

Shapelet Discovery for Atrial Fibrillation Detection

Saman Parvaneh, PhD^a, Yale Chang, PhD^a

^a*Philips Research North America, Cambridge, MA, USA*

Background: ECG-based Atrial Fibrillation (AF) detectors analyze atrial activities and/or ventricular responses. Different features from RR intervals such as entropy are used to differentiate between AF and Normal Sinus Rhythm (NSR). Shapelets are time series subsequences that are maximally representative of a class. In this paper, we aimed to use shapelet discovery to distinguish between AF and NSR. The shape of the discovered shapelet will enable the interpretation of results.

Methods: AF and NSR from PhysioNet/Computing-in-Cardiology Challenge 2017 training dataset were used in this study. RR intervals were extracted using GQRS. The stratified split was applied to create a training set (NSR:1521 and AF:239) and test set (NSR:1527 and AF:234). Shapelets were extracted by scanning RR time series in the training set and identifying statistically significant patterns (S3M algorithm). Minimum and maximum shapelet length were set to 5 and 30 RR intervals, respectively. For classification, an XGboost model was trained using the presence or absence of the top 100 shapelets. The performance of the classifier was evaluated on the test set using the area under the ROC curve (AUC) and the area under the precision recall curve (AUPRC).

Results: A total of 138,492 significant shapelets were extracted from the training set. Using the top 100 significant shapelets (Shapelet length between 5 and 29), we achieved AUC and AUPRC of 0.94 and 0.77 in discrimination between AF and NSR. Among the top 100 significant shapelets, we used all shapelets with length no greater than a certain threshold (maximum acceptable Shapelet length) for training different models. Increasing the number of shapelet features by varying the threshold from 5 to 30 in the model training improved AUC/AUPRC 0.91/0.68 to 0.94/0.77.

Conclusions: The promising performance of the trained model demonstrates that shapelet-based features have a great potential to discriminate between AF and NSR.