# Trend Prediction and Anomaly Detection on ITOps Series

Dr. Wan-Lei *Zhao*

Sep. 8 2022

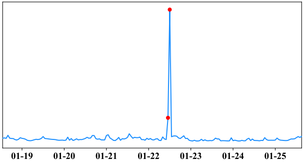
Contact: wlzhao@xmu.edu.cn

# Outline

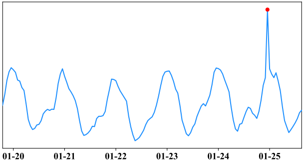
1. Anomaly Detection and Trend Prediction by LSTM

2. Online Matrix Profile for Anomaly Detection

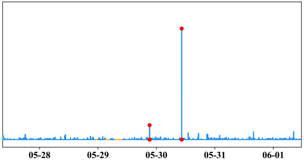
# Anomaly Detection and Trend Prediction: the problem



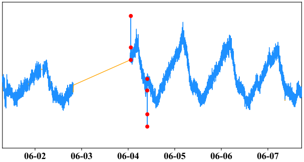

(a) hour-level & stable




(b) hour-level & periodic




(c) minute-level & stable



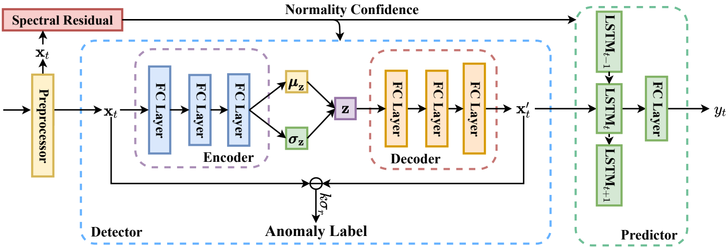

(d) minute-level & periodic

- Anomalies are marked in red
- Aims: detect the anomalies and predict the normal trend

# Existing Solutions

1. Prediction
   - ARIMA: Autoregressive Integrated Moving Average model
   - LSTM: Long-term Short Term Memory
   - VAE: Variational Auto-Encoder
   - Prophet: developped by Facebook
2. Detection
   - SPOT: based on Isolation Forest
   - SR: Spectral Residual

# Framework of our Solution




- The VAE block shown inside the blue box in charge of detection
- SR is integrated to associate a normality score for each timestamp
- Then re-encoded signal by VAE is fed into LSTM for robust prediction

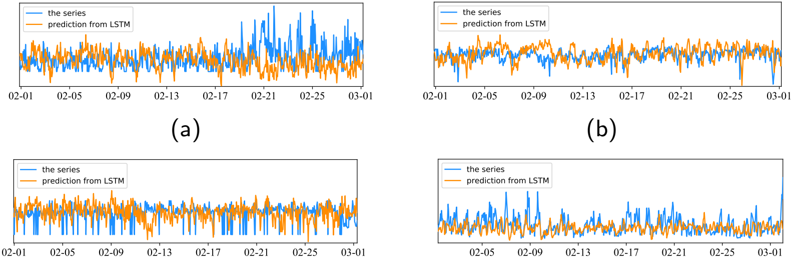
# Performance Evaluation: the dataset (1)

Table: Summary over the datasets

| Dataset | # Series | # Time-stamps | # Anomalies | Gran. |
|---|---|---|---|---|
| **KPI** | 29 | 5,922,913 | 134,114 (2.26%) | Minute |
| **Yahoo** | 367 | 572,966 | 3,896 (0.68%) | Hour |

- **KPI** is built by "AIOps challenge Competition"
- **Yahoo** is built by Yahoo, data both from real scenarios and synthesized

# Performance Evaluation: the dataset (2)




(a) (b) (c) (d)

- Four sample series from **Yahoo**
- Along with the prediction from original LSTM
- LSTM fails on irregular patterns

# Performance Evaluation: the measures

- Measures for Prediction task

$$\mathrm{MSE} = \frac{1}{n-\Omega}\sum_{t=\Omega}^{n-1}(x_{t+1} - y_t)^2 \tag{1}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n-\Omega}\sum_{t=\Omega}^{n-1}(x_{t+1} - y_t)^2} \tag{2}$$

$$\mathrm{MAE} = \frac{1}{n-\Omega}\sum_{t=\Omega}^{n-1}|x_{t+1} - y_t|, \tag{3}$$

- Measures for Detection task

$$\text{precision} = \frac{\#\text{True positive}}{\#\text{True positive} + \#\text{False positive}} \tag{4}$$

$$\text{recall} = \frac{\#\text{True positive}}{\#\text{True positive} + \#\text{False negative}} \tag{5}$$

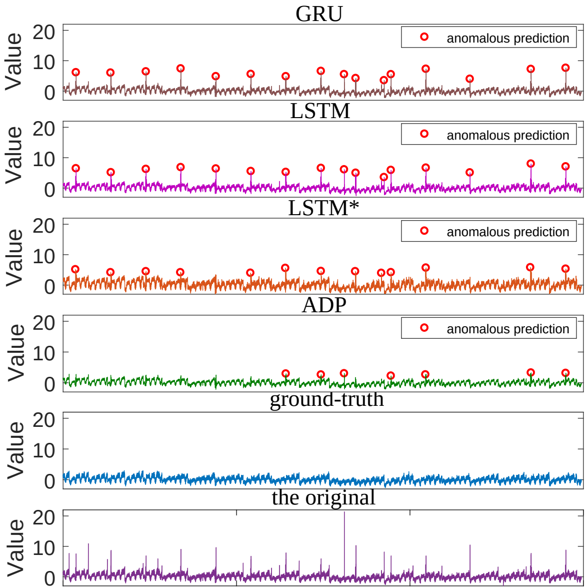
$$\mathrm{F}_1\text{-score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{6}$$

# Performance on Prediction task

Table: The prediction performance of ADP in comparison to ARIMA, Prophet, GRU, LSTM and LSTM* on **KPI** dataset

| | ARIMA | Prophet | GRU | LSTM | LSTM* | AD-P | $ADP^{-}$ | ADP |
|---|---|---|---|---|---|---|---|---|
| **$MSE_1$** | 0.8489 | 1.5447 | 0.2790 | 0.2750 | **0.2628** | 0.3068 | 0.2884 | 0.2906 |
| **$RMSE_1$** | 0.8735 | 1.1053 | **0.3669** | 0.3698 | 0.3682 | 0.3967 | 0.3849 | 0.3859 |
| **$MAE_1$** | 0.6149 | 0.8415 | 0.1803 | 0.1836 | 0.2353 | 0.1819 | 0.1726 | **0.1723** |
| **$MSE_2$** | 0.6278 | 1.3101 | 0.1850 | 0.1866 | 0.3215 | 0.1298 | 0.1107 | **0.1086** |
| **$RMSE_2$** | 0.7448 | 0.9777 | 0.3293 | 0.3349 | 0.4227 | 0.2957 | 0.2761 | **0.2724** |
| **$MAE_2$** | 0.6048 | 0.8253 | 0.1870 | 0.1904 | 0.2475 | 0.1828 | 0.1740 | **0.1704** |
| **$MSE_3$** | 0.6345 | 1.3227 | 0.1446 | 0.1441 | 0.2463 | 0.1201 | 0.1069 | **0.1059** |
| **$RMSE_3$** | 0.7606 | 0.9910 | 0.2632 | 0.2680 | 0.3502 | 0.2709 | 0.2602 | **0.2598** |
| **$MAE_3$** | 0.6014 | 0.8240 | 0.1727 | 0.1758 | 0.2309 | 0.1717 | 0.1629 | **0.1625** |

# Performance on Prediction: result samples



# Performance on Detection task

Table: Performance comparison on Anomaly Detection on **KPI** and **Yahoo**. The supervised approach is marked with '*'

| | **KPI** | | | **Yahoo** | | |
|---|---|---|---|---|---|---|
| Approach | $F_1$-score | Precision | Recall | $F_1$-score | Precision | Recall |
| **OCSVM** | 0.183 | 0.144 | 0.251 | 0.026 | 0.013 | 0.803 |
| **VAE-LSTM** | 0.061 | 0.033 | 0.423 | 0.026 | 0.014 | 0.244 |
| **SPOT** | 0.217 | 0.786 | 0.126 | 0.338 | 0.269 | 0.454 |
| **DSPOT** | 0.521 | 0.623 | 0.447 | 0.316 | 0.241 | 0.458 |
| **DONUT** | 0.595 | 0.735 | 0.500 | 0.501 | 0.669 | 0.401 |
| **SR** | 0.622 | 0.647 | 0.598 | 0.563 | 0.451 | 0.747 |
| **VAE** | 0.685 | 0.725 | 0.648 | 0.642 | 0.773 | 0.549 |
| ***SR-CNN** | **0.771** | 0.797 | 0.747 | 0.652 | 0.816 | 0.542 |
| **AD** | 0.726 | 0.884 | 0.615 | 0.737 | 0.806 | 0.678 |
| **ADP$^{-}$** | 0.711 | 0.757 | 0.670 | 0.734 | 0.881 | 0.630 |
| **ADP** | 0.739 | 0.839 | 0.660 | **0.755** | 0.837 | 0.688 |

# Summary

- Advantages
  1. High precisions are achieved for both detection and prediction tasks
- Disadvantages
  1. One model should be trained for one time series
  2. It is not adaptive to the moving trend
- Publication
  1. Run-Qing Chen, Guang-Hui Shi, Wan-Lei Zhao*, Chang-Hui Liang, "A Joint Model for IT Operation Series Prediction and Anomaly Detection", Neurocomputing'21

,

# Outline

1. Anomaly Detection and Trend Prediction by LSTM

2. Online Matrix Profile for Anomaly Detection

# Matrix Profile

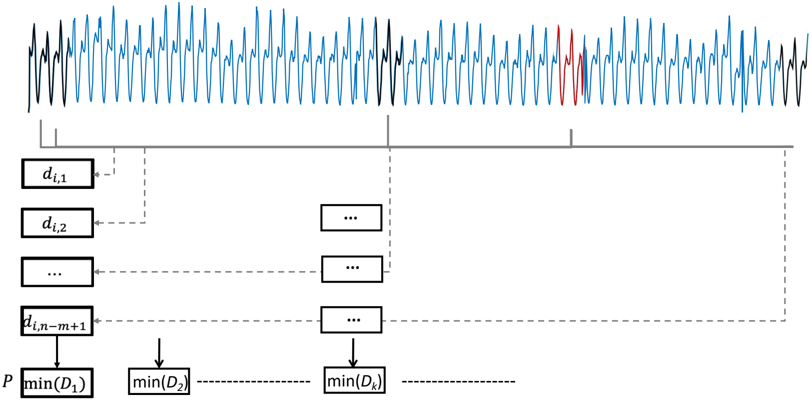
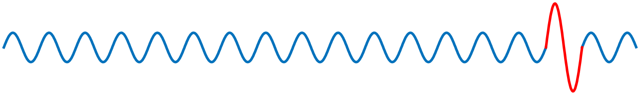



Figure: The demonstration of Matrix profile.

- Find out the closest subcurve for each subcurve (taking $\min(\cdot)$)

# Left Matrix Profile

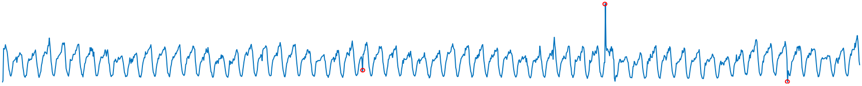


(a) Time series

(b) Left matrix profile without standard deviation alignment

- Anomaly on sin curve marked in red
- We only know the timestamps already occurred
- Left matrix profile is caculated

# Left Matrix Profile

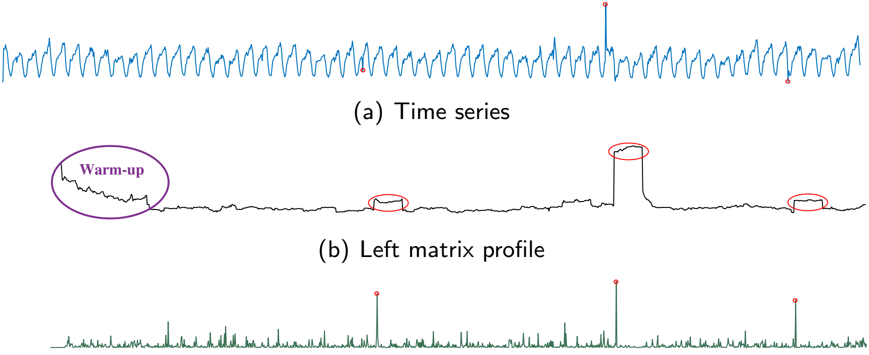


(a) Time series


(b) Left matrix profile

- Only timestamps on the left are known
- Distance significance on left matrix profile

# Distance Significance




(c) Distance significance

- Distance significance shows high response on low energy signal
- It is more precise than original left matrix profile

# Performance Evaluation (1)

Table: Ablation study about OMP on **KPI** and **Yahoo**. cache: cache strategy, DS: distance significance, SR: spectral residual

| | KPI | | | Yahoo | | |
|---|---|---|---|---|---|---|
| Approach | $F_1$-score | Precision | Recall | $F_1$-score | Precision | Recall |
| **SR** | 0.622 | 0.647 | 0.598 | 0.563 | 0.451 | 0.747 |
| **MP** | 0.525 | 0.424 | 0.687 | 0.599 | 0.679 | 0.536 |
| **MP**$^*$ | 0.597 | 0.565 | 0.633 | 0.752 | 0.750 | 0.753 |
| **MP**$^*$**+cache** | 0.541 | 0.495 | 0.597 | 0.752 | 0.710 | 0.799 |
| **MP**$^*$**+cache+DS** | 0.632 | 0.697 | 0.578 | 0.790 | 0.878 | 0.718 |
| **OMP** | **0.709** | 0.758 | 0.667 | **0.815** | 0.842 | 0.790 |

- All approaches shown here work online (no training is required)
- OMP: **MP**$^*$**+cache+DS** integrated with **SR** performs the best

# Compared to SOTA

Table: Comparison with the state-of-the-art approaches on testing data. The neural network-based approaches are marked with '‡', OL: Online, Prec.: Precision

| | | KPI | | | Yahoo | | |
|---|---|---|---|---|---|---|---|
| Approach | OL | $F_1$-score | Prec. | Recall | $F_1$-score | Prec. | Recall |
| **SPOT** | √ | 0.217 | 0.786 | 0.126 | 0.338 | 0.269 | 0.454 |
| **DSPOT** | √ | 0.521 | 0.623 | 0.447 | 0.316 | 0.241 | 0.458 |
| **SR** | √ | 0.622 | 0.647 | 0.598 | 0.563 | 0.451 | 0.747 |
| ‡**SR-CNN** | | **0.771** | 0.797 | 0.747 | 0.652 | 0.816 | 0.542 |
| ‡**DONUT** | | 0.595 | 0.735 | 0.500 | 0.501 | 0.669 | 0.401 |
| ‡**VAE** | | 0.685 | 0.725 | 0.648 | 0.642 | 0.773 | 0.549 |
| ‡**PAD** | | 0.739 | 0.839 | 0.660 | 0.755 | 0.837 | 0.688 |
| ‡**Online-VAE** | √ | 0.686 | 0.716 | 0.657 | 0.541 | 0.694 | 0.443 |
| ‡**Online-PAD** | √ | 0.731 | 0.806 | 0.669 | 0.681 | 0.711 | 0.653 |
| **OMP** | √ | 0.709 | 0.758 | 0.667 | **0.815** | 0.842 | 0.790 |

- Among all **online** approaches, OMP achieves the best performance
- Only 1.5 ms is required to process one timestamp

# Summary

- Advantages
  1. No training is required
  2. One procedure works for all types of time series
- Disadvantages
  1. Detection precision is inferior to deep learning approach
- Publication
  1. Shi-Ying Lan, Run-Qing Chen, Jie Zhao, Wan-Lei Zhao*, "Anomaly Detection on IT Operation Series via Online Matrix Profile", under review

,

Q & A

Thanks for your attention!