Robust Anomaly Detection for Multivariate Time Series through Stochastic Recurrent Neural Network

Ya Su, Youjian Zhao, Chenhao Niu, Rong Liu, Wei Sun, Dan Pei

SIGKDD 2019
Outline

- Background
- Algorithm
- Evaluation
- Conclusion
Outline

Background  Algorithm  Evaluation  Conclusion
Anomaly Detection

- **Graph** [SIGKDD 2018, AI Magazine 2014]

- **Log Messages** [SIGKDD 2016, SIGKDD 2017]

- **Time Series** [SIGKDD 2015, SIGKDD 2017, SIGKDD 2018]

  - Univariate Time Series
  - Mutivariate Time Series
Entities with monitored multivariate time series

Entities
- Server Machine
- SpaceCraft
- Robot-assisted System
- Engine

Multi-metrics
- CPU Load
- Network Usage
- Memory Usage
- Radiation
- Temperature
- Power
- Kinematic
- Visual
- Haptic
- Accelerator
- Torque
- Temperature
Entities with monitored multivariate time series

Entities

Server Machine

Multi-metrics

CPU Load
Network Usage
Memory Usage
...

More intuitive
More effective
More efficient

Multivariate time series
Machine with monitored multivariate time series

- TCP Active Opens
- TCP Retransmissions
- Memory Usage
- CPU Load
- Disk Write
- ETH1 inflow
- UDP out
- TCP timeout

Time (hour)
Machine with monitored multivariate time series
Motivations

How to detect the anomalies?

How to interpret the anomalies?
Challenges

• How to deal with the temporal dependence of multivariate time series?

• How to deal with the stochasticity of multivariate time series?

• How to provide interpretation to the detected entity-level anomalies?
## Related work

<table>
<thead>
<tr>
<th>Deterministic models</th>
<th>Stochastic based models</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSTM, LSTM-based Encoder-Decoder</td>
<td>DAGMM, LSTM-VAE</td>
</tr>
<tr>
<td>Deterministic models without stochastic variables</td>
<td>Ignore the dependence of time series or stochastic variables.</td>
</tr>
</tbody>
</table>
Outline

- Background
- Algorithm
- Evaluation
- Conclusion
OmniAnomaly

Helps answer the questions
Structure of OmniAnomaly

Offline Model Training

Data Pre-processing \rightarrow Model Training \rightarrow Threshold Selection

Multivariate time series data

Online Anomaly Detection

Model \rightarrow Anomaly Detection \rightarrow Anomaly Result

Threshold \rightarrow Anomaly Interpretation

Online Model Training

Multivariate time series data
Model Architecture of OmniAnomaly

(a1) qnet

(a2) pnet
Model Architecture of OmniAnomaly

- **Input Sequence data**
- **Reconstructed data**
- **GRU cells for capturing temporal dependence**
- **Stochastic cells for modeling data distribution**
- **GRU cells for capturing temporal dependence**
- **Input Sequence data**
Core idea of OmniAnomaly

A good $z_t$ can represent $x_t$ well no matter $x_t$ is anomalous or not.

When $x_t$ is anomalous, its $z_t$ can still represent its normal pattern and $x'_t$ will be normal too.
Anomaly detection of OmniAnomaly

Anomaly Score $S_t = \frac{\text{Reconstruction probability of } x_t \text{ > Threshold}}{Y \rightarrow \text{Normal} \quad \text{Anomalous} \quad N}$
Anomaly detection of OmniAnomaly

Anomaly Score $S_t =$ Reconstruction probability of $x_t$

$$x_t = [x_t^1, x_t^2, ..., x_t^M], \ M \text{ is the dimension}$$

$$S_t = \sum_{i=1}^{M} S_t^i$$

Sort the $[S_t^1, S_t^2, ..., S_t^M]$ in ascending order, and the Top K dimensions can interpret the anomaly.
### Datasets

<table>
<thead>
<tr>
<th>Dataset name</th>
<th>Number of entities</th>
<th>Number of dimensions</th>
<th>Training set size</th>
<th>Testing set size</th>
<th>Anomaly ratio(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMAP</td>
<td>55</td>
<td>25</td>
<td>135183</td>
<td>427617</td>
<td>13.13</td>
</tr>
<tr>
<td>MSL</td>
<td>27</td>
<td>55</td>
<td>58317</td>
<td>73729</td>
<td>10.72</td>
</tr>
<tr>
<td>SMD</td>
<td>28</td>
<td>38</td>
<td>708405</td>
<td>708420</td>
<td>4.16</td>
</tr>
</tbody>
</table>
F1-best of OmniAnomaly and baselines
F1-best of OmniAnomaly and variants

[Graph showing F1-best results for different datasets and models, including OmniAnomaly, C1-RNN, C1-LSTM, C2-qnet, C2-no z connected, and C2-pnet.]
F1 obtained through POT vs. F1-best

<table>
<thead>
<tr>
<th>Evaluation metrics for OmniAnomaly</th>
<th>SMAP</th>
<th>MSL</th>
<th>SMD</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1 obtained through POT</td>
<td>0.8434</td>
<td>0.8989</td>
<td>0.8857</td>
</tr>
<tr>
<td>F1-best</td>
<td>0.8535</td>
<td>0.9014</td>
<td>0.9620</td>
</tr>
</tbody>
</table>
F1-best of OmniAnomaly with different z dimension
OmniAnomaly

• The first multivariate time series anomaly detection method that deal with explicit temporal dependence among stochastic variables

• The first anomaly interpretation approach for stochastic based multivariate time series anomaly detection algorithms

• Achieve an overall F1-score of 0.86 in three real world datasets.

• The interpretation accuracy is up to 0.89.
Lessons for time series data learning

• A combination of stochastic deep Bayesian model and deterministic RNN model is necessary

• The connection of stochastic variables is necessary and effective

• It is necessary to assume non-Gaussian distributions in z-space
Lessons for multivariate time series anomaly detection

- Reconstruction-based models are more robust than prediction-based models.

- It is critical to obtain robust latent representations which can accurately capture the normal patterns of time series.

- Reconstruction-based stochastic approaches offer an opportunity to interpret the anomalies.
Thanks

su-y16@mails.tsinghua.edu.cn