

Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View that Includes Motifs, Discords and Shapelets

著者:hin-Chia Michael Yeh, Yan Zhu, Liudmila Ulanova, Nurjahan Begum, Yifei Ding,
Hoang Anh Dau, Diego Furtado Silva, Abdullah Mueen, and Eamonn Keogh

会議名, 年: IEEE 16th International Conference on Data Mining (ICDM), 2016

1 概要

1.1 どんな論文?

1 本の時系列から部分時系列とその最近傍の部分時系列との距離・インデックスを格納する Matrix Profile を提案

1.2 Matrix Profile とは

長さ n の時系列 T から部分時系列の長さ m を決め, 任意の部分時系列とその最近傍の部分時系列の距離とインデックスを格納する。

インデックス i で始まる長さ m の時系列を $t_{i,m}$ と表すと,

$$MP_i = \min_j ED(\text{zscore}(T_{i,m}), \text{zscore}(T_{j,m})),$$

$$MPIndex_i = \arg \min_j ED(\text{zscore}(T_{i,m}), \text{zscore}(T_{j,m}))$$

となる ($\text{zscore}(t)$ は, t を正規化することを指す)。

- Matrix Profile の例

matrix profile の値が低いところ → 類似した部分時系列が存在する

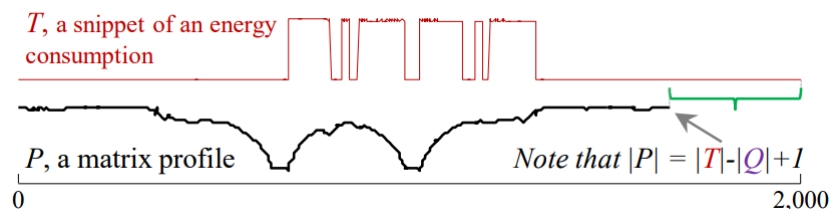


図1 energy consumption データの matrix profile(上:時系列データ, 下:MatrixProfile)

反対に matrix profile の値が高いところは他に類似した部分時系列がないので、異常検知などにも応用できる可能性がある

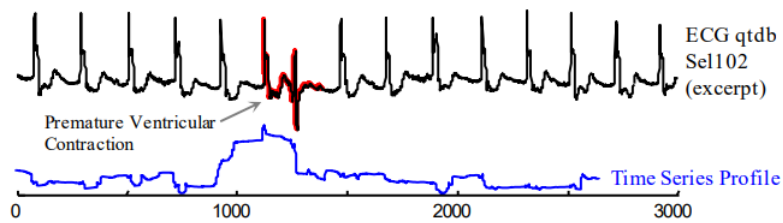


図2 ECG データの matrix profile(上:時系列データ, 下:MatrixProfile)

1.3 似ている手法

- k 近傍法

1.4 メリット

- 視覚的にモチーフになり得る箇所が分かる
- 愚直に計算すると $O(mn^2)$ だが、高速化により $O(n^2)$ に削減できる

2 感想・メモ

- モチーフとなり得る箇所などが視覚的に理解できるという点で、時系列マイニングに非常に有効だと感じた
- 高速化をしても計算オーダーが $O(n^2)$ なので、長すぎる時系列に適用するにはデータを削減するなどして対応する必要がある
- 高速化の部分がまだ理解していないので勉強中