Pseudo Periodic Time Series Segmentation: An Optimization Approach

Hendri Sutrisno

Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

Abstract

Time series segmentation has been one of the most discussed problems in data mining. With a proper approach, the motifs and anomalies in time series can be obtained based on the segmentation results. By scanning the whole data, existing methods consider the fixed-length segmentation based on sliding window approaches to find the motifs and anomalies in the time series; however, this process is time-consuming due to the need to scan the long time series. Also, a dynamic approach is preferable for real-world problems due to its flexibility and effectiveness according to the problem domain. In this work, we formalize the time series segmentation problem as an optimization task. Given a set of cutting points, we aim to locate the cutting points optimally so that the total distance between the segmentation results to the target motifs is minimized. Suppose that the time series has a main motif with limited anomalies; the proposed optimization task reduces the distances between the subsequences with the main motif. As a consequence, the anomalies get higher distances. Under the proposed formulation, existing optimization algorithms can be applied to solve the time series segmentation problem. Experimental results show the computational advantage of our study and the comparison of multiple optimization algorithms for solving the proposed formulation. This is a joint work with Dr. Frederick Kin Hing Phoa (Institute of Statistical Science, Academia Sinica).