33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Customizing the Training Dataset to an Individual for Improved Heartbeat Recognition Performance in Long-Term ECG Signals

Can Ye [1,2], Johannes Pallauf [2], B.V.K. Vijaya Kumar [1], 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* — This work presents an investigation of the potential benefits of customizing the analysis of long-term ECG signals, collected from individuals using wearable sensors, by incorporating small amount of data from these individuals in the training set of our classifiers. The global training dataset selected was from the MIT-BIH Arrhythmias Database. This proposal is validated on long-term ECG recordings collected via wearable technology in unsupervised environments, as well on the MIT-BIH Normal Sinus Rhythm Database. Results illustrate that heartbeat classification performance could improve significantly if short periods of data (e.g., data from the first 5-minutes of every 2 hours) from the specific individual are regularly selected and incorporated into the global training dataset for training a customized classifier.**

I. Introduction

As life expectancies increase, improvements in quality and efficiency of healthcare and telemedicine, both at home and in hospital, are becoming more and more necessary. Wearable technology is widely viewed as having the potential to revolutionize the healthcare industry, by providing low-cost solutions for ubiquitous, continuous, unobtrusive personal health monitoring. Such continuous monitoring can lead to better healthcare for the elderly population, patients suffering from chronic diseases (such as cardiovascular disease and diabetes), people working in harsh environments (such as soldiers and firemen), and even for the common population [1].

Recent advances in the miniaturization of biosensors, wearable technology and microelectronics have enabled continuous ambulatory monitoring of physiological signals through the deployment of wearable sensors. These biosensors can come in different forms, including skin patches, sensors embedded into wearable devices such as clothing, wrist devices, arm bands, chest belts, shoes, etc.

One good example of such wearable devices is VitalJacket® (VJ), shown in Fig.1, from BioDevices S.A.[1]. This user-friendly designed smart textile is as light-weight, wearable and inconspicuous as a normal T-shirt. However, based on miniature biosensors, VJ is capable of continuously recording high-quality electrocardiogram (ECG) signals in different clinical and regular life scenarios, in hospitals, home or on the move. The collected data could either be stored in a SD memory card for offline analysis, or preferably in most cases, be transmitted using Bluetooth technology to mobile devices or a PC for online processing and analysis, so as to enable real-time monitoring of cardiac activity.

Figure 1: VitalJacket: real-time ECG monitor.

During the past decade, there has been significant research on heart arrhythmia detection and classification based on ECG signals [2-7]. These works explored characterization of heartbeats using a variety of features, including wavelets [2][3], autoregressive features [3][6], and waveform shape features [4][5]. In addition, a number of machine learning algorithms have been proposed for classification, such as neural networks [2], linear discriminant functions [4] decision trees [5] and support vector machines [3][6][7]. Most of the work focused on traditional scenarios where ECG signals are collected in hospitals using conventional ECG recorders and where patients are usually required to stay immobilized. Besides, beat classification performance is typically reported on the baseline MIT-BIH Arrhythmias Database. However, none of these seems to have addressed this interesting, yet unexplored question, *i.e.*, analysis of long-term ECG signals collected via wearable devices in unsupervised environments. Our contribution in this paper is an initial exploration of the potential of customizing classifiers trained using a reference database, by incorporating small amounts of data for a specific individual in the training set of the resulting classifiers. This work serves as a preliminary investigation for the ultimate research question, *i.e.*, the development of an individual and context adaptable ECG heartbeat classifier, for online analysis of long-term ECG signals collected via wearable devices, so as to realize real-time health monitoring in unsupervised environments based on bio-sensing technology.

[1] BioDevices S.A.: http://www.biodevices.pt/

II. Methodologies

The basic heartbeat feature extraction and classification methodologies were adapted from our previous work [8], which reported the best published performance so far in heartbeat classification on the MIT-BIH Arrhythmias Database [9], to the best of our knowledge. Fig. 2 provides an overview on the framework of methodology flow.

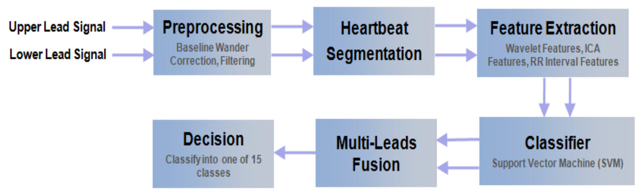

Figure 2: Methodology overview.

The MIT-BIH Arrhythmias Database includes 48 half-hour two-lead ECG recordings, sampled at 360 Hz. The raw ECG signals were first preprocessed to correct baseline wander and band-pass filtered to remove other low and high-frequency artifacts. Heartbeat segmentation follows preprocessing. Each heartbeat segment is normalized, by dividing the values by its absolute peak value. The basic assumption behind is that morphological information, such as shape of QRS complex waveform, rather than absolute physical magnitude information, is important for the decision of the heartbeat classification. This peak normalization also allows us to not worry about amplifier gains, analog-to-digital converter (ADC) settings, etc.

We are motivated by the observation that arrhythmia heartbeats are different from the normal heartbeats in both morphology and dynamics. Two categories of features are extracted for each heartbeat to represent the morphological and dynamic information of the heartbeat. Wavelet Transform (WT) and Independent Component Analysis (ICA) are applied separately to each heartbeat segment to extract corresponding coefficients, categorized as morphological features. In addition, as shown in Fig.3, four RR interval features are computed as dynamic features, to characterize the heartbeat rhythm. Principal Component Analysis (PCA) is employed to reduce the feature dimensionality. The final representation for each heartbeat consists of 26 morphological features and 4 dynamic features, *i.e.*, a total of 30 features. The features were scaled into [0, 1] before classification.

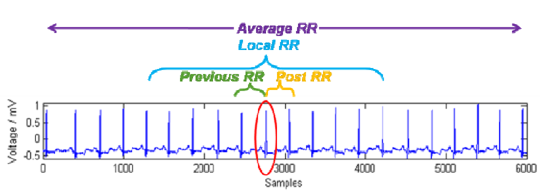

Figure 3: The previous RR, post RR, local RR and average RR features for one particular heartbeat.

Following feature extraction, a Support Vector Machine (SVM) classifier is utilized for classifying heartbeats into 16 classes, defined in the baseline database. This procedure is independently applied to signals from the upper-lead and the lower-lead, such that two independent 'decisions' are made for each heartbeat. For each single-lead decision, probabilities of being any of 16 classes are estimated and the most probable class is selected. If the two single-lead decisions give out different answers, the final decision is to select the answer that corresponds to higher probability.

The entire MIT-BIH Arrhythmias Database was utilized for the selection of the global training dataset. Following [8], approximately 13% of normal heartbeats and 50% of each of the other 15 arrhythmia classes were randomly selected, for a total 21.8% of about 110,000 beats from the database, to constitute a representative global dataset covering a variety of arrhythmias. This dataset was utilized to train a global SVM classifier. All relevant settings are trained based on the training dataset, including ICA bases, PCA bases and SVM classifier, etc. The test on the rest of the baseline database results in heartbeat classification accuracy of 99.61%, the best reported to our knowledge.

III. Dataset & Configuration

In this work, we explored the deployment of the pre-determined global settings to both long-term ECG signals collected via Vital Jacket in unsupervised environments as well as collected in a conventional way in hospital (from the MIT-BIH Normal Sinus Rhythm Database [10], *i.e.*, NSRDB). The Vital Jacket ECG data used for test is a 13-hour single-lead ECG recording sampled at 200 Hz, collected in unsupervised settings when the subject (one first-responder professional) was doing routine work. The long-term ECG data from NSRDB was used to supplement validation due to the limited availability of fully-annotated VJ ECG signals. The NSRDB recordings are two-lead ECG signals sampled at 128Hz. The records 16265, 16272 and 16483 were selected and the first 20 hours of each was utilized due to the availability of beat-by-beat annotation. Basic information of the dataset is summarized in Table 1. Although the ECG recordings were found to contain no significant arrhythmias, this preliminary investigation provides some interesting insights into deploying a classifier based on pre-determined global settings for the analysis of long-term ECGs collected in unsupervised envi-

Index	Name	Source	Duration (Hours)	Sampling Rate (Hz)	Leads Number
1	16265	NSRDB	20	128	2
2	16272	NSRDB	20	128	2
3	16483	NSRDB	20	128	2
4	VJ	VJ	13	200	1

Table 1: Information summary of the four long-term ECG recordings: three were from NSRDB and one were collected via VJ.

ronment, namely in the worrying number of false positives that traditional setups tend to produce when applied in these.

The long-term ECGs were at first re-sampled to 360 Hz and preprocessed to remove artifacts; R peaks were localized using WFDB tool [11]. Heartbeat segmentation, feature extraction and classification basically followed the procedure described in Section II. The global settings are pre-trained based on the global training dataset, including ICA bases, PCA bases, scaling vectors, etc. These settings actually characterize a certain feature space, for which the global SVM classifier has been 'optimized'. The basic idea is to project new data into the feature space and apply the corresponding optimum SVM classifier for classification.

Three testing configurations are under investigation. In Test A, the pre-determined global settings and classifier are directly applied to the individual ECG data. In Test B and C, certain small portion data from the individual long-term recording is selected as the 'individual' training dataset, which are combined with the 'global' training dataset for retraining to obtain a set of customized settings, including new PCA bases, ICA bases, scaling vectors, SVM classifier, etc. The customized settings are tested on the rest of the individual ECG data. In Test B, the first N minutes data is selected as the individual training dataset, while in Test C, the first N minutes data of every M hours is chosen to adapt the global settings. The difference between Test A, B and C lies in the training dataset of the resulting classifier. In Test B and C, the global settings as well as the classifier are customized by incorporating individual training dataset, so as the characterized feature space; the SVM classifier is re-trained and adapted accordingly.

	Training Dataset
Test A	Global
Test B	Global + Individual (First N-min)
Test C	Global + Individual (First N-min of every M hrs)

Table 2: The summary of training datasets of Test A, B, C.

IV. Results

Fig. 4 shows Test A, B and C performance for the four long-term ECG signals, including both the single-lead (*i.e.*, upper-lead) performance as well as the one after the two-lead fusion. From Fig. 4(a), we observe that classification performance improves significantly, by incorporating the first 5-minute data (*i.e.*, N=5) from the specific long-term recording into the training set of the resulting customized classifier for the subject. It seems that the global classifier trained on the reference database generalized poorly to some 'unobserved' data (record 16265 and 16272). Such underperformance could most probably be attributed to the large inter-individual variation of the morphologies of ECG waveforms. The small amount (*i.e.*, the first 5-minute) of data selected seems to have tackled such 'inter-individual difference' and customized the training dataset as well as the resulting classifier. Besides, interestingly, we notice that, when more individual data are incorporated for the training (*i.e.*, the N increased from 5 to 10, 15, etc.), the performance degraded for the three NSRDB recordings, namely, number of false positives increased. First, this observation seems to indicate that continuously adding more and more normal heartbeats of the individual to the training set does not seem to be beneficial after a certain amount of individual data is incorporated. Secondly, such degraded performance could possibly be explained by Fig. 5, supporting the existence of influence of another type of difference, *i.e.*, 'context difference'. Subject switching from one context to another is inevitable, when it comes to the scenario of long-term and ubiquitous health monitoring in unsupervised environments. In Fig.5, the upper three sub-figures present the number of false positives over time, for three different cases (*i.e.*, N=0,5,10) of record 16483 based on the first-lead performance. As we can see, when N increased from 0 to 5, the number of false positives was reduced over the whole period of the recording; while N increased to 10, the degraded performance is mainly caused by the increased number of error at the latter period of the

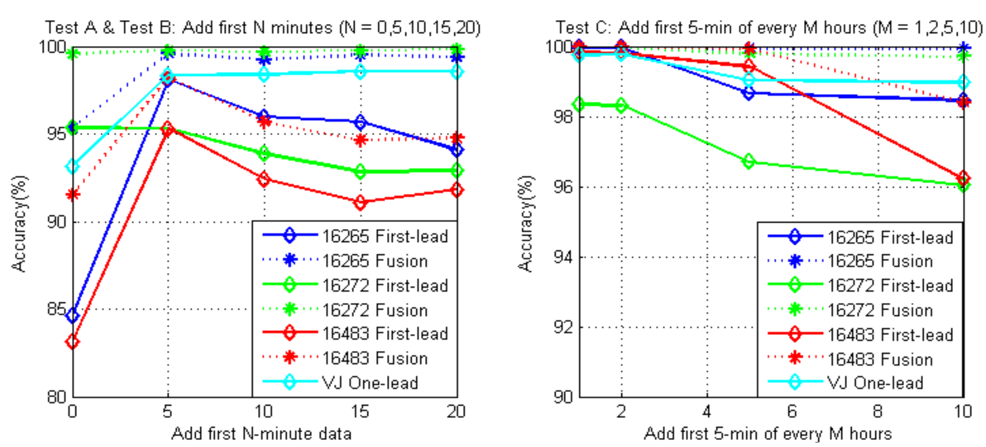

Figure 4: Test A, B & C results for the four long-term ECG signals, including the first-lead performance and the two-lead fusion performance (The VJ ECG recording contains only one-lead signal, thus no fusion was made).

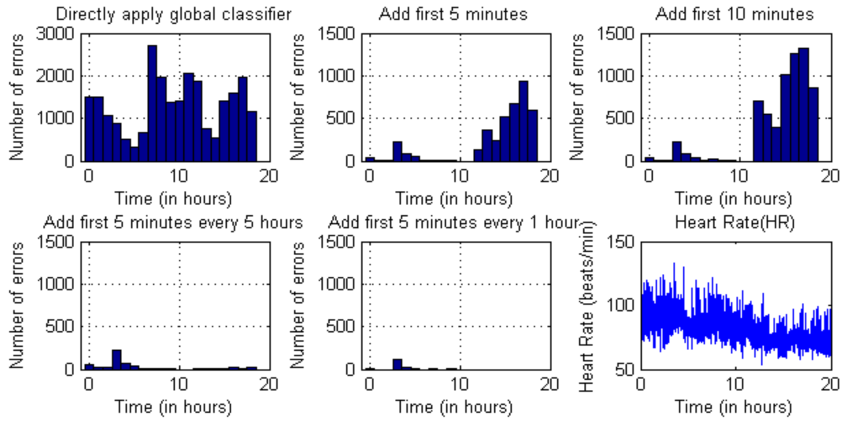

Figure 5: The analysis of error histograms the heart rate plot over time based on first-lead performance of record 16483.

record. Relating with the heart rate sub-figure, the heart rate dynamics of this period is very different from that of the beginning period. The averagely lower heart rate of the latter period could possibly be associated with a different context where the subject tends to act more 'quietly', *e.g.*, sleeping. In other words, by adding the first 10-minute period instead of the first 5-minute, the resulting classifier has actually been 'specialized' too much to a single context. Therefore, to cope with the dynamic and morphological difference brought by the 'context' difference, the classifier is regularly re-trained, by incorporating small amount of data consistently (*e.g.*, the first 5-minute of every M hour) to the training dataset of the resulting context-customized classifier. The results are presented in Fig. 4(b), from which, we find that, more frequently the classifier is customized, better it is capable to cope with the context difference, so is the resulted heartbeat classification performance, which could also been seen in the two bottom error histograms over time in Fig.5. Besides, we observe in Fig. 4, the two-lead fusion generally tends to generate better performance than using only single lead, since classification confidence has been boosted by the fusion between the two leads of ECG signals. However, when the classifier is adapted frequently 'enough' (*e.g.*, M=1 in Fig.4(b), that is, every one hour), even the single-lead performance tends to produce very good results. In real-time applications such as Vital Jacket, it might be adequate to utilize only the single-lead signal, to halve the computational cost of the algorithm while still providing sufficient accuracy.

V. Conclusion & Discussion

Given that the scenario of long-term, real-time and ubiquitous health monitoring in unsupervised environments is concerned, our study hints that the long-term ECG signal analysis requires frequent adaptation of the pre-determined global settings to tackle both the individual and context differences. This preliminary investigation shows that, regular re-training of the classifier, by consistently incorporating consistently short periods of data as individual training dataset, can significantly reduce the number of false positives. However, such re-training procedure is never meant to be the final answer, for two obvious reasons: the first is that regular re-training of a huge classifier is undesired as far as real-time application is concerned; the other is that the shift from context to context is unpredictable in unsupervised environments, thus making constant adaptation a non-ideal solution. In future work, it will be desirable to develop a self-adaptable classifier to cope with these variations.

References

[1] A. Pantelopoulos et al., "A Survey on Wearable Sensor-Based Systems for Health Monitoring and Prognosis," *IEEE Trans. Syst., Man, Cybern.,* vol.40, no.1, pp.1-12, 2010.

[2] G.K. Prasad *et al.*, "Classification of ECG arrhythmias using multi-resolution analysis and neural networks," *IEEE TENCON Conf.*, pp. 227-231, 2003.

[3] Q.B. Zhao *et al.*, "ECG Feature Extraction and Classification Using Wavelet Transform and Support Vector Machines," *Neural Networks & Brain*, vol. 2, pp.1089-1092, 2005.

[4] D.C. Philip *et al.*, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1196-1206, 2004.

[5] J. Rodriguez *et al.*, "Real-time classification of ECGs on PDA", *IEEE Trans. Info. Tech. in Biomed.*, vol. 9, no. 1, pp. 23-34, 2005.

[6] D.F. Ge *et al.*, "Cardiac arrhythmia classification using autoregressive modeling," *Biomed. Eng. Online*, vol. 1, no. 5, pp. 1-12, 2002.

[7] X. Jiang *et al.*, "ECG Arrhythmias Recognition System Based on Independent Component Analysis Feature Extraction," *IEEE TENCON Conf.*, pp. 1-4, 2006.

[8] C. Ye *et al.*, "Arrhythmia detection and classification using morphological and dynamic features of ECG signals," *IEEE EMBS Conf.*, pp. 1918-1921, 2010.

[9] The MIT-BIH Arrhythmias Database (Cited in April 2011): http://physionet.org/physiobank/database/mitdb/

[10] The MIT-BIH Normal Sinus Rhythm Database (Cited in April 2011): http://physionet.org/physiobank/database/nsrdb/

[11] The WFDB Software Package: http://www.physionet.org/physiotools/wfdb.shtml