Mini-Project (ML for Time Series) - MVA 2023/2024

Machine Learning for Time Series

Antoine Ratouchniak, Hugo Queniat

March 20, 2024

ENS Paris-Saclay

Table of contents

Introduction

Introduction

We studied **Anomaly Detection in Time Series: A Comprehensive Evaluation** by Schmidl, Wenig et al. [12].
An anomaly in a time series refers to an instant or a sequence of instants that deviate from the regular patterns of the series.

Introduction

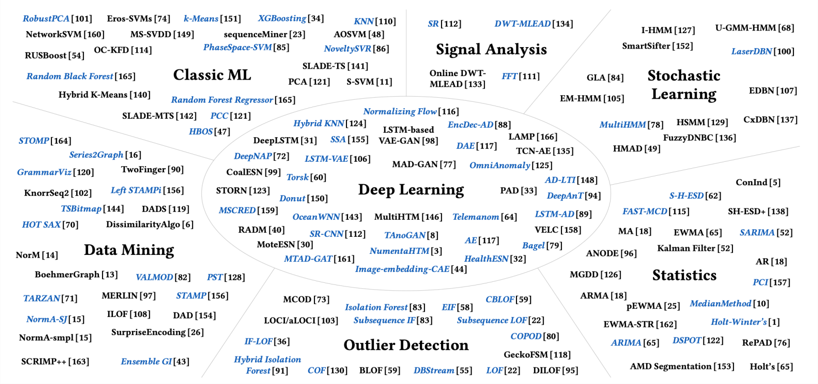


Figure 1: All 158 collected anomaly detection methods by [12] for time series data structured by their method family. Methods directly evaluated in [12] are highlighted in blue italics.

Introduction

- Chose a handful of the presented methods.
- Tested them on datasets of our choice to verify the paper's conclusions.
- Adjusted model's parameters and observed their influence on its performance.
- Tried to tweak the methods and find improvements.

Method

Method: preliminaries

Notations

We denote by $x = \{x[1], \ldots, x[N]\} \in \mathbb{R}^N$ a time series with $x[t],\ t \in \{1, \ldots, N\}$ a value at a specific timestamp

Performances evaluation

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{Recall} = \frac{TP}{TP+FN} \quad \text{Expectancy} = \frac{FP}{TN+FP}$$

AUC-ROC: area under the ROC curve where the x-axis is the Precision and the y-axis is the Expectancy

AUC-PR: area under the precision-recall curve where the x-axis is the Precision and the y-axis is the Recall

Method: Median method (1)

Median method [2]

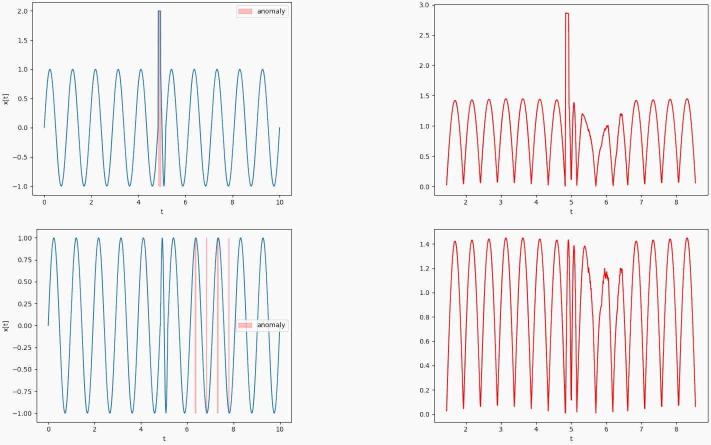
We let $\gamma[t] = \{x[t-\kappa], \dots, x[t-1], x[t+1], \dots, x[t+\kappa]\}$ for $t \in \{\kappa+1, N-\kappa\}$ a neighborhood of points of size 2κ

$$|x[t] - \mathrm{med}(\gamma[t])| > \tau\sigma(\gamma[t])$$

where $\mathrm{med}(\cdot)$ is the median, $\sigma(\cdot)$ the standard deviation and τ a threshold.

Method: Median method (2)

Example: Median method



Anomalies detected

Anomaly scores

Figure 2: Anomaly detection using the Median method

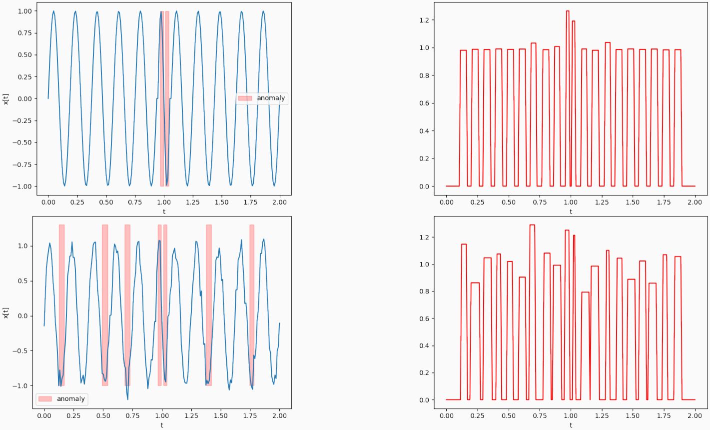
Method: FFTBSOM (1)

Fourier Transform Based Spatial Outlier Mining (FFTBSOM) [9]

1. Apply a low-pass filter on the signal, giving $\hat{x}$
2. Compute $s[t] = |\hat{x}[t] - x[t]|$ for each data point
3. If $s[t] > \tau_1$, with τ_1 a threshold, store $s[t]$ as a potential local outlier
4. If $(s[t] - \mu(s))/\sigma(s) > \tau_2$ with τ_2 another threshold, store $s[t]$ as a local outlier
5. If $sign(s[t]) \neq sign(s[t+1])$ return $s[t]$ and $s[t+1]$ as outliers

Method: FFTBSOM (2)

Example: FFTBSOM



Anomalies detected   Anomaly scores

Figure 3: Anomaly detection using FFTBSOM

Spectral Residual [10]

$$\text{Fourier Transform} : A(f) = \mathrm{Amplitude}(\mathfrak{F}(x))\ ,\ P(f) = \mathrm{Phase}(\mathfrak{F}(x))$$
$$\text{Spectral Residual} : R(f) = \log A(f) - \overline{\log A(f)}$$
$$\text{Saliency Map} : S(x) = \left\| \mathfrak{F}^{-1}(\exp(R(f) + iP(f))) \right\|$$

Scoring is computed through the local relative amplitude :

$$O(x_i) = \frac{S(x_i) - \overline{S(x_i)}}{\overline{S(x_i)}}$$

Method: Spectral Residual (2)

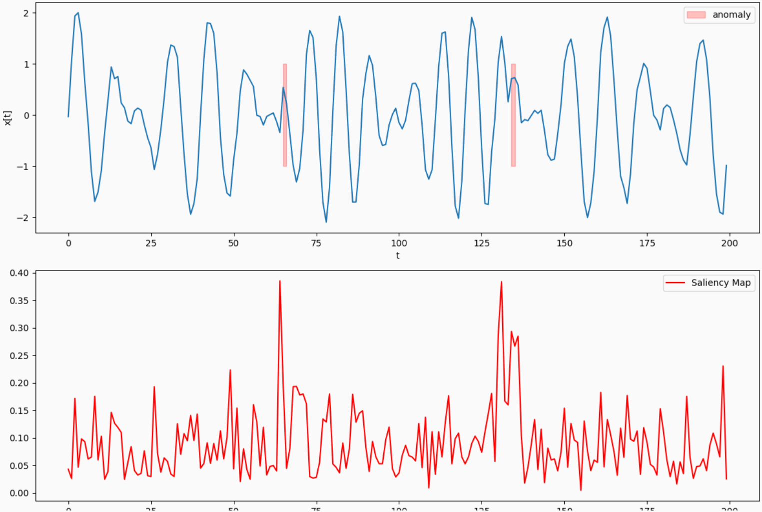


Figure 4: The Saliency Map against a signal and its anomalies

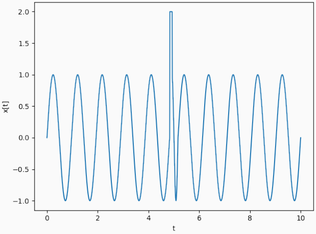
Phase Space SVM (1) [8]

Let

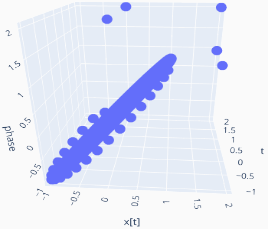
$$K : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+ \tag{1}$$

$$(x, y) \mapsto \langle \phi(x), \phi(y) \rangle \tag{2}$$

a kernel and $\phi : \mathbb{R} \to \mathcal{X}$ a feature map. We first unfold the time series using a time-delay embedding process [1]



An arbitrary time series



The time series and its vector set in the phase space

[1] Packard, N. H., Crutchfield, J. P., Farmer, J. D., Shaw, R. S. (1980). Geometry from a time series. Physical review letters, 45(9), 712.

Method: Phase Space SVM (2)

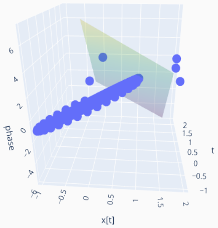
We then want to solve the one-class SVM problem

$$\begin{aligned} \underset{\mathbf{w},\rho,\xi}{\arg\min} \quad & \|\mathbf{w}\|_2^2 + \frac{1}{\nu(N-d)} \sum_{i=1}^{N-d} \xi_i - \rho \\ \text{s.t.} \quad & \langle \mathbf{w}, \phi(S_d[i]) \rangle \geq \rho - \xi_i \\ & \xi_i \geq 0,\ i = 1, \dots, N-d \end{aligned}$$

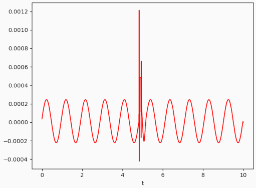
where $S_d[i]$ is the shifted time series, ν is a hyperparameter for the training error and fraction of support vectors, ξ_i are slack variables and ρ is the threshold. The dual can be shown to be [2]

$$\begin{aligned} \underset{\boldsymbol{\alpha}}{\arg\min} \quad & \frac{1}{2}\boldsymbol{\alpha}^T K \boldsymbol{\alpha} \\ \text{s.t.} \quad & \|\boldsymbol{\alpha}\|_1 = 1 \\ & 0 \leq \alpha_i \leq \frac{1}{\nu(N-d)},\ i = 1, \dots, N-d \end{aligned}$$

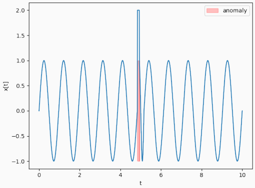
[2] Schölkopf, B., Williamson, R. C., Smola, A., Shawe-Taylor, J., Platt, J. (1999). Support vector method for novelty detection. Advances in neural information processing systems, 12.



Hyperplane of the one-class SVM (linear kernel)



Anomaly scores



Anomalies detected

Note: in practice we use a Gaussian kernel.

Method : STAMPi (1)

Scalable Time series Anytime Matrix Profile Incremental [14]
Solve the all-pair-similarity-search problem :

$$\theta_{1\mathrm{nn}} : A \times B \longrightarrow \{0, 1\}$$
$$(A[i], B[j]) \longmapsto \begin{cases} 1 & \text{if } B[j] \text{ is the closest neighbor to } A[i] \text{ in } B \\ 0 & \text{otherwise} \end{cases}$$

The Matrix Profile P_{AB} is then defined as the vector of Euclidean distances between each pair of sub-sequences in $A \bowtie_{\theta 1\mathrm{nn}} B$.
Thus, consider the scoring P_{AA} for

$$A \bowtie_{\theta_{1\mathrm{nn}}} A \text{ with } A = \{x[i : i + m] \mid 0 \leq i \leq |x| - m\}$$

Method : STAMPi (2)

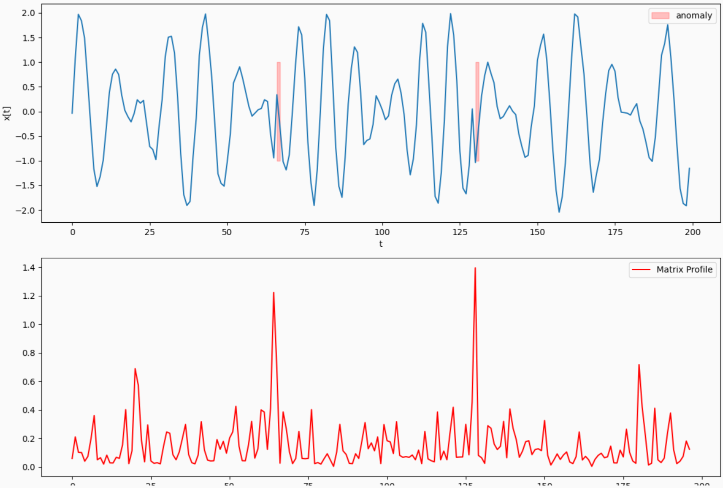


Figure 7: The Matrix Profile against a signal and its anomalies

Method: Sub-LOF (1)

Subsequence Local Outlier Factor [4]

Let $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$ a distance and z an arbitrary point

- $N_k(z) = \{p \in \mathbb{R} \mid d(z,p) \leq k - distance(z)\}$
- $\mathrm{RD}_k(z,p) = \max(k - distance(p), distance(z,p))$ where $p \in \mathbb{R}$ is an arbitrary point
- $LRD_k(z) = 1/\left(\frac{\sum_{p \in N_k(z)} RD_k(z,p)}{|N_k(z)|}\right)$
- $LOF_k(z) = \frac{\sum_{p \in N_k(z)} \frac{LRD(p)}{LRD(z)}}{|N_k(z)|}$
- z is an outlier if $LOF_k(z) > 1$

To apply Sub-LOF to time series, we split them into sequences.

Discrete Wavelet Transform MLE Anomaly Detection [13]

- Compute DWT for levels $l \in \{l_0, ..., L\}, L = \log_2(m), m = 2^{\lceil \log_2(|x|) \rceil}$
- Sliding a window of size ω to get $\mathbf{D}^{(l)} = \begin{pmatrix} d_{1,l} & \dots & d_{\omega,l} \\ \vdots & \vdots & \vdots \\ d_{2^l-\omega+1,l} & \dots & d_{2^l,l} \end{pmatrix}$
- For each $\mathbf{D}^{(l)}$ and $\mathbf{C}^{(l)}$ compute MLE estimators of (μ, Σ) for a Gaussian distribution. Associate the likelihood to each row of the matrices.

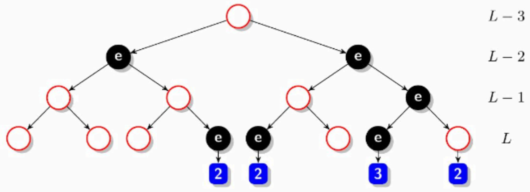


Figure 8: Leaf counters for anomalies [13]

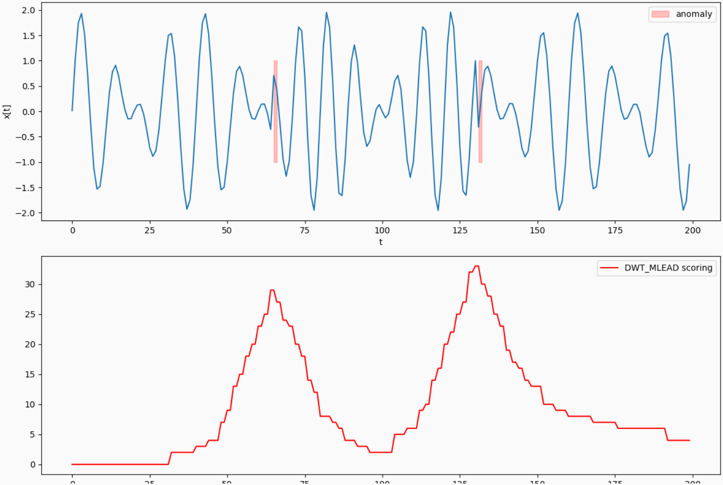


Figure 9: Example of anomaly detection with **DWT_MLEAD**

Data

Data

As the article is a survey, we use the dataset provided by the authors for accurate comparison:

- Dodges [7]: real data
- NAB [1]: real and synthetic data
- NASA-MSL, NASA-SMAP [6]: real and synthetic data
- NormA [3]: real and synthetic data

Results

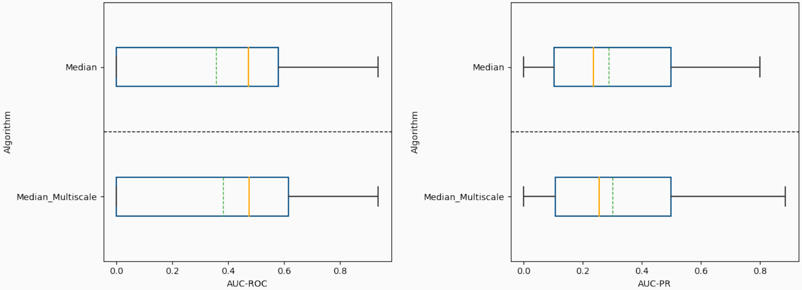
Results: Median method (1)

Instead of using only one window $\gamma_\kappa[t] = \{x[t-\kappa], \ldots, x[t-1], x[t+1], \ldots, x[t+\kappa]\}$, we try using three such that

$$\gamma[t] = \frac{\gamma_\kappa[t] + \gamma_{2\kappa}[t] + \gamma_{3\kappa}[t]}{3}$$

to have a multiscale view of the signal.

Results: Median method (2)



AUC-ROC obtained on the datasets

AUC-PR obtained on the datasets

Figure 10: Results obtained with the Median method by using a multiscale version

Results: FFTBSOM (1)

The algorithm is sensitive to noise. We can denoise the signal by using

- Moving average:

$$x[t] = \frac{x[t - k/2] + \cdots + x[t] + \cdots + x[t + k/2 - 1]}{k}$$

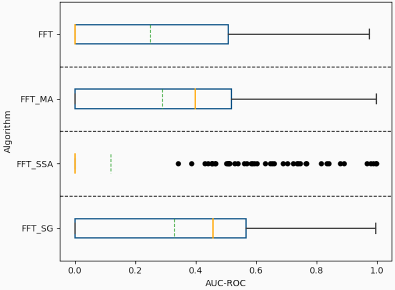
- Dictionary learning [5, 11]:

$$\arg\min_z \frac{1}{2}\|x - Dz\|_2^2 + \lambda\|z\|_1$$

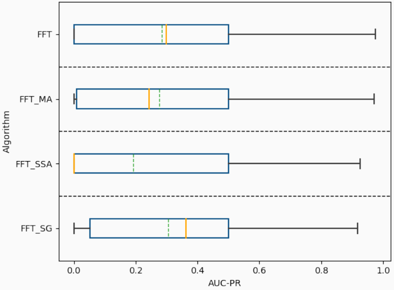
- Savitzky-Golay [3]: polynomial regression within a frame on the signal

---

[3] Savitzky, A., Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. Analytical chemistry, 36(8), 1627-1639.



AUC-ROC obtained on the datasets



AUC-PR obtained on the datasets

Figure 11: Results obtained with FFTBSOM by smoothing the data

Results: Spectral Resdiual

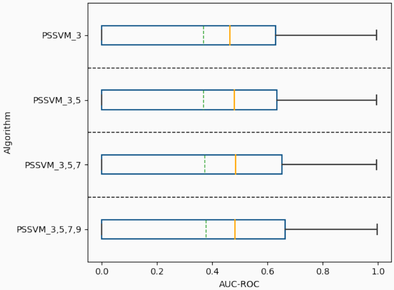
All three parameters ω, m and z are window size parameters.
Only z proved to have a significant global influence on our dataset.

Value of z	ROC AUC	PR AUC
3	0.47	0.13
23	0.50	0.17
43	0.50	0.19
63	**0.51**	**0.20**

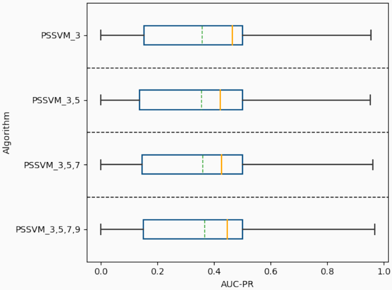
Table 1: Means obtained for different values of the score window size, z, in the Spectral Residual algorithm.

Results: Phase Space SVM

We average the scores over the phase space



AUC-ROC obtained on the datasets



AUC-PR obtained on the datasets

Figure 12: Results obtained with PS-SVM by using different time-delay embeddings

Results: STAMPi

The model exhibits a single parameter *m*, the length of the subsequences to compare.

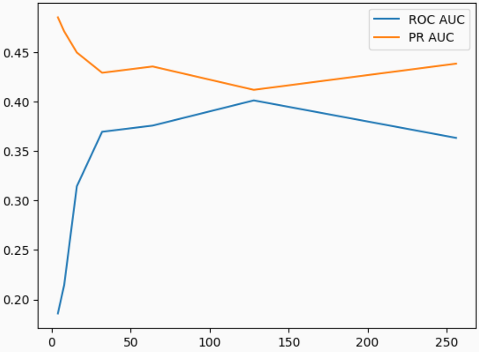


Figure 13: Means of the performance scores obtained for different values of the subsequence length, *m*, in the STAMPi algorithm.

Results: DWT_MLEAD

The method offers to fit a Gaussian distribution: What about other distributions?
Heavy tail vs Light tail :

Method / Score	ROC AUC	PR AUC
DWT	**0.78**	**0.50**
DWT-Laplace	0.76	**0.50**
DWT-T_{19}	0.77	**0.50**

Table 2: Medians obtained for the different distributions used

Results: Sub-LOF (1)

Due to the complexity ($\mathcal{O}(n^2)$) of the algorithm, the algorithm is not scalable with custom metrics on Python.
We try three different metrics:

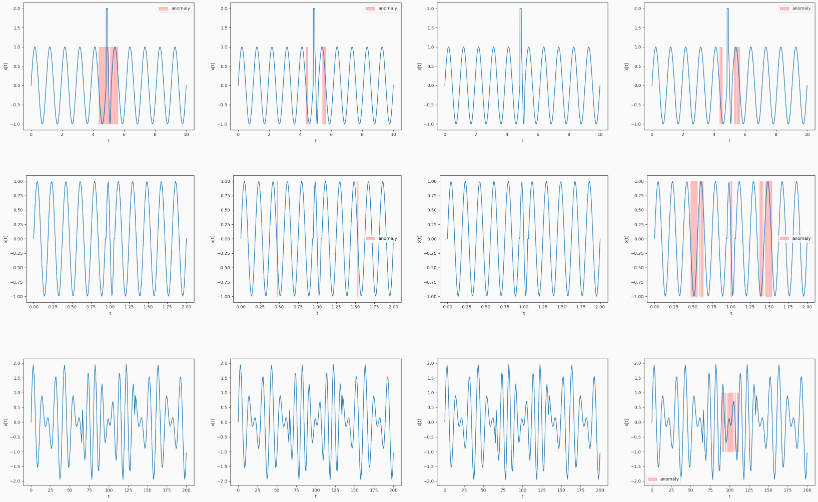
- DTW: measures the similarity between time series by finding an optimal matching by dynamic programming
- Soft-DTW: a differentiable version of the DTW
- Wasserstein-Fourier [4]

$$\mathrm{WF}_2^2(s_x, s_y) = \left(\min_{\sigma \in \mathfrak{S}_N} \sum_{i=1}^{N} (s_x[i] - s_y[\sigma(i)])^2 \right)$$

where s_x and s_y are the normalized power spectral density given by $s_z(\xi) = \frac{S_z(\xi)}{\sum S_z(\xi)}$ where z is an arbitrary signal

[4] Cazelles, E., Robert, A., Tobar, F. (2020). The Wasserstein-Fourier distance for stationary time series. IEEE Transactions on Signal Processing, 69, 709-721.

Results: Sub-LOF



Minkowski | DTW | Soft-DTW $\gamma = 1$ | WF

Figure 14: Anomaly detection on toy signals with Sub-LOF and different metrics

Results

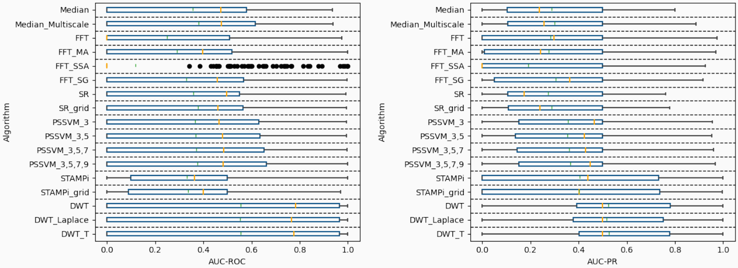


Figure 15: Boxplots performances of the algorithms and modified algorithms we used for the AUC-ROC and AUC-PR metrics. The mean value is in green and the median in orange.

References i

S. Ahmad, A. Lavin, S. Purdy, and Z. Agha.
Unsupervised real-time anomaly detection for streaming data.
Neurocomputing, 262:134–147, 2017.

S. Basu and M. Meckesheimer.
Automatic outlier detection for time series: an application to sensor data.
Knowledge and Information Systems, 11:137–154, 2007.

P. Boniol, M. Linardi, F. Roncallo, and T. Palpanas.
Automated anomaly detection in large sequences.
In *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, pages 1834–1837, 2020.

M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander.
Lof: identifying density-based local outliers.
In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pages 93–104, 2000.

S. S. Chen, D. L. Donoho, and M. A. Saunders.
Atomic decomposition by basis pursuit.
SIAM review, 43(1):129–159, 2001.

K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom.
Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding.
In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 387–395, 2018.

J. Hutchins.
Dodgers Loop Sensor.
UCI Machine Learning Repository, 2006.
DOI: https://doi.org/10.24432/C51P50.

J. Ma and S. Perkins.
Time-series novelty detection using one-class support vector machines.
In *Proceedings of the International Joint Conference on Neural Networks, 2003.*, volume 3, pages 1741–1745. IEEE, 2003.

F. Rasheed, P. Peng, R. Alhajj, and J. Rokne.
Fourier transform based spatial outlier mining.
In *Intelligent Data Engineering and Automated Learning-IDEAL 2009: 10th International Conference, Burgos, Spain, September 23-26, 2009. Proceedings 10*, pages 317–324. Springer, 2009.

H. Ren, B. Xu, Y. Wang, C. Yi, C. Huang, X. Kou, T. Xing, M. Yang, J. Tong, and Q. Zhang.
Time-series anomaly detection service at microsoft.
2019.

R. Rubinstein, A. M. Bruckstein, and M. Elad.
Dictionaries for sparse representation modeling.
Proceedings of the IEEE, 98(6):1045–1057, 2010.

S. Schmidl, P. Wenig, and T. Papenbrock.
Anomaly detection in time series: A comprehensive evaluation.
volume 15, 2022.

M. Thill, T. Bäck, and W. Konen.
Time series anomaly detection with discrete wavelet transforms and maximum likelihood estimation.
Proceedings ITISE, 2017.

C. C. M. Yeh, Y. Zhu, L. Ulanova, N. Begum, Y. Ding, H. A. Dau, D. F. Silva, A. Mueen, and E. Keogh.
Matrix profile i: All pairs similarity joins for time series: A unifying view that includes motifs, discords and shapelets.
volume 0, 2016.

References

[?]

Questions?