Signal Based Methods

The Matrix Profile

Riccardo M.G. Ferrari

May 8th 2025

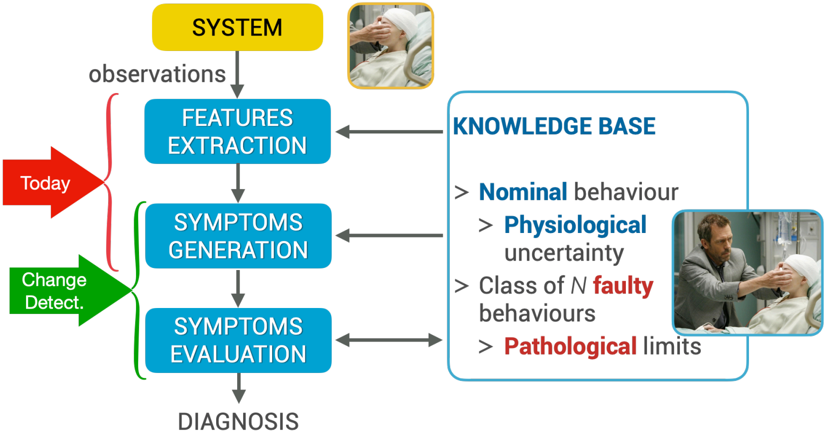
TU Delft

Lecture Summary

Introduction

Introduction

Remember the parallel with medical diagnosis?



INTRODUCTION

THE MATRIX PROFILE: LOOKING FOR SELF SIMILARITIES IN TIME SERIES

Scene taken from the movie “The Matrix” (2003). Photo: Courtesy of Warner Bros

Introduction

The Matrix Profile: looking for self similarities in time series

Intuition: conservation is key

- If a pattern is conserved, there must be some mechanism that conserve it
- Question: what is conserved in a time series?
- Conservation $\Rightarrow$ **healthy**

Examples

Bengali:	bābā	*Norwegian:*	papa
Mandarin:	baba	*Spanish:*	papá
Indonesian:	baba	*English:*	papa
Turkish:	baba	*Hindi:*	papa
Polish:	tata	*Xhosa:*	-tata

https://www.cs.ucr.edu/~eamonn/MatrixProfile.html

Introduction

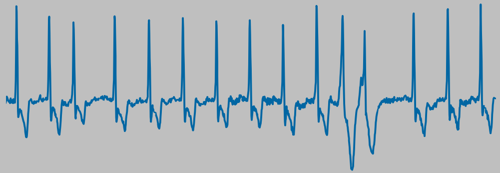
The Matrix Profile: looking for self similarities in time series

Intuition: conservation is key

- If a pattern is conserved, there must be some mechanism that conserve it
- Question: what is conserved in a time series?
- Conservation $\Rightarrow$ **healthy**

Examples

ECG signal



https://www.cs.ucr.edu/~eamonn/MatrixProfile.html

Introduction

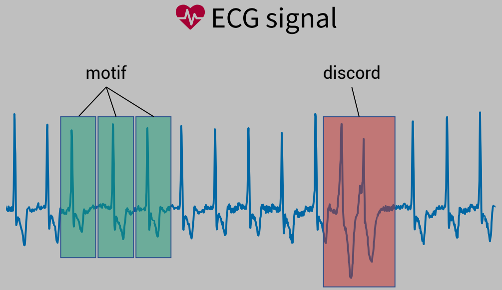
The Matrix Profile: looking for self similarities in time series

Intuition: conservation is key

- If a pattern is conserved, there must be some mechanism that conserve it
- Question: what is conserved in a time series?
- Conservation ⇒ **healthy**

https://www.cs.ucr.edu/~eamonn/MatrixProfile.html

Examples

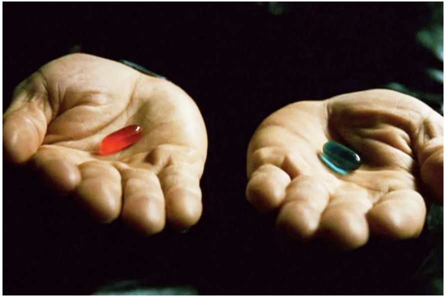


Introduction

The Matrix Profile (MP) in pills

3 easy steps

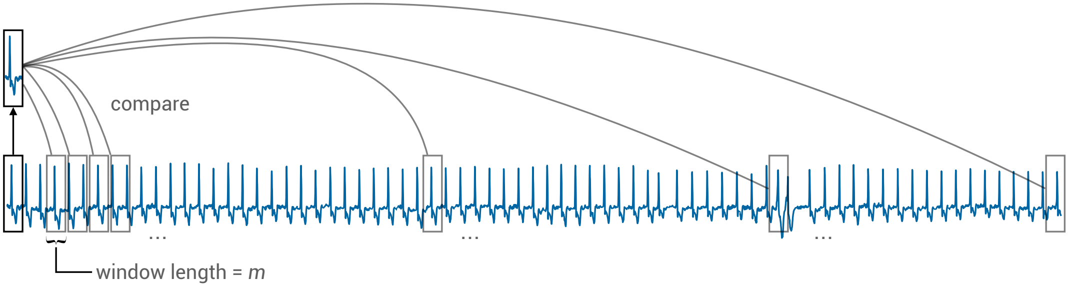
- cut up the time series in pieces
- compare each piece with every other piece
- very similar $\Rightarrow$ you found a pattern (**motif**)
- very different $\Rightarrow$ you found an anomaly (**discord**)



"Remember – all I am offering is the truth, nothing more."
Photo: Courtesy of Warner Bros."

Introduction

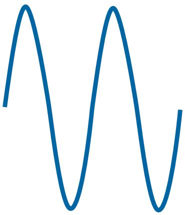
The Matrix Profile (MP) in pills



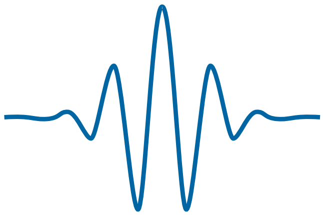
Introduction

The Matrix Profile (MP) as a generalization

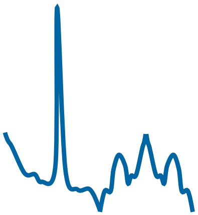
Sinusoid
(FFT)



Wavelet
(WT)



Shapelet
(MP)



Introduction

The Matrix Profile (MP): why?

PROS

- ✚ domain agnostic
- ✚ fast
- ✚ only requires a single parameter.

CONS

- − Basic assumption: repetition
- − Basic assumption: steady-state
- − Empirical choice of threshold

Definition of the MP

Definition of the MP

Distance between subsequences

- Let us consider two subsequences z_1 and z_2
  - Can be taken from same time series, or from different ones
- The **comparison** step in the MP is based on computing their distance

Definition (z-normalized Euclidean distance)

The **z-normalized Euclidean distance** of $z_1, z_2 \in \mathbb{R}^m$ is

$$d(z_1, z_2) \triangleq \sqrt{\sum_{i=1}^{m} (\bar{z}_1(i) - \bar{z}_2(i))^2} = \|\bar{z}_1 - \bar{z}_2\|_2$$

where the z-normalized values $\bar{z}_1$ and $\bar{z}_2$ are defined as

$$\bar{z}_1(i) \triangleq \frac{z_1(i) - \mu_1}{\sigma_1}, \quad \bar{z}_2(i) \triangleq \frac{z_2(i) - \mu_2}{\sigma_2}$$

and μ and σ denote, respectively, means and std. dev.

Definition of the MP

Distance between subsequences

It is all Pearson's

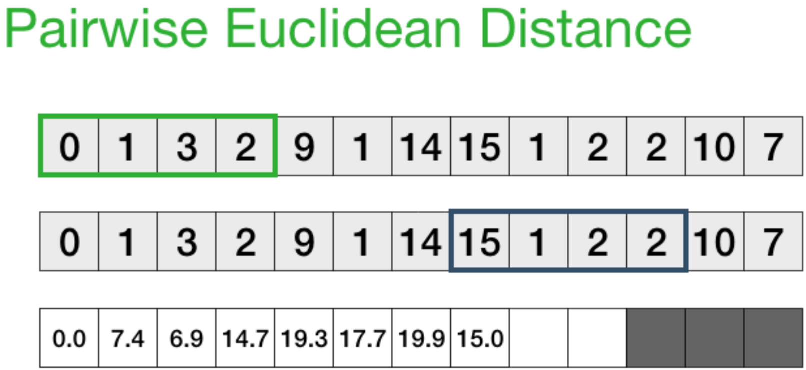
- It holds $d(z_1, z_2) = \sqrt{2m(1 - \mathbf{corr}(z_1, z_2))}$
- where **corr** is the **Pearson's Correlation Coefficient**

$$\mathbf{corr}(z_1, z_2) \triangleq \frac{\sum_{i=1}^{m} z_1(i) z_2(i) - m\hat{\mu}_1\hat{\mu}_2}{(m-1)\hat{\sigma}_1\hat{\sigma}_2}$$

- with $\hat{\mu}$ and $\hat{\sigma}$ being, as usual, sample means and std. devs.

Definition of the MP

Pairwise distance computation – example

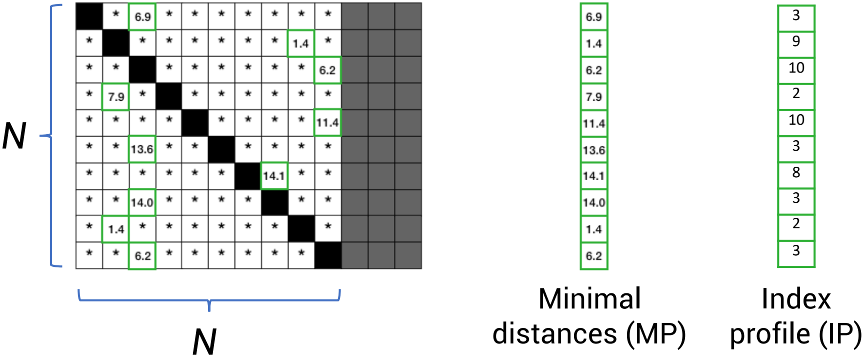


https://stumpy.readthedocs.io/en/latest/Tutorial_The_Matrix_Profile.html

Definition of the MP

Distance matrix computation – example

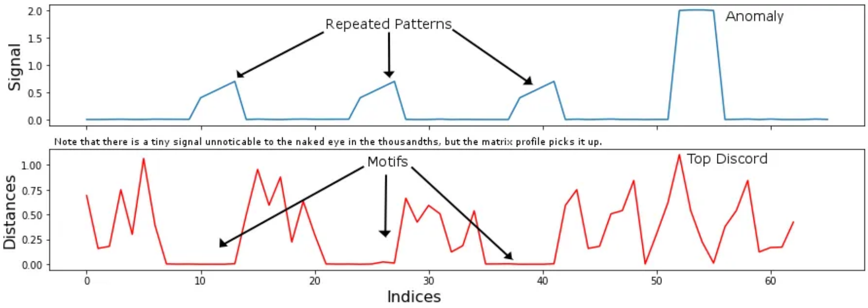
Matrix Profile



https://stumpy.readthedocs.io/en/latest/Tutorial_The_Matrix_Profile.html

Definition of the MP

Matrix Profile – example



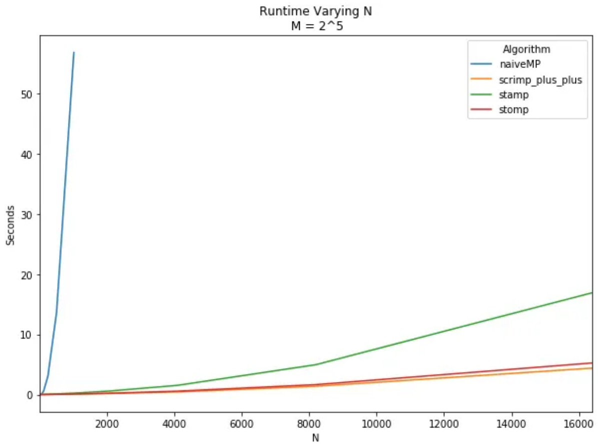
---

https://towardsdatascience.com/introduction-to-matrix-profiles-5568f3375d90

Definition of the MP

Matrix Profile – computational complexity

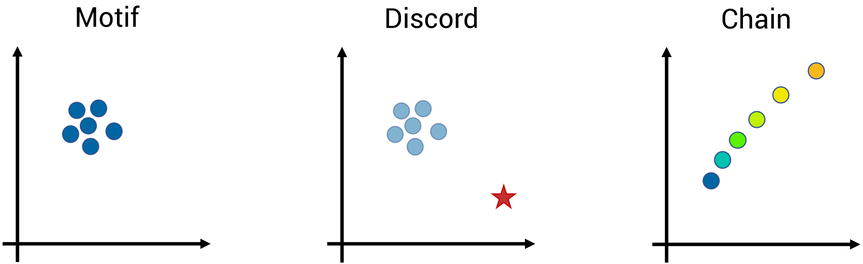
- Vanilla computation is $\mathcal{O}(n^2)$
- Efficient algorithms exist with $\mathcal{O}(n \log n)$ (similar to FFT)
- i.e. SCRIMP, STAMP, STOMP and GPU versions of that



https://towardsdatascience.com/introduction-to-matrix-profiles-5568f3375d90

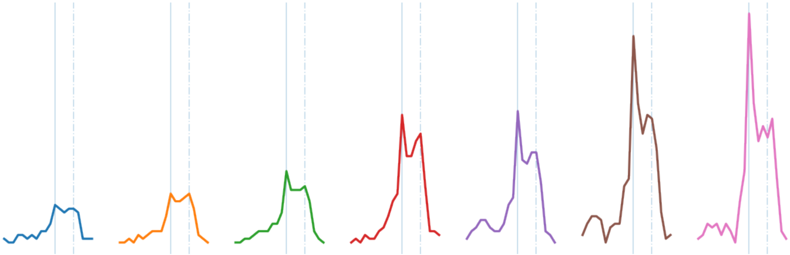
Definition of the MP

Motifs, Discords and Chains



Definition of the MP

Chain – example



https://stumpy.readthedocs.io/en/latest/Tutorial_Time_Series_Chains.html

Conclusions

Conclusions

In this lecture we covered

- **Introduction to MP**
- MP can compute very fast the self similarity of a signal to itself or to a nominal one
- Assuming your process is in/moving across a finite number of steady states, the MP can detect anomalies

Next lecture: **AI & ML for signal-based detection**

Conclusions

Further reading

- STUMPY, a Python toolbox for MP
- MP page at UC Riverside
- Chin-Chia Michael Yeh et al. "Matrix Profile I: all pairs similarity joins for time series: a unifying view that includes motifs, discords and shapelets". In: *2016 IEEE 16th international conference on data mining (ICDM)*. Ieee. 2016, pp. 1317–1322
- . . .
- Sadaf Tafazoli and Eamonn Keogh. "Matrix Profile XXVIII: Discovering Multi-Dimensional Time Series Anomalies with K of N Anomaly Detection". In: *Proceedings of the 2023 SIAM International Conference on Data Mining (SDM)*. SIAM. 2023, pp. 685–693

Conclusions

Thank you for your attention!

For further information:

Course page on **Brightspace**

or

our **MS Team**