

Matrix Profile XII: MPdist

(An Ultra-Fast Time Series Distance Measure to allow Data Mining in more Complex Real-World Deployments)

著者: Shaghayegh Gharghabi, Shima Imani, Anthony Bagnall²,
Amirali Darvishzadeh, Eamonn Keogh

会議名, 年: IEEE International Conference on Data Mining, 2018, pp. 17–20.

1 概要

1.1 どんな論文?

2つの時系列に似たような部分時系列が含まれていれば距離が小さくなるような、新しい距離尺度 (MPdist) を提案 (計算には Matrix Profile を用いる)。

1.2 比較対象

- ユークリッド距離
- DTW

1.3 メリット

- 時系列の長さが違っていても測定可能
- ベースラインの異なる時系列や spike 等に対してロバスト

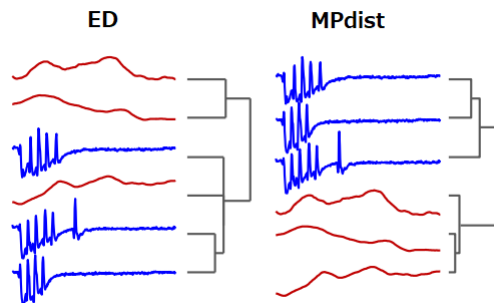


図1 ユークリッド距離と MPdist を用いて階層的クラスタリング (最長距離法) を行った結果

1.4 MPdist の計算

1. 時系列 A と B の Matrix Profile である MP_{AB} と MP_{BA} を計算
2. MP_{AB} と MP_{BA} のうち, k 番目に小さい値が MPdist となる (論文中では $k = 2 \times \text{len}(A) \times \text{len}(B) \times 0.05$ を使っている)

1.5 MPdist ベクトル

長さ n の時系列 Q と, それよりも長い長さ m の時系列 T が与えられたとき, Q と $T_{i,n}$ の MPdist を $i = 1 \sim m - n - 1$ の範囲全てで求めたものを MPdist ベクトルと呼ぶ ($T_{i,n}$ はインデックス i で始まる長さ n の部分時系列). 式で書くと以下で表せる.

$$MPdistvec_i = MPdist(Q, T_{i,n})$$

普通に計算 (matrix profile は高速化していると仮定) すると $O(n^2m)$ だが, 高速化により $O(nm)$ で計算できる.

1.6 感想・メモ

- ユークリッド距離や DTW 距離みたいに時系列全体を使って距離を計算すると, ノイズや振幅方向の歪みに敏感になってしまうので, 部分時系列を比較することは有意義と感じた.
- 長さ n と m の時系列の MPdist の計算量は $O(nm)$ なので, 長い時系列に適用するには工夫が必要.
- MPdist ベクトルは Snippets を求める際に用いられる.