Identification of Heart Disorders Using Symbolic Aggregate Approximation (SAX)

Motif identification in Electrocardiogram (ECG) time series is challenging due to the nature of the data as well as the computational complexity of the algorithms used to identify such patterns. This work utilizes the Symbolic Aggregate Approximation (SAX) on 1000 fragments of ECG signals from 45 patients (42% females aged between 23 and 89 years and 58% males aged 32 to 89 years). using data obtained from the MIT-BIH Arrhythmia database to recognize cardiac health disorders. Data include a Normal Sinus Rhythm, and ECG readings for 15 heart disorders, plus normal pacemaker rhythm, making 17 in total. Disorders with less than 3 patients were dropped resulting in only 12 classes being used.

The aim is to use SAX to identify heart disorders using the ECG signals, by first analyzing QRS-complexes, splitting the time series into smaller equally sized segments using the Piecewise Aggregate Approximation (PAA) approach. SAX is then used to discretize the normalized PAA series to a string of arbitrary length alphabets. The algorithm is simulated for a range of values of the parameters used in the SAX execution; alphabet size, window size and number of PAA divisions. The corresponding errors for these values are compared and the set of parameter values with the least error is chosen as optimal parameters values. Using a sliding window approach in combination with the optimal values, a list of word bags is created that is the result of SAX on our data set – this gives the words (patterns) as well as frequency counts for each class. Weights for these word bags that give an indication of the proportion of these word sequences in each heart disorder of our input time series is computed.

This will be repeated for all 11 disorders together with NSR so that we are able to determine which exact patterns are present in the ECG signals. We check for discords (subsequences with the maximum distance) in the data, thereby facilitating the identification of different heart disorders. Plots of identified discords to depict the varying positions of these discords across all disorders will be made. SAX misclassification errors are computed to examine the performance of the algorithm. Long Short Term Memory (LSTM) neural network approach was also used to classify the ECG signals and results were in line with that obtained for SAX.