21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Combining General Multi-class and Specific Two-class Classifiers for Improved Customized ECG Heartbeat Classification

Can Ye[1,2], B.V.K. Vijaya Kumar[1] and Miguel Tavares Coimbra[2]

[1] Department of Electrical & Computer Engineering, Carnegie Mellon University, USA

[2] Instituto de Telecomunicações, Faculdade de Ciências da Universidade do Porto, Portugal

Abstract

We present an approach for customized heartbeat classification of electrocardiogram (ECG) signals, based on the construction of one general multi-class classifier and one specific two-class classifier. The general classifier is trained on a global training dataset, containing examples of all possible classes and patterns. On the other hand, the individual-specific classifier is built using a small amount of individual data, which is a binary one-against-the-rest classifier, providing discrimination between normal and abnormal patterns from that individual. Such an individual-specific classifier can be a two-class classifier or a one-class classifier, depending on the availability of abnormal patterns in the individual training dataset. The classifications from the two classifiers are fused to obtain a final decision. The proposed approach is applied to the study of ECG heartbeat classification problem, significantly outperforming state-of-the-art methods. The proposed method can also be useful in anomaly detection of other biomedical signals.

1. Introduction

Cardiac arrhythmias refer to a large group of conditions associated with abnormal activity or behavior of the heart. An essential step toward detecting and classifying arrhythmias is the classification of heartbeats, given that heart rhythm can be determined by recognition of classes of consecutive heartbeats [1]. One major challenge for automatic ECG analysis comes from the significant variation in the morphologies of ECG waveforms between different patients and patient groups [2]. One general ECG heartbeat classifier, which is trained on and performs well on a large training database, could fail in analyzing a different subject's ECG due to such inter-individual variation.

In the recent past, there have been a few works [2]-[5] exploring in the design of patient-specific heartbeat classifiers. Overall, these works were all based on a customized general classifier approach. One general classifier is first built using a global training set. When testing a record, the general classifier is adapted using the first 5-min data (termed as the individual training period) from the given subject along with the manual labels for heartbeat classes, in accordance with ANSI/AAMI EC57:1998 standard [6]. In this work, the first 5-min individual training data from the given subject is used not only to adapt a general multi-class classifier, but also to construct a specific binary classifier. The proposed specific classifier is built to give an emphasis to the available individual evidence, aiming to provide a better discrimination between the individual normal and abnormal patterns. Probabilistic estimates from the general and specific classifiers are fused to make the final decision. Under this case study of heartbeat classification, the specific classifier turns out to be very effective in complementing the general classifier by solving the inconsistency of same class caused by inter-individual variation. The approach of combining proposed general and specific classifiers yielded significantly better classification accuracies than the published works [2]-[5].

The major contributions of this work are as follows:

- A new approach based on combining a general multi-class classifier and a specific two-class classifier, for the individual-specific heartbeat classification of ECG signals.
- The construction and adaptation of the general multi-class classifier using incremental support vector machine (SVM) method.
- The construction of the specific classifier – a two-class SVM classifier or a one-class SVM classifier, depending on the availability of abnormal patterns in individual training set.

Although proposed in the context of individual-specific heartbeat classification of ECG signals, this approach can be generalized to not only the anomaly detection of other biomedical signals, but also general pattern recognition problems, such as face recognition, given the similar nature as well as that inter-individual variation is a common challenge for state-of-the-art pattern recognition.

2. Experimental Data

The benchmark MIT-BIH Arrhythmias Database [7], is used as the data set in this work. This database consists of 48 half-hour two lead ECG records from 47 subjects, exhibiting a variety of arrhythmias. As the literature [2]-[5], the four paced records (i.e., 102, 104, 107 and 217) are excluded for test, as also suggested by ANSI/AAMI EC57:1998 standard. The remaining 44 recordings are divided into the training and testing datasets, each consisting of 22 records, as in [2]-[5]. The ANSI/AAMI 5-class scheme is utilized, re-clustering the original 16 heartbeat classes into the 5 bigger classes, namely 'N' (i.e., any heartbeat not in S, V, F or Q classes), 'S' (i.e., supra-ventricular ectopic beat), 'V' (i.e., ventricular ectopic beat), 'F' (i.e., fusion beat) and 'Q' (i.e., unknown beat). Class 'N' is deemed as the normal class; classes 'S' and 'V' are regarded as the two major ectopic (abnormal) classes.

3. Methodology

The global training dataset is utilized to derive the general classifier, i.e., a multi-class SVM classifier, based on labeled samples of all classes and patterns. In the testing phase, the general multi-class classifier is adapted using the first 5-min individual training data. Simultaneously, the same individual training data is utilized to obtain a specific normal-against-abnormal classifier. The class-wise probability estimates from the general and the specific classifier are fused to make a final decision.

3.1. Preprocessing

Basic preprocessing methodology is adopted from our previous work [8]. Baseline wander was corrected and the ECG signals were band-pass filtered to remove artifacts. The provided annotations of R-peak locations from the database were utilized to obtain heartbeat segments. The annotation of heartbeat class is used as the ground truth for evaluation of the proposed algorithm. Given the sampling rate of 360 Hz, each heartbeat segment consists of 100 samples before the R peak location and 200 samples after the R peak, as a total of 300 samples corresponding to 0.83 sec. The size is selected to contain most, if not all, of the information in one heart cycle, i.e., P, QRS, T wave etc. The obtained heartbeat segments are used in training and testing phases.

3.2. General Multi-Class Classifier

The construction of the general multi-class classifier is also adopted from our previous work [8]. Wavelet transform (WT) and independent component analysis (ICA) are separately applied to each heartbeat; the resulting coefficients are concatenated and represented in lower-dimensional subspace using principal component analysis (PCA). In addition, a set of RR interval features are derived to characterize rhythm information, including the previous RR, post RR, local RR and average RR interval features. The final representation consists of 18 morphological features, plus 4 dynamic features.

Incremental SVM was first proposed to tackle the memory constraint problem when training SVMs on large datasets [9]. The large dataset is partitioned. At each incremental step, the SVM classifier is updated by training the new batch of data and the support vectors from the previous step; only the support vectors of resulting SVM classifier are retained for updating the model at the next step. The nature of incremental SVM is that the representation of data so far is given by the set of support vectors characterizing the decision boundary (along with corresponding weights). This property makes incremental SVM suitable for training SVMs for streaming data.

In this work, incremental SVM method is employed to customize the pre-trained general classifier. The general classifier is characterized by the corresponding support vectors, providing a compact representation of the global training dataset. In customizing the general heartbeat classifier to a specific individual, the set of support vectors and a small amount of labeled data from the individual (i.e., the first 5-min) are combined to train the customized general multi-class classifier.

3.3. Specific Two-Class Classifier

The proposal for the specific two-class classifier is motivated as follows. Fig.1 presents 9 example heartbeats extracted from 9 different subjects, including 3 beats from each of the three sub-classes, namely, the normal sinus rhythm (NSR), the left branch bundle block (LBBB), the premature ventricular contraction (PVC). The NSR and LBBB are two sub-classes from the normal class 'N' and the PVC is a sub-class of the ventricular ectopic class 'V'. The existence of inter-individual variation is obvious for the beats from the same sub-class, e.g., the 2nd and the 3rd NSR beats have a bigger T wave and a P wave component respectively, compared with the 1st NSR beat. Besides, we observe the LBBB beats can have quite different appearance from the NSR beats, though belonging to the same class, i.e., class 'N'. We also see that some LBBB beats even appear close to some PVC patterns, e.g., 2nd LBBB and 2nd PVC beats, 3rd LBBB and 3rd PVC beats. Such inconsistency can cause problems for the customized general classifier, especially when individual abnormal training samples are under-present (e.g., as in records 117, 234), or when the individual patterns appear differently (e.g.,

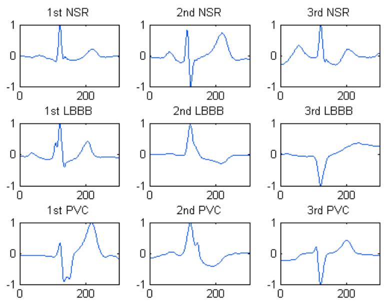

Fig. 1: The 9 beats from the NSR, LBBB, PVC sub-classes, extracted from 9 different subjects. NSR and LBBB are from the normal class 'N'; PVC belongs to the abnormal class 'V'.

as in records 105, 231) from the counterparts in the global training dataset, leading to the failure in discriminating among a few morphologically similar patterns from different classes, e.g., between LBBB and PVC.

The specific classifier is proposed to complement the general classifier by giving an emphasis to the available individual evidence. Basically, an individual normal pattern is learnt on normal samples in the individual training period. If the test beat appears 'consistently' or 'similarly' as the derived individual normal pattern, we tend to believe it as normal even if it looks close to certain abnormal patterns (e.g., PVC) in the global training dataset; otherwise, we tend to believe it as abnormal even if it looks close to certain normal patterns (e.g. LBBB). Specially, the first five normal beats from the subject are averaged as the individual normal pattern. The dynamic time warping (DTW) method is used as the metric of similarity between two heartbeat patterns. DTW is a method for measuring similarity between two sequences which may vary in time or speed. A well known application is automatic speech recognition, handling different speaking speeds. In this work, DTW is employed to alleviate dissimilarity through same beat class caused by varying heart rates. The DTW distance between the beat and the derived individual normal pattern is hereby utilized as the measure of the 'similarity'. In addition to the DTW distance feature, the pre-RR and pro-RR features are also used for representation of heartbeats in building the specific two-class classifier.

The construction of the specific two-class classifier can be generally divided into two cases. In the case that there are abundant abnormal examples available in the individual training period, a two-class SVM classifier is trained using the individual training data, with normal and abnormal samples clustered as the two classes. If there are no abnormal samples or abnormal samples are under-present, which can be very typical in real-world applications, the one-class SVM (OC-SVM) method is utilized. The OC-SVM was first proposed by Scholkopf [10], which aims to separate the dataset from the origin with maximum margin and estimate the support vectors of one 'small' region capturing most the data points, resulting in a decision function f that takes +1 corresponding to 'normal', and -1 corresponding to the outliers, i.e., 'novelty' or 'abnormal' class. OC-SVM is suitable for anomaly detection when abnormal training samples are under-present. Thus, in our case, an OC-SVM classifier is trained using the individual training data. An essential parameter of OC-SVM training is γ, which represents an upper bound on the fraction of outliers and a lower bound on the fraction of support vectors. Given the under-presence of individual abnormal samples, γ is set as 0.01. In a few cases (e.g., records 212, 231) in the testing dataset, there exist two different normal patterns (e.g., NSR and RBBB for record 231) in the individual training period. Given morphological dissimilarity between the two sub-classes, we derive two different individual normal patterns respectively and two OC-SVM classifiers are trained using samples of the corresponding normal pattern. It is worth noting that the specific classifier is constructed using only the individual training dataset.

3.4. Probabilistic Fusion

The probability estimates for the general multi-class classifier and specific two-class classifier are denoted as $P_{GE}(y=i \mid \boldsymbol{x}_{GE}),\ i=1,\cdots,N$ and $P_{SE}(z=j \mid \boldsymbol{x}_{SE}),\ j=1,2$. $\boldsymbol{x}_{GE}$ and $\boldsymbol{x}_{SE}$ represents the feature vectors fed into the general and specific classifiers, respectively; N is number of possible classes for the general classifier, i.e., $N=5$; the specific classifier has only two possible classes (i.e., normal and abnormal). The derivation of probability estimates for the general classifier and specific classifier (in the case of two-class SVM) follows Wu et. al [11].

We propose to obtain a probability estimate for OC-SVM by fitting the decision value into a logistic regression model as follows.

$$P_{SE}\left(z=1 / \boldsymbol{x}_{SE}\right) = 1/(1+\exp(b_0 + b_1 f(\boldsymbol{x}_{SE}))) \qquad (1)$$

where $f(\boldsymbol{x}_{SE})$ is the value of the decision function f, and b_0 and b_1 are intercept and regression coefficient. Following [10], the decision function f is

$$\operatorname{sgn}(\textstyle\sum_{i=1}^{l} \alpha_i K(\boldsymbol{x}_i, \boldsymbol{x}) - \rho) \qquad (2)$$

given the training samples $\boldsymbol{x}_i \in \mathbf{R}^n, i=1,\cdots,l$ without any class information.

In case of two individual normal patterns, we train two specific classifiers and probability estimates are derived respectively, as P_{SE1} and P_{SE2}. We tend to treat a test beat as normal if it is close to any of the two normal patterns. The probability estimate is given as,

$$P_{SE}(z=1) = \max\{P_{SE1}(z=1), P_{SE2}(z=1)\} \quad (3)$$

Probability estimates from the two classifiers are merged. The final probabilities of the normal class N (indexed as 1) and the four abnormal classes (indexed as 2~5) are given as

$$P_F(w=1) = P_{GE}(y=1) \cdot P_{SE}(z=1) / T \quad (4)$$

$$P_F(w=i) = P_{GE}(y=i) \cdot P_{SE}(z=2) / T, \ i = 2, \cdots, 5 \quad (5)$$

where T is the normalizing factor. The winning class corresponds to the highest probability. Probability estimates of two classifiers are fused by giving an emphasis to the available individual evidence. The specific binary classifier enhances the discrimination capability between the individual normal pattern and abnormal patterns, serving the primary concern of anomaly detection problems, i.e., the discrimination of abnormal patterns from normal ones.

4. Results

As discussed, the specific binary classifier is built only using the individual training data (the first 5-min) and the construction strategy depends on the availability of abnormal samples in the individual training dataset. Fig. 2 shows the histogram of the number of abnormal samples in the 'first-5min' of the 22 test records. The threshold is set as '10 beats' for determining the 'availability' of abnormal samples. Therefore, for the 10 records, in which there are more than 10 abnormal training beats, a two-class SVM classifier is constructed; for the remaining 12 records, in which the individual abnormal training samples are under-present, the OC-SVM strategy is used. For each test record, the general multi-class classifier is adapted using the individual training dataset. The resulting general and specific classifiers are applied to the individual testing data (i.e., the remaining 25-min). Probability estimates from the two classifiers are merged to make the classification decision for each heartbeat. Table I presents a comparison of the results of the proposed method with the published results on respective metrics. The method I (only the customized general classifier) yielded a better performance than the literature. The method II (combining the general and specific classifiers) outperforms not only the published results, but also that of only using the general classifier, indicating that the specific binary classifier is effective in complementing the general multi-class classifier.

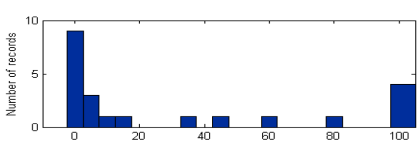

Fig. 2: The histogram of number of abnormal (arrhythmias) beats in the 'first 5-min' for 22 test ECG records.

TABLE I. PERFORMANCE COMPARISON

Method	Class 'V'				Class 'S'			
	Acc[c]	***Se***[d]	***Sp***[e]	***Ppr***[f]	***Acc***	***Se***	***Sp***	***Ppr***
Hu [2]	94.8	78.9	96.8	75.8	N/A	N/A	N/A	N/A
Chazal [3]	96.4	77.5	98.9	90.6	92.4	76.4	93.2	38.7
Jiang [4]	98.8	94.3	99.4	95.8	97.5	74.9	98.8	78.8
Tucker [5]	97.9	90.3	98.8	92.2	96.1	81.8	98.5	63.4
Method I[a]	**98.8**	**95.7**	**98.6**	**87.6**	**98.8**	**74.9**	**99.8**	**95.3**
Method II[b]	**99.3**	**96.6**	**99.5**	**92.5**	**98.9**	**79.1**	**99.8**	**94.7**

a. Method I – only the customized general classifier; b. Method II – merging the general and specific classifiers; c. Accuracy of any class = (TP+TN)/(TP+TN+FP+FN), where TP, TN, FP, FN represents the number of true positive, true negative, false positive and false negative, for the given class; d. Sensitivity; e. Specificity; f. Positive prediction rate.

5. Conclusions

This work proposes to construct a specific binary classifier to handle the inconsistency of patterns from same class caused by inter-individual variation. In this study, the approach of combining general and specific classifiers yielded an improved heartbeat classification performance. The proposed approach can possibly be extended to other pattern recognition problems to cope with inter-individual variation.

Acknowledgements

This research is under the support of Portuguese Foundation for Science and Technology through the CMU-Portugal Program under the student grant: SFRH/BD/33519/2008, the "Vital Responder" project grant: CMU-PT/CPS/0046/2008 and the "DigiScope" project grant: PTDC/EIA-CCO/100897/2008.

References

[1] J. A. Kastor, *Arrhythmias*, 2nd ed., W.B. Saunders, London, 1994.
[2] H. Yuhen et. al., "A patient-adaptable ECG beat classifier using a mixture of experts approach," *IEEE Trans. Biomed. Eng.*, vol.44, no.9, pp.891-900, Sept. 1997.
[3] P. de Chazal et. al. "A Patient-Adapting Heartbeat Classifier Using ECG Morphology and Heartbeat Interval Features," *IEEE Trans. Biomed. Eng.*, vol.53, no.12, pp.2535-2543, Dec. 2006.
[4] J. Wei et. al., "Block-Based Neural Networks for Personalized ECG Signal Classification," *IEEE Trans. Neural Networks*, vol.18, no.6, pp.1750-1761, Nov. 2007.
[5] T. Ince et al., "A Generic and Robust System for Automated Patient-Specific Classification of ECG Signals," *IEEE Trans. Biomed. Eng.*, vol.56, no.5, pp.1415-1426, May 2009.
[6] Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms, 1998. Association for Advancement of Medical Instrumentation.
[7] MIT-BIH Arrhythmias Database, available at Physionet http://www.physionet.org/physiobank/database/mitdb/
[8] C. Ye et. al, "Arrhythmia Detection and Classification using Morphological and Dynamic Features of ECG Signals," *IEEE Proc. EMBC*, vol., no., pp.1918-1921, 2010.
[9] P. Mitra et. al., "Data Condensation in Large Databases by Incremental Learning with Support Vector Machines," *IEEE Conf. ICPR*, vol.2, no., pp.708-711, 2000.
[10] B. Scholkopf et. al., "Estimating the support of a high-dimensional distribution," ACM Neural Computation, vol.13, no.7, pp.1443-1471, 2001.
[11] T.F. Wu, et. al. "Probability estimates for multi-class classification by pairwise coupling," *Journal of Machine Learning Research*, vol. 5, pp. 975-1005, 2004.