2012 2nd International Conference on Biomedical Engineering and Technology
IPCBEE vol. 34 (2012) © (2012) IACSIT Press, Singapore

Classification of Cardiac Arrhythmia via SVM

Thaweesak Yingthawornsuk[1]

[1]Division of Biomedical Media Technology, Department of Media Technology,
King Mongkut's University of Technology Thonburi, Bangkok, Thailand

Abstract. This paper shows the classification results in term of performance of SVM classifier in separating the multi-class arrhythmia dataset. By pre-selecting sets of feature classifying the training dataset in pairwise-class fashion can provide most accurate score of separation. Selection of dimensional features can be unrestrictedly allowed in grouping classifications which are less complex in computation as compared to that using the constant dimension in classifying feature parameters.

Keywords: ECG, PCA, SVM, Classification.

1. Introduction

The electrocardiogram (ECG) provides significant clinical information of patients who have abnormal activity of heart. By using the ECG record physicians can classify the abnormality into which class the disorder belongs. However, in the normal case the ECG is recorded in a long time period. This is the time consuming and inconvenience for the physician since he needs to be in alert at all time. More importantly, arrhythmia events can be missed by the human detection due to the fatigue of long hours of work. The ECG record can be interpreted using a special machine that has an artificial intelligence inside. The detected arrhythmia, a pattern of ECG signal considered as abnormal heart functioning, event can be used as information for a pre-screening procedure so that the physician can pay attention to the suitable remedy. There are several approaches for classifying the ECG arrhythmia record [1-8]. Support vector machine (SVM) is a classification tool that outperforms several classification methods. Many researches including our previous work [9] show the procedure of applying the SVM to the classification of arrhythmia. As the two-class classification method, we investigate the performance of SVM whose input feature dimensions, at each classification step, are fixed and freely-selected.

2. Methods

2.1. Database

In this paper, the ECG signal is obtained from MIT-BIH Arrhythmia database [10]. There are three types of the arrhythmias considered here including normal beat (NORMAL), premature ventricular contraction (PVC) and arterial premature contraction (APC). We select 6 files from the arrhythmia database including file 100, 109, 107, 118, 208 and 232 which contain enough beats of normal, PVC and APC arrhythmia for the experiment. The provided ECG signal is sampled with 360Hz. The classification of aforementioned types is investigated since they are more likely to be incorrectly interpreted by computation machine compared to other waveforms. Fig. 1 shows general waveform of three arrhythmia beats.

The ECG signal from Modified Lead II (ML-II) is chosen for processing. The database is prepared with the R-peak locations and arrhythmia annotations so that we can skip the QRS detection and perform the feature extraction as described in later section. In our experiment, we use the R-peak of QRS complex to be the significant sampling point. By selecting from 120 samples prior to the R-peak to 179 samples after the R-peak, we have a pre-classified ECG record of 300 samples. This sequence occupies 333ms prior to R-peak sample and 497ms after the R-peak sample. In our experiment, each simulation, ECG beats are randomly selected from those 6 files so that dataset is composed of 1,000 beats of each type.

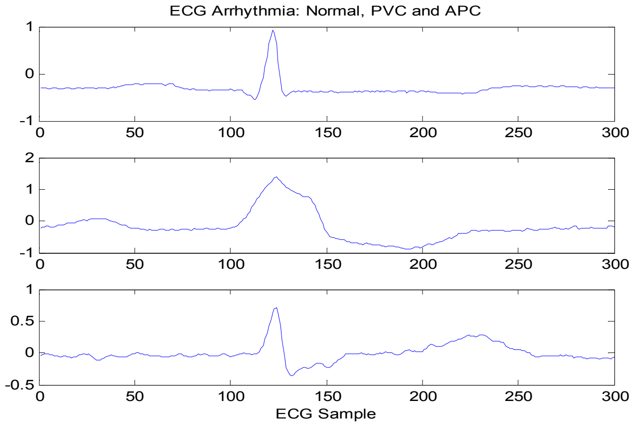

Fig. 1: Waveforms of ECG: Normal beat (N), premature ventricular contraction (PVC) and arterial premature contraction (APC)

2.2. Feature extraction

The 300-sample sequence is then decomposed using the Daubechies 4 (DB4) wavelet algorithm. To extract the feature of the ECG signal, we select the lower-frequency component for our consideration, since higher-frequency component usually is noise component. In this paper, with the effect of down-sampling, we take four levels of DB4 so that the sequence of 360-sample ECG is reduced to 21-sample wavelet coefficients of lower-frequency components. Fig. 2 shows the ECG signal of NORMAL arrhythmia and its wavelet coefficients.

The training and validation dataset are randomly selected from the whole dataset of row vectors so that there are 300 (100 for each type) and 2,700 (900 for each type) row vectors for training and validation dataset, respectively. The row vectors, as features, in training dataset are formed into a matrix $\boldsymbol{X}_{\text{tr}}$. The principal component analysis (PCA) is applied to matrix $\boldsymbol{X}_{\text{tr}}$ so as it provides the transformation matrix that can be used to reduce the row vector size to a smaller size vector. In our investigation the row vector size of 2, 3 and 4 are selected. The signal distributions of 2 and 3 dimensional features are shown in Figs. 3~4, respectively.

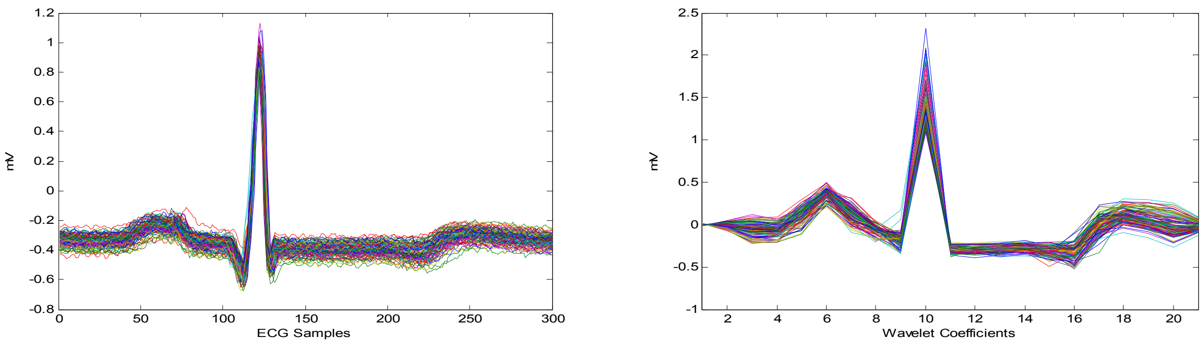

Fig. 2: NORMAL arrhythmia (left) and its wavelet coefficients (right)

2.3. Signal classification

The support vector machine or SVM [11] separates an input $\boldsymbol{x} \in \boldsymbol{R}^d$ into two classes. A decision function of SVM separates two classes by $f(\boldsymbol{x}) > 0$ or $f(\boldsymbol{x}) < 0$. The training data which is used in training phase is $\{\boldsymbol{x}_i, y_i\}$, for $i = 1, \ldots, l$ where $\boldsymbol{x}_i \in \boldsymbol{R}^d$ is the input pattern for the ith sample and $y_i \in \{-1, +1\}$ is the class label. Support Vector Classifiers map $\boldsymbol{x}_i$ into some new space of higher dimensionality which depends on a nonlinear function $\phi(\boldsymbol{x})$ and look for a hyperplane in that new space. The separating hyperplane is optimized by maximization of the margin. Therefore, SVM can be solved as the following quadratic programming problem,

$$\max_{\alpha_i} \{\sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum\sum \alpha_i \alpha_j y_i y_j K(\boldsymbol{x}_i, \boldsymbol{x}_j)\} \quad (1)$$

Subject to $0 \le \alpha_i \le C$ and $\sum_{i=1}^{l} \alpha_i \boldsymbol{y}_i = 0$

where C is parameter to be chosen by user, a larger C corresponding to assigning a higher penalty to errors, and $\alpha \ge 0$ are Lagrange multipliers. When the optimization problem has solved, system provides many $\alpha_i > 0$ which are the required Support vector.

Note that kernel function $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi^T(\boldsymbol{x}_i)\phi(\boldsymbol{x}_j)$ where $\phi(\cdot)$ is a non linear operator mapping input vector $\boldsymbol{x} \in \boldsymbol{R}^d$ to a higher dimensional space. In this work, we choose the radius basis function (RBF) as the kernel function since it provides the best performance compared to other well-known kernels to our application.

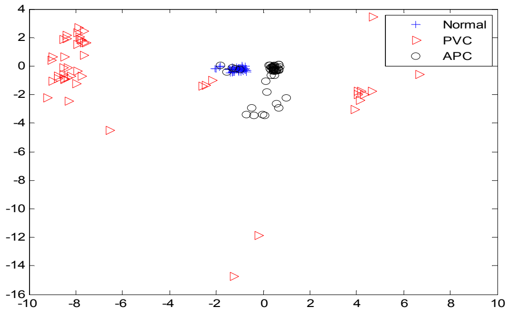

Fig. 3: Distribution of the 2-dimensional features.

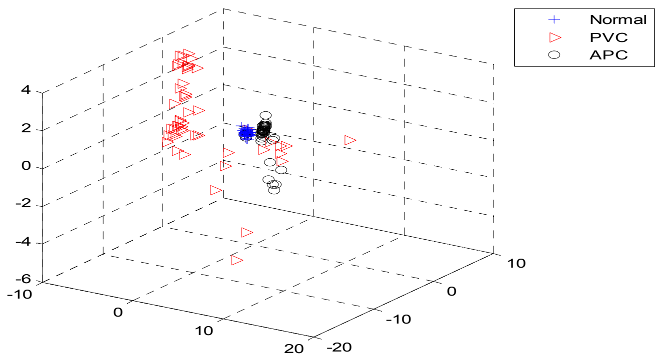

Fig. 4: The 3-dimensional distribution

3. Results

Classification consists of two steps: training and validation. In the training phase, SVM receives features as input. In this investigation, features extracted from heartbeat are represented by 2, 3 and 4 dimensional row vectors. The system under investigation is configured as follows. The dataset is categorized into two-class dataset as NORMAL beats and non-NORMAL beats where training and validation set are randomly selected as described above. The classification is performed by using 2, 3 and 4 dimensional row vectors as features input to the SVM. The similar procedure is applied to dataset so that there are the classifications of PVC beats versus non-PVC beats and APC beats versus non-APC beats. Fig. 5 shows the SVM classification of PVC against non-PVC classes where the input feature is two dimensional. Fig.6 shows the classification performance of the system under investigation by the SVM classifier in boxes which have lines at the lower quartile, median, and upper quartile values.

As inspected from the figure, the validation result from the classification of PVC versus non-PVC beats with 2-dimensional features provides the best accuracy. This fact leads us to set up further experiment where the first step is to classify with two-class method of PVC versus non-PVC beats with 2-dimensional features, then the second step is to classify the group of non-PVC into normal and APC beats. The second step uses the 3-dimensional features as input to the SVM. Fig. 7 shows the comparison of validation results of SVM using 2 and 3 dimensional features in the PVC against non-PVC classification step. The figure shows that their performances are comparable.

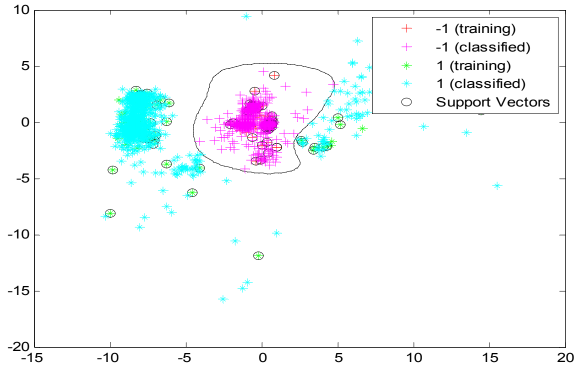

Fig. 5: SVM classification of the 2-dimensional PVC versus non-PVC classes.
PVC and non-PVC classes respectively labeled as 1 and -1.

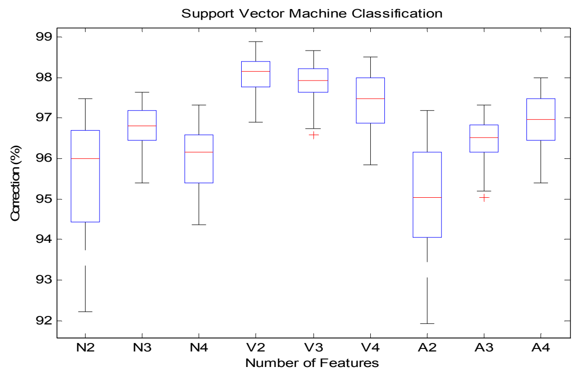

Fig. 6: Validation of two-class classification scores in different feature dimension selection:
normal (N), premature ventricular contraction (V) and arterial premature contraction (A).

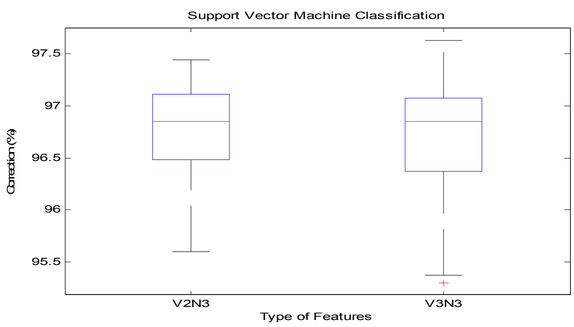

Fig. 7: SVM classification scores: V2N3 refers to 2-dimensional classification in first stage and then 3-dimensional in second stage, V3N3 refers to 3-dimensional classification and then followed by 3-dimensional.

4. Discussion and Conclusion

In this paper, the feature dimension has been used in the support vector machine (SVM). By using the two-class classification, the feature dimension can be freely selected in different levels of classification. The result shows that the classification performance as the feature dimension reduced is comparable to one that uses constant feature dimension to all levels. This can also be considered as the reduction of the computation complexity. The observation can further be applied to other implementations of multi-level SVM classification.

5. References

[1] P. Chazal, and R. B. Reilly, “Automatic classification of ECG betas using waveform shape and heart beat interal features”, *International Conference on Acoustics, Speech and Signal Processing (ICASSP’03)*, vol.2, pp. 269-272, 2003.

[2] O. T. Inan, L. Giovangrandi, and G. T. A. Kovacs, “Robust Neural-Network Based Classification of Premature Ventricular Contractions Using Wavelet Transform and Timing Interval Features, *IEEE Transactions on Biomedical Engineering*, Dec. 2006, Vol. 53, Part 1, pp. 2507-2515.

[3] M. H. Kadbi, J. Hashemi, H. R. Mohseni, A. Maghsoudi, "Classification of ECG Arrhythmias Based on Statistical and Time-Frequency Features", *Advances in Medical, Signal and Information Processing, 2006. MEDSIP 2006. IET 3rd Intr. Conf.*, July 2006, pp. 1-4.

[4] Q. Zhao, and L. Zhang, "ECG Feature Extraction and Classification Using Wavelet," *International Conference on Neural Networks and Brain, 2005, (ICNN&B '05)*, Vol. 2, pp. 1089- 1092, 2005.

[5] D. Ge, N. Srinivasan, and S. M. Krishnan, “Cardiac Arrhythmia Classification Using Autoregressive Modeling”, *BioMedical Engineering OnLine 2002,* http://www.biomedical-engineering-online.com /content/1/1/5

[6] Y. H. Hu, S. Palreddy, and W. J. Tompkins, “A Patient-Adaptable ECG Beat Classifier Using a Mixture of Experts Approach”, *IEEE Transactions on Biomedical Engineering*, vol. 44, pp. 891-900, 1997.

[7] I. Atsushi, M. Hwa, A. Hassankhani, T. Liu, and S. M. Narayan, “Abnormal Heart Rate Turbulence Predicts the Initiative of Ventricular Arrhythmias”, *Pacing Clinical Electrophysiology*, vol. 11, pp. 1189-97, Nov. 28, 2005.

[8] H. J. L. Marriott, N. L. Schwartz, and H. H. Bix, “Ventricular Fusion Beats”, *Circulation*, vol 26, pp. 880-884, 1962.

[9] C. Thanawattano and S. Tan-a-ram, “ECG classification using modified support vector machine”, *JICT2007*, Vientiane, Lao PDR, 2007

[10] R. Mark, and G. Moody, MIT-BIH Arrhythmia Database [Online], Available: http://ecg.mit.edu/ dbinfo.html.

[11] V. N. Vapnik, “The Nature of Statistical Learning Theory”, 2nd ed., Springer-Verlag, New York, 1999