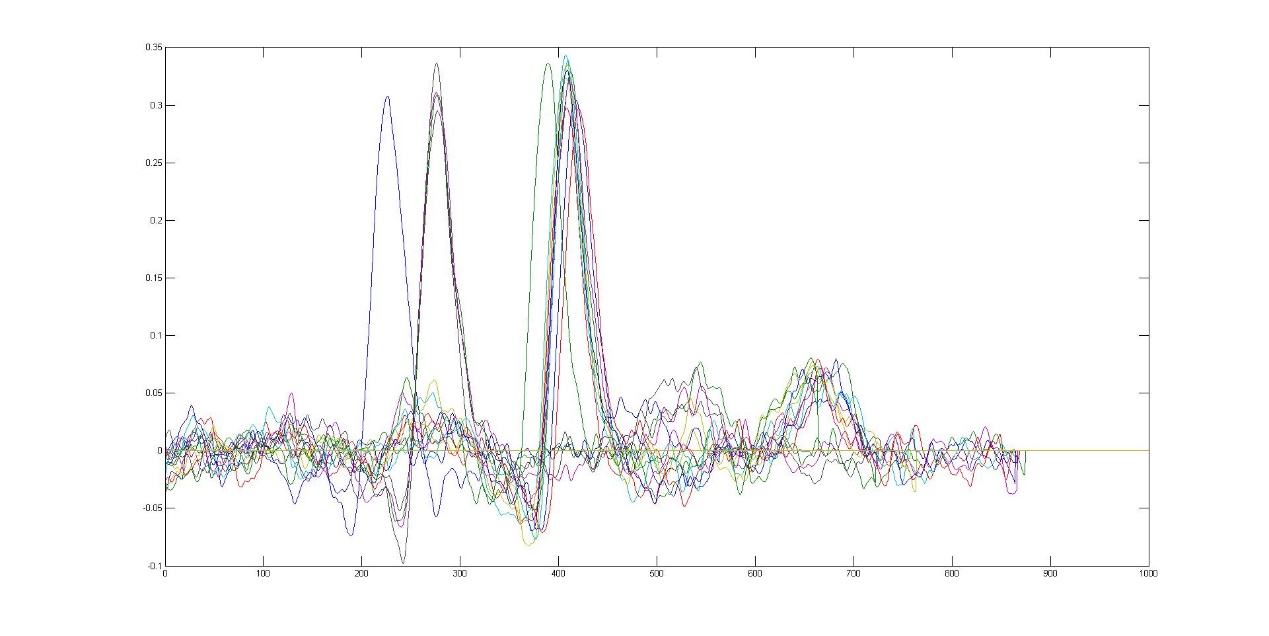


Fig3.k=3(Left ventricular hypertrophy)

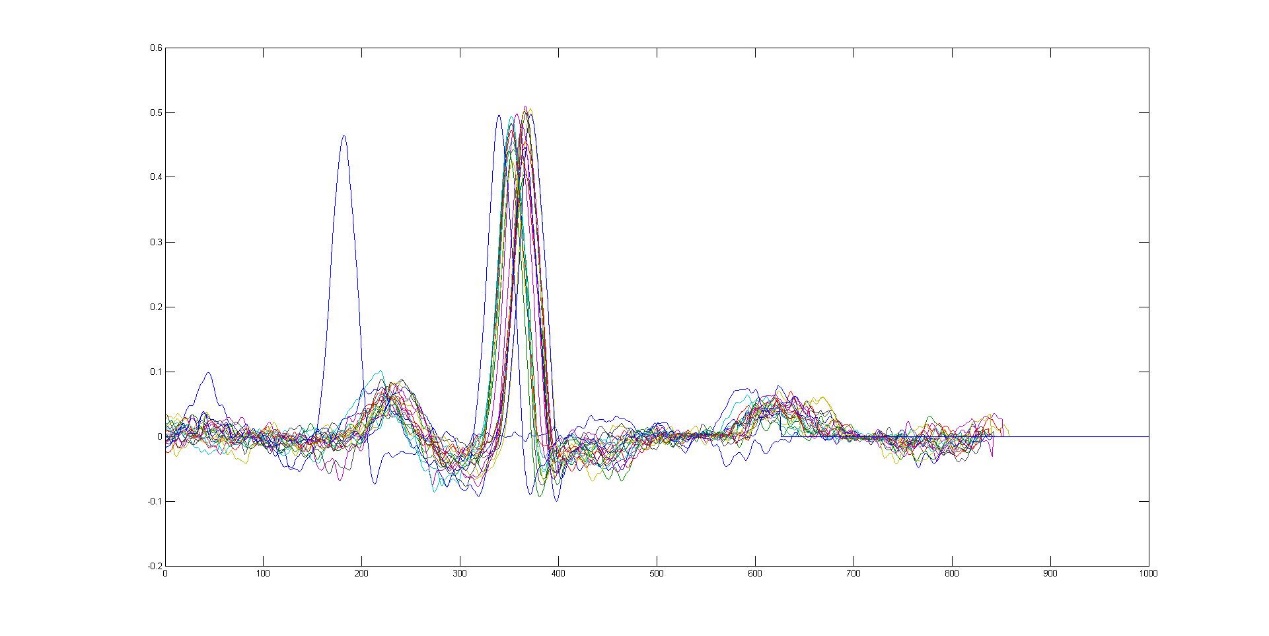


Fig4.k=4(Normal)

From the 4 figure we deduce that

* The curves of heartbeats is simalar according to different classes, but the difference between the classes are obvious
* There are always delayed or forward curves in the dataset,but there waveforms share simalarities, which means we need to detect features from the frequency domain rather than time domain
* The peaks of the wave may contain the most important information,but the whole wave also contian some information.We need to use middle frequency band of the wave to distinguish them.

Considering to the above all,we decide to use the Discrete Cosine Transform(DCT) to extract features from the graph.Compared to Discrete Fourier Transform(DFT), DCT coefficients are all real and easy to process.The first several coefficients of DCT display the approximatiom information while the left coefficients display the detailed information.After so many attempt we used the first 5 components of the DCT coefficients to represent the feature of a Graph.Then I joint the 12 group coefficients together to represent the feature of a heartbeat.

1. SVM classification

Thanks to the LIVSVM library,I have implemented the classification easily.As mentioned above, I used the 5\*12=60 coefficient to represent a heartbeat, then use the first to fifth .mat data of each class to train the SVM hyperplane.In the SVM funtion

model =svmtrain(training\_label\_vector, training\_instance\_matrix ) t raining\_label\_vector is the label vector of M\*1 matrix while training\_instance\_matrix is the M\*P matrix.In my configuration M=620 and P= 60 and this makes the best accuracy of 80.6452%.

Test data are the sixth to nineth .mat data of each class.The test number is 496.

Here is the comparison of different coefficients selection.

Table 1 Comparison of selection of DCT coefficients

|  |  |
| --- | --- |
| DCT coefficients | Accuracy |
| 1-5, P= 60 | 80.6452% |
| 2-6, P= 60 | 80.8468% |
| 3-7, P= 60 | 78.8306% |
| 1-6, P= 72 | 81.6532% |
| 1-7, P= 84 | 81.6532% |

We find that when using the 1-6 coefficients and P=6\*12=72 we get the highest accuracy of 81.6532%.Continuing to add the dimension of P does not increase the accuracy.

1. BP neural network classification

In the next step I used the BP neural network to do the clssification job.