

Data Analytics and Management in Data Intensive Domains (DAMDID) 2024 23-25 October 2024, HSE University, Nizhny Novgorod, Russia



# Time reveals all things: Leveraging behavioral patterns for anomaly detection and event prediction in time series

Andrey Goglachev, Yana Kraeva, Alexey Yurtin, and Mikhail Zymbler

{goglachevai, kraevaya, iurtinaa, mzym}@susu.ru

South Ural State University, Chelyabinsk, Russia

This work is financially supported by the Russian Science Foundation (grant No. 23-21-00465)

## We measure everything over time to predict the course of events



Smart manufacturing, Predictive maintenance



Internet of Things



**Prediction** of natural disasters



Weather forecasting, Climate modelling



Agriculture and farming



Personal healthcare



Chemoand bioinformatics







High-performance subsequence anomaly discovery in long time series

13.06.2025

2/37

# Behavioral patterns are the key

#### Preprocessing

Take a representative fragment of time series to be processed





# $\bigcirc$

#### **Anomaly/event prediction**

Determine a behavioral pattern to which the subsequence that came from a sensor, is most similar

#### Deep learning model for event prediction:

According to all the patterns, what should be the next subsequence?

#### Deep learning model for anomaly detection:

How much does the subsequence differ from all the patterns?

# Similarity measure to find patterns in time series



# Euclidean distance is for the structural similarity



Structural similarity compares time series point-by-point

# Matrix Profile distance is for the behavioral similarity



Behavioral similarity is proportional to the number of subsequences that are close w.r.t. the Euclidean distance (no matter their locations)

\* Gharghabi S. et al. An ultra-fast time series distance measure to allow data mining in more complex real-world deployments. DMKD. 2020. (34). pp. 1104-1135. DOI: 10.1007/s10618-020-00695-8

# MPdist: Step 1, matrix profile AB



Time reveals all things: Leveraging behavioral patterns for anomaly detection and event prediction in time series © 2024 M. Zymble

© 2024 M. Zymbler *et al.* October 24, 2024 **7/37** 

## MPdist: Step 2, matrix profile BA



#### MPdist: Step 3, matrix profile ABBA



Time reveals all things: Leveraging behavioral patterns for anomaly detection and event prediction in time series

© 2024 M. Zymbler *et al.* October 24, 2024 9/37

#### MPdist: Step 4, sorting and selection



10/37

# MPdist: phase-invariant, robust to spikes, warping, etc.



High-performance subsequence anomaly discovery in long time series

# Distance profile of the potential behavioral pattern



**Distances to all the subsequences w.r.t. MPdist** 



# Distance profile of the potential behavioral pattern



Distances to all the subsequences w.r.t. MPdist

Time reveals all things: Leveraging behavioral patterns for anomaly detection and event prediction in time series © 2024 M. Zymbler *et al.* October 24, 2024 13/37

# Snippets formalize behavioral patterns of any domain



# PSF (Parallel Snippet Finder) algorithm for GPU This result was reported at DAMDID'2022



Time reveals all things: Leveraging behavioral patterns for anomaly detection and event prediction in time series

© 2024 M. Zymbler *et al.* October 24, 2024 15/37

### Step forward: Can we set the snippet length without an expert?



Time reveals all things: Leveraging behavioral patterns for anomaly detection and event prediction in time series

© 2024 M. Zymbler *et al.* October 24, 2024 **16/37** 

# PaSTiLa (Parallel Snippet-based Time series Labeling)



Time reveals all things: Leveraging behavioral patterns for anomaly detection and event prediction in time series

© 2024 M. Zymbler *et al.* October 24, 2024 **17/37** 

#### PaSTiLa: Predictor



Time reveals all things: Leveraging behavioral patterns for anomaly detection and event prediction in time series

© 2024 M. Zymbler *et al.* October 24, 2024 **18/37** 

#### PaSTiLa: Selection heuristics



Time reveals all things: Leveraging behavioral patterns for anomaly detection and event prediction in time series

© 2024 M. Zymbler *et al.* October 24, 2024 **19/37** 

#### PaSTiLa: Selection heuristics



## PaSTiLa outperforms s.o.t.a. rivals



<sup>1)</sup> Ermshaus A. *et al.* ClaSP: Parameter-free time series segmentation. Data Min. Knowl. Discov. 37, 1262–1300 (2023). DOI: <u>10.1007/S10618-023-00923-X</u>

- <sup>2)</sup> Truong C. et al. Selective review of offline change point detection methods. Signal Process 167, 107299 (2020). DOI: 10.1016/J.SIGPRO.2019.107299
- <sup>3)</sup> Gharghabi S. *et al.* Domain agnostic online semantic segmentation for multi-dimensional time series. Data Min. Knowl. Discov. 33, 96–130 (2019). DOI: <u>10.1007/S10618-018-0589-3</u>
- <sup>4)</sup> Rakshitha G. *et al.* Solar Power Dataset (4 Seconds Observations). DOI: <u>10.5281/zenodo.4656027</u>.

# SANNI (Snippet & ANN-based Imputation)



Time reveals all things: Leveraging behavioral patterns for anomaly detection and event prediction in time series

© 2024 M. Zymbler *et al.* October 24, 2024 22/37

#### SANNI outperforms s.o.t.a. rivals



Time reveals all things: Leveraging behavioral patterns for anomaly detection and event prediction in time series

© 2024 M. Zymbler *et al.* October 24, 2024 23/37

# SALTO (Snippet & Autoencoder Labeling of Time series Online)



Time reveals all things: Leveraging behavioral patterns for anomaly detection and event prediction in time series

© 2024 M. Zymbler *et al.* October 24, 2024 24/37

# SALTO outperforms s.o.t.a. rivals



<sup>1)</sup>Ermshaus A. *et al.* ClaSP: Parameter-free time series segmentation. Data Min. Knowl. Discov. 37, 1262–1300 (2023). DOI: <u>10.1007/S10618-023-00923-X</u>

<sup>2)</sup> Wang Z. *et al.* Time series classification from scratch with deep neural networks: A strong baseline. IEEE IJCNN 2017. DOI: <u>10.1109/ijcnn.2017.7966039</u>
3) Dempster A. *et al.* ROCKET: Exceptionally fast and accurate time series classification using random convolutional kernels. Data Min. Knowl. Discov. 34(5) 1454–1495 (2020). DOI: 10.1007/s10618-020-00701-z

<sup>4)</sup> He K. et al. Deep residual learning for image recognition. IEEE CVPR 2016. pp. 770–778. DOI: <u>10.1109/cvpr.2016.90</u>

<sup>5)</sup> Ismail H. et al. InceptionTime: Finding AlexNet for time series classification. Data Min. Knowl. Discov. 34(6), 1936–1962 (2020). DOI: 10.1007/s10618-020-00710-y

Time reveals all things: Leveraging behavioral patterns for anomaly detection and event prediction in time series

© 2024 M. Zymbler *et al.* October 24, 2024 **25/37** 

# Discord formalizes anomaly of any domain

- *Discord*<sup>\*</sup> is the given-length subsequence whose distance to its nearest neighbor is greatest
- *Nearest neighbor* is the same-length subsequence whose distance to the given subsequence is smallest



\* Keogh E. et al. HOT SAX: Efficiently finding the most unusual time series subsequence. ICDM 2005. pp. 226-233. DOI: 10.1109/ICDM.2005.79

Distance matrix: the close neighbors, the similar they are



Distance matrix with calculated distances to neighbors

Homer	Marg	ge Bart	Selma	Patty	Barney	Quasimodo
					A CONTRACTOR	R
0	5	2	4	4	6	8
5	0	2.5	3	3	6	10
2	2.5	0	4	4	6	9
4	3	4	0	0.5	5	8
4	3	4	0.5	0	5	8
6	6	6	5	5	0	7
8	10	9	8	8	7	0

Distance matrix with **distances to their nearest neighbors** (i.e. column-wise minima)

	Homer	Marge Bart		Selma	Patty	Barney	Quasimodo	
<u></u>								
	0	5	2	4	4	6	8	
	5	0	2.5	3	3	6	10	
	2	2.5	0	4	4	6	9	
	4	3	4	0	0.5	5	8	
	4	3	4	0.5	0	5	8	
	6	6	6	5	5	0	7	
	8	10	9	8	8	7	0	

Distance matrix with the **farthest distance to the nearest neighbor** (i.e. maximum among column-wise minima)



**Discord** is an object with the **farthest nearest neighbor** 

(i.e. argument

of the maximum among column-wise minima)

	Homer	Marg	ge Bart	Selma	Patty	Barney	Quasimodo
	0	5	2	4	4	6	8
	5	0	2.5	3	3	6	10
	2	2.5	0	4	4	6	9
	4	3	4	0	0.5	5	8
	4	3	4	0.5	0	5	8
	6	6	6	5	5	0	7
R	8	10	9	8	8	7	0

# Discords grab anomalies in real time series



# PALMAD and PADDi outperform s.o.t.a. rivals This result was reported at DAMDID'2023



<sup>1)</sup> Thuy T.T.H. *et al.* A new discord definition and an efficient time series discord detection method using GPUs. ICSED 2021. pp. 63-70. DOI: <u>10.1145/3507473.3507483</u> <sup>2)</sup> Zhu B. *et al.* A GPU acceleration framework for motif and discord based pattern mining. IEEE TPDS. 2021. 32(8). 1987-2004. DOI: <u>10.1109/TPDS.2021.3055765</u>

PADDi is the only algorithm for discord discovery on HPC clusters with multi-GPU nodes



Time reveals all things: Leveraging behavioral patterns for anomaly detection and event prediction in time series

© 2024 M. Zymbler et al. October 24, 2024 33/37

# How to differ normal behavior from the opposite one



Time reveals all things: Leveraging behavioral patterns for anomaly detection and event prediction in time series

#### © 2024 M. Zymbler *et al.* October 24, 2024 **34/37**

#### DiSSiD: Discord, Snippet & Siamese net-based anomaly Detection



Time reveals all things: Leveraging behavioral patterns for anomaly detection and event prediction in time series

#### 35/37 October 24, 2024 © 2024 M. Zymbler et al.

## DiSSiD outperforms s.o.t.a. rivals

#### Anomaly detection accuracy, VUS-PR (higher is better)

Method		Time series										
		SMD	OPP	Daphnet	ECG-1	ECG-2	ECG-3	MITDB	IOPS	YAHOO	Average accuracy	Average rank
Unsupervised	IForest	0.0673 (13)	0.9731 (2)	0.3255 (5)	0.6559 (4)	0.5893 (3)	0.5519 (4)	0.2618 (6)	0.8595 (5)	0.7360 (1)	0.5578 (2)	4.78 (2)
	LOF	0.1492 (5)	0.0395 (14)	0.5055 (1)	0.3195 (12)	0.4048 (10)	0.3583 (9)	0.1864 (7)	0.4887 (11)	0.6789 (8)	0.3479 (11)	8.56 (9)
	MP	0.2762 (3)	0.0392 (15)	0.3552 (3)	0.2910 (14)	0.4582 (9)	0.2981 (12)	0.3608 (2)	0.7471 (8)	0.7353 (2)	0.3957 (8)	7.56 (7)
	DAMP	0.0879 (11)	0.0581 (13)	0.2366 (8)	0.2464 (16)	0.2699 (13)	0.2347 (14)	0.1034 (11)	0.2234 (15)	0.1583 (17)	0.1799 (17)	13.11 (15)
	NormA	0.3412 (2)	0.0361 (16)	0.2058 (12)	0.2263 (17)	0.3086 (11)	0.1470 (16)	0.1565 (9)	0.7227 (9)	0.5276 (13)	0.2969 (13)	11.67 (12)
	PCA	0.1443 (6)	0.9946 (1)	0.1219 (17)	0.6115 (5)	0.5685 (4)	0.5214 (5)	0.3013 (5)	0.9253 (1)	0.6806 (7)	0.5410 (3)	5.67 (4)
	POLY	0.1243 (8)	0.9063 (3)	0.2213 (10)	0.5582 (6)	0.5113 (7)	0.4797 (6)	0.3070 (4)	0.5567 (10)	0.6975 (6)	0.4847 (5)	6.67 (6)
	AE	0.0767 (12)	0.1979 (18)	0.2160 (11)	0.7758 (2)	0.5589 (5)	0.7651 (2)	0.0759 (15)	0.3720 (12)	0.7238 (4)	0.4181 (7)	7.89 (8)
	Bagel	0.0559 (15)	nan (17)	0.2269 (9)	0.3302 (11)	0.1878 (17)	0.2988 (11)	0.0833 (12)	0.2678 (13)	0.4871 (14)	0.2422 (14)	13.22 (16)
	DeepAnT	0.0522 (16)	0.0605 (12)	0.2573 (7)	0.3350 (10)	0.2346 (14)	0.2906 (13)	0.0795 (14)	0.1834 (16)	0.5659 (12)	0.2288 (15)	12.67 (14)
Semi-supervised	IE-CAE	0.1297 (7)	0.9002 (4)	0.3079 (6)	0.5234 (8)	0.5397 (6)	0.4739 (7)	0.1713 (8)	0.9163 (2)	0.7050 (5)	0.5186 (4)	5.89 (5)
	LSTM-AD	0.0653 (14)	0.0650 (11)	0.1711 (14)	0.2897 (15)	0.1934 (16)	0.2330 (15)	0.0799 (13)	0.1595 (17)	0.4478 (16)	0.1894 (16)	14.56 (17)
	OceanWNN	0.1075 (9)	0.4678 (7)	0.1812 (13)	0.5544 (7)	0.2003 (15)	0.3596 (8)	0.1058 (10)	0.9085 (4)	0.6126 (10)	0.3886 (9)	9.22 (10)
	OCSVM	0.0119 (17)	0.1795 (9)	0.1388 (16)	0.3548 (9)	0.3069 (12)	0.3315 (10)	0.0474 (17)	0.7533 (7)	0.6639 (9)	0.3098 (12)	11.78 (13)
	TAnoGAN	0.0965 (10)	0.8090 (5)	0.1609 (15)	0.3002 (13)	0.4634 (8)	0.1430 (17)	0.0714 (16)	0.9130 (3)	0.4591 (15)	0.3796 (10)	11.33 (11)
	DiSSiD (L1)	0.1543 (4)	0.1222 (10)	0.4124 (2)	0.7477 (3)	0.8008 (1)	0.7505 (3)	0.3718 (1)	0.2464 (14)	0.5961 (11)	0.4669 (6)	5.45 (3)
	<b>DiSSiD</b> (MPdist)	0.4889 (1)	0.5340 (6)	0.3332 (4)	0.7801 (1)	0.7927 (2)	0.8124 (1)	0.3544 (3)	0.7922 (6)	0.7306 (3)	0.6243 (1)	3.00 (1)

# Do you have time series to predict events/anomalies in?

- Parallel unsupervised algorithms, which outperform s.o.t.a. rivals
  - *Snippet discovery:* PSF, PaSTiLa (on GPU and multi-GPU clusters, respectively)
  - Discord discovery: PALMAD, PADDi (on GPU and multi-GPU clusters, respectively)
- Deep learning models, which outperform s.o.t.a. rivals
  - Prediction: SANNI, SALTO
  - Anomaly detection: DiSSiD







Yana Kraeva Cand.Sci.



Andrey Goglachev PhD student MSc



Alexey Yurtin PhD student MSc