Multiple Representation-Based Ensembles for Time Series Classification

Abstract

Time series classification has gained more attention due to its potential in various fields, such as healthcare (e.g., diagnosing diseases based on patient vitals), industrial monitoring (e.g., detecting machine faults), and environmental studies (e.g., weather pattern recognition). Many algorithms have been specifically designed for time series classification. Those techniques can be categorized based on the fundamental data representation employed. Feature-based approaches depend on global features extracted through a straightforward pipeline and fed into an appropriate classifier. Dictionary-based methods transform real-valued time series into discrete symbol sequences, thereby exploiting the frequency of recurrent patterns. Interval-based approaches generate features from specific time segments within the series, revealing temporal characteristics of time series. Shapelet based approaches identify phase-independent subsequences to effectively discriminate between time series. The current attractive method in classifying time series is to utilize two or more representations.

This thesis introduces MuRBE (Multiple Representation-Based Ensembles), a novel heterogenous ensemble for time series classification. The MuRBE leverages diverse representation domains, including feature-based, dictionary-based, interval-based, and shapelet-based methods. Exploiting complementary information from different representations makes it particularly effective in improving classification performance. We additionally present two innovative classifiers that serve as the components within the MuRBE structure, namely feature-based autoregressive fractional integrated moving average with random forest (ARFIMA-RF) and dictionary-based symbolic aggregate approximation with stacking gated recurrent unit and convolutional neural networks (SAX SGCNN). We demonstrate that MuRBE is significantly more accurate than the base classifiers and achieves competitive performance across the current state-of-the-art methods on 40 UCR/UEA archive datasets.