Combining Low-dimensional Wavelet Features and Support Vector Machine for Time Series Classification

Karndeep Singh Rai-Bhatti
Kraibhat@eng.ucsd.edu

William Argus
wargus@eng.ucsd.edu

1. Project Overview

Classification of time series data is an important tool that can be used in a wide variety of applications. The goal of this project is to expand the applications of an existing classificier previously used for EKG classification, and reuse it on a broad assortment of applications. We will begin by replicating the original system consisting of wavelet analysis coupled to a support vector machine(SVM), and then more forward to implementing the system on new types of data sets. Additionally, wavelet optimization via a cost function and feedback will be explored if possible.

2. Problem Description

With the growing quantity of time-series data available, classification algorithms are necessary to quickly and accurately take advantage of this data in a meaningful way. We hypothesize that such a classification system that uses wavelet analysis to extract features and a SVM to classify a time-series can be applied to various different types of data. Appropriate wavelets will be selected for each data type, the overall system will be retrained, and the systems efficacy will be analyzed.

3. Current Approaches

Most current approaches use a unique approach to the feature extraction part of the problem and then use a neural network or SVM to classify the signal based on the feature. Fatin et al. [2] used linear and nonlinear features as input to an SVM classifier. This allowed for a performance improvement, particularly in noisy conditions over existing systems that relied solely on linear features for classification.

The approach of Martis et al.[4] was to identify and segment each event (in this case each beat) in the time series data followed by a Discrete Cosine Transform on each of the segmented beats. This allowed for the DCT coefficients to be treated as a feature vector for each segmented beat, thereby solving the feature extraction portion of the problem. This work in particular utilized a feed forward neural network, SVM, and a probabilistic neural network for the classification following feature extraction.

Our project's approach is based on the work of Zhang et al [5], which uses an 8-level multiresolution wavelet analysis using bior (6,8) wavelets on the ECG signal. This is advantageous over the other techniques mentioned due its accurate time-frequency localization. This work then performed dimension reduction via Principal Component Analysis (PCA), intended to reduce the computational cost of classification. Classification is performed to label

ECG waveforms as normal or abnormal (one of 5 types of arrhythmia). This method is shown in figure 1 below.

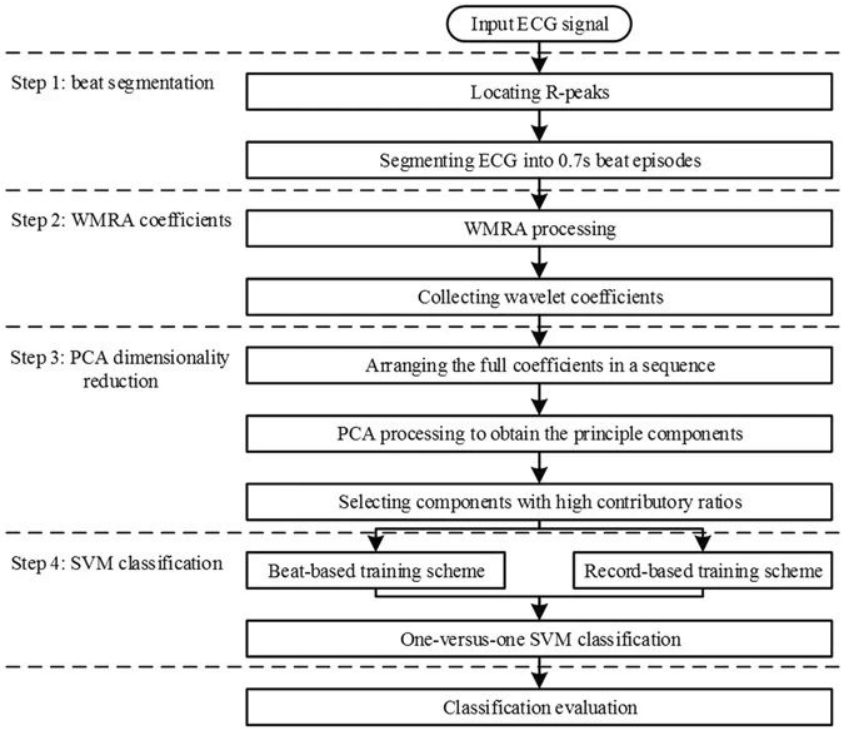

Figure 1: Block diagram of the feature extraction algorithm proposed by Zhang et al.

4. Proposed Approach

We are interested in utilizing wavelet analysis in conjunction with support vector machine classification in applications for automatic feature extraction and classification tasks. We intend on utilizing the classification scheme described by Zhang et al, and reapplying it to different datasets. Firstly, we plan to replicate Zhang et al.'s, procedure and results for classifying ECG waveforms as normal or abnormal (one of 5 types of arrhythmia). Once this initial step is successful, we would like to apply this same analysis scheme on another unrelated data set, such as datasets [1] and [6]. The key effort for this portion of the project would be to alter the wavelet being used in the analysis. Initially, we would reference literature to implement a wavelet, additionally if time permits, we would like to generate the wavelet as an optimization problem, solving for the coefficients.

5. References

[1] Alves de Souza, Vinícius. (2018). Asphalt pavement classification using smartphone accelerometer and Complexity Invariant Distance. Engineering Applications of Artificial Intelligence. 74. 198-211. 10.1016/j.engappai.2018.06.003.

[2] Fatin, A. E. *et al*. Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals. *Comput. Methods Programs Biomed.* 127, 52–63 (2016).

[3] Martis, R. J., Acharya, U. R. & Min, L. C. ECG beat classification using PCA, LDA, ICA and discrete wavelet transform. *Biomedical Signal Processing and Control* 8, 437–448 (2013).

[4] Martis, R. J. *et al*. Characterization of ECG beats from cardiac arrhythmia using discrete cosine transform in PCA framework. *Knowledge-based Systems* 45, 76–82 (2013).

[5] Qin, Q., Li, J., Zhang, L. *et al.* Combining Low-dimensional Wavelet Features and Support Vector Machine for Arrhythmia Beat Classification. *Sci Rep* **7,** 6067 (2017). https://doi.org/10.1038/s41598-017-06596-z

[6] Ye, L., Keogh, E. Time series shapelets: a novel technique that allows accurate, interpretable and fast classification. *Data Min Knowl Disc* **22,** 149–182 (2011). https://doi.org/10.1007/s10618-010-0179-5