Time series with STUMPY

Report and Analysis

Analyzed time series dataset contains four columns: drum pressure; excess oxygen; water level; steam flow related to the operation of a steam generator.

The collection of data consists of 9600 instances.

Report analyses the code in which motif and discords are detected through the stumpy matrix profile .

1. Data Loading and Preprocessing:

* The dataset is loaded into a Pandas DataFrame.
* The ‘steam flow’ column is extracted as the primary time series for analysis.

1. Matrix Profile Calculation:

•The ‘stumpy.stump’ function is used to compute the matrix profile with a subsequence length of 640.

* The matrix profile provides the minimum distance for each subsequence of length 640 to its nearest neighbor in the time series.
* The output matrix profile is stored in a DataFrame with columns: profile, profile index, left profile index, and right profile index.

Matrix profile is calculated with STUMPY using the function stumpy.stump. By calculating matrix profile we can can insights into the common behaviors of the data and later on to discover motifs and anomalies

1. Motif Discovery:

* The best motif is identified as the subsequence with the smallest matrix profile value which represents recurring pattern in the steam flow, indicating stable operational cycles in the steam generator.
* The location of this motif is highlighted on the time series plot.

The best motif is the one where the profile is the smallest, by running the code :

best\_motif = matrix\_profile\_df[matrix\_profile\_df['profile'] == matrix\_profile\_df['profile'].min()]

best\_motif

A screenshot of a black screen

Description automatically generated

output is:

A screenshot of a graph

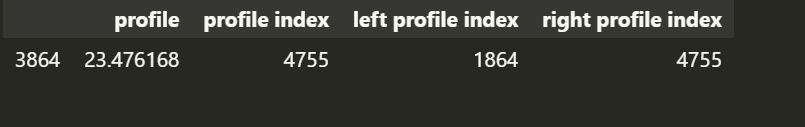
Description automatically generatedThe subsequences starting at indices 643 and 8724 are identified as the most similar, forming the best motif in the steam flow time series.

1. Discord Discovery:

* The discord is identified as the subsequence with the largest matrix profile value which indicates subsequence that is most dissimilar to all other subsequences, highlighting potential anomalies.
* The location of the discord is highlighted on the time series plot.

discord = matrix\_profile\_df[matrix\_profile\_df['profile'] == matrix\_profile\_df['profile'].max()]

discord

output: 

A screenshot of a graph

Description automatically generated