The Abnormal vs. Normal ECG Classification Based on Key Features and Statistical Learning

Jun Dong[1], Jia-fei Tong[1], Xia Liu[2]
[1]Suzhou Institute of Nano-technology and Nano-bionics, Chinese Academy of Sciences, China
[2]Shanghai Ruijin Hospital, China

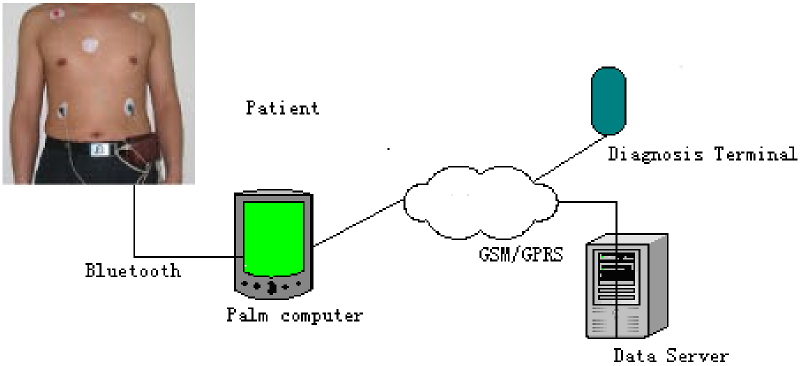
1 Background

Cardiovascular diseases as one of the most frequent and dangerous problems in modern society is much more difficult to take immediate measures without real time ECG (electrocardiogram) information.

Chinese government has planned to establish basic community medical insurance system(BCMIS) before 2020, where remote medical service is one of core issues.

2 Our Solution(1/3)

Therefore, we have developed the “remote network hospital system” which includes data server and diagnosis terminal by the aid of wireless detector to sample ECG.



2 Our Solution(2/3)

Originally, the physicians who are in charge of taking response in according to server request have to spend much time for distinguishing abnormal vs. normal ECG. It will be better and more efficient if normal ECG can be recognized automatically by computer and abnormal ECG left to be diagnosed specially by physicians. We are intended to solve the classification issue based on key features and statistical learning.

2 Our Solution(3/3)

In general, computer-aided ECG diagnosis includes preprocess, features recognition and classification processes. We have got better results when concerning imagery thinking and morphology features in recognition phase.

3 Features and Data(1/2)

In fact, some features are not used practically by physicians for diagnosing common diseases. Through experienced physicians, we got most important 14 features (third column, set Ⅱ) from physicians' experience.

Open source tool——ECGPuwave is used to extract features (second column), where, from 1 to 10 are amplitude features, and from 11 to 23 are interval features (set Ⅰ).

number	23 features(set Ⅰ)	14 feature(set Ⅱ)
1	P wave	√
2	R wave	√
3	R wave start point	√
4	R wave end point	
5	T1wave	√
6	T2wave	√
7	P wave start point	
8	P wave endt point	
9	T wave start point	
10	T wave endt point	
11	PP	√
12	QRS	√
13	PR	√
14	ST	√
15	QT	√
16	RR	√
17	TT	
18	P	√
19	T	
20	Total beat	√
21	T1 T2	
22	ST	
23	PR	√

3 Features and Data(2/2)

MIT-BIH Arrhythmia Database (lead Ⅱ)is used too. 101837 beats of 45 records are selected. With some unrecognized beats being ignored, 94315 beats are ready, among which 79110 beats are normal and 15205 beats are abnormal separately.

4 Statistical Learning

Open source SVM tool Libsvm is selected here. Two parameters, C and σ should be determined first of all. So two kinds of experiments are designed: C is fixed with 8000 and σ is fixed with $\frac{\sqrt{2}}{2}$. In Libsvm, $\gamma = \frac{1}{2\sigma^2}$ is set for RBF. Several results are demonstrated in table 2 to table 5 respectively.

4.1 Result(1/2)

Table 2. C = 8000 with features set I .

γ	0.01	0.03	0.05	0.09	1	10	50	100
SP	98.85	99.24	99.29	99.27	98.62	99.48	99.96	99.99
SE	89.7	91.98	93.3	93.88	94.87	89.23	60.65	38.86
GCR	97.36	98.06	98.31	98.39	98.01	97.81	93.55	90.02

Table 3. C = 8000 with features set II .

γ	0.01	0.03	0.05	0.09	1	10	50	100
SE	84.42	88.63	90.38	91.87	94.43	94.20	89.10	81.22
SP	97.73	98.30	98.55	98.69	98.93	98.36	99.29	99.65
GCR	94.90	96.25	96.82	97.25	97.97	97.48	97.13	95.74

4.1 Result(2/2)

Table 4. $\sigma = \frac{\sqrt{2}}{2}$ with features set I .

C	0.01	0.05	0.1	1	10	100	1000	5000
SP	99.90	99.45	99.39	99.36	99.35	99.24	98.95	98.75
SE	35.89	76.13	84.07	91.56	94.37	95.03	94.68	94.96
GCR	89.46	95.65	96.89	98.09	98.54	98.55	98.25	98.13

Table 5. $\sigma = \frac{\sqrt{2}}{2}$ with features set II

C	0.01	0.05	0.1	1	10	100	1000	5000
SE	43.53	76.75	82.75	90.10	92.43	93.34	94.00	94.26
SP	98.65	98.14	98.23	98.57	98.93	99.07	99.07	98.93
GCR	86.96	93.61	94.95	96.77	97.55	97.86	97.99	97.94

4.2 Parameters

As the tables show, when C is fixed, as γ increased, the correctness is improved. The result is optimized when $\gamma = 1$. On the other hand, when σ is fixed, as C increased, the correctness for normal beats is down till C=1. Then the correctness for abnormal beats is up fast.

4.3 Compared References

1. Übeyli E. D.: ECG Beats Classification Using Multiclass Support Vector Machines with Error Correcting Output Codes. Digital Signal Processing. 17, 675—684(2007)
2. Osowski S., Hoai L. T., Markiewicz T.: Support Vector Machine based Expert System for Reliable Heartbeat Recognition. IEEE Transactions on Biomedicine Engineering. 51(4), 582—589(2004)
3. Acır N.: A Support Vector Machine Classifier Algorithm Based on a Perturbation Method and Its Application to ECG Beat Recognition Systems. Expert Systems with Applications. 31, 150—158(2006)

5 Conclusion(1/2)

SVM needs only limited samples, give global optimized solution avoiding the local optimized problem using neural network. At the same time, its generalization capability is good. But how to select the kernel function is an intractable issue. Furthermore, large train set is left unsolved.

5 Conclusion(2/2)

The classification approach has been applied in “remote network hospital system” of medicine school of Shanghai Jiaotong University, China.

Now large amount of normal ECG have been filtered by computer automatically and abnormal ECG is left to be diagnosed specially by physicians. We are managing to combine several classifiers to get better results.

Thanks a lot!