Orchestrating a brighter world NEC

A Deep Neural Network for Unsupervised Anomaly Detection and Diagnosis in Multivariate Time Series

Chuxu Zhang[#$], Dongjin Song[$], Yuncong Chen[$], Xinyang Feng[$&], Cristian Lumezanu[$], Wei Cheng[$], Jingchao Ni[$], Bo Zong[$], Haifeng Chen[$], Nitesh V. Chawla[#]

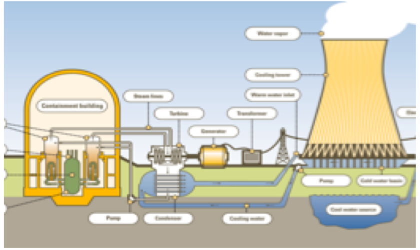
[#]University of Notre Dame, [$]NEC Labs America, [&]Columbia University

NEC Laboratories America
Relentless passion for innovation

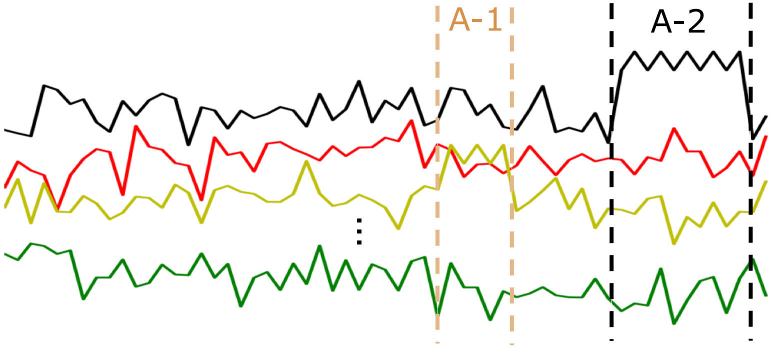
Motivation

- Large Scale Complex Systems/Multivariate Time Series

Power Plants



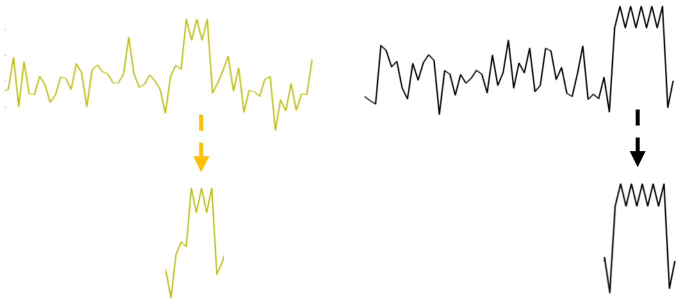
IT Systems

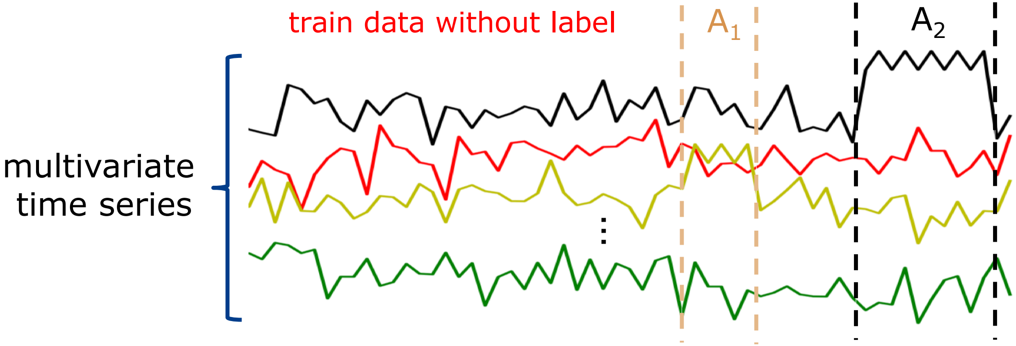


❖ Normal period: little/no label
❖ Abnormal period: few root causes, multi-scale (duration) anomalies

Goal

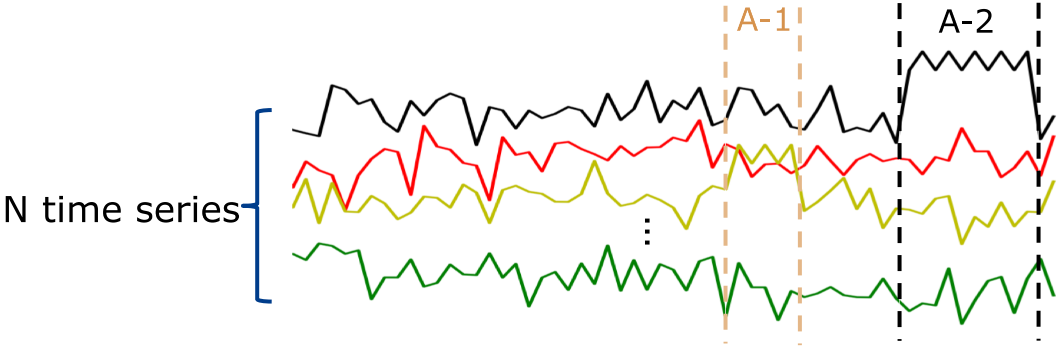
- Unsupervised Anomaly Detection: A_1, A_2
- Anomaly Diagnosis
  - ✓ Root cause identification
    find causal sensor
  - ✓ Anomaly scale analysis
    interpret anomaly duration





Challenge

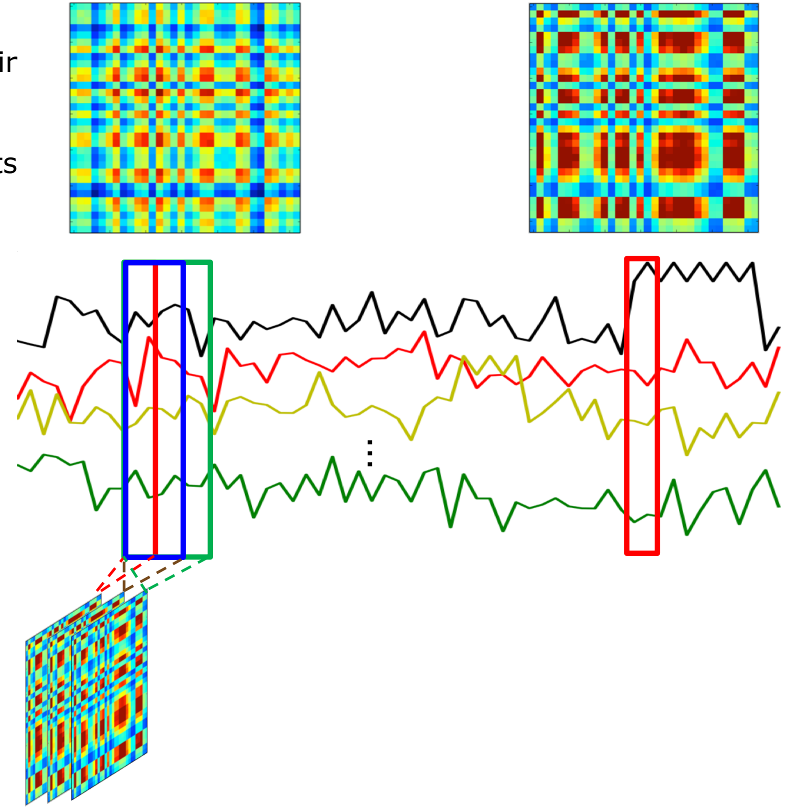
- C1: Time series contain noise
  false positive of temporal dependency based models, e.g., SIAT[1], ARMA, LSTM-AE[2]
- C2: Multi-dimensional input, Temporal dependency
  density based models can not capture, e.g., OC-SVM[3], DMM[4]
- C3: Multi-scale (duration) anomalies
  both temporal dependency/density based models can not handle



[1] Exploiting local and global invariants for the management of large scale information systems, ICDM 2008
[2] A Dual-stage attention-based recurrent neural network for time series prediction, IJCAI 2017
[3] One-class SVMs for document classification, JMLR 2001
[4] Deep autoencoding gaussian mixture model for unsupervised anomaly detection, ICLR 2018

Model

- System Signature for C1 (avoid noise)
  - Signature matrix: compute inner-product between every pair of sensors on each time segment
  - Capture both shape and range
  - Robust to noise as the noise of individual time series impacts little on the signature of the whole system



- Auto-Encoder for C2 (multi-dimen, temporal)
  - Signature matrix pattern encoding: CNN
  - Temporal dependency modeling: RNN
  - Signature matrix pattern decoding: CNN
  - Profiling the normal period for model training, test the abnormal period

- Multi-Scale Matrices for C3 (multi-scale)
  - Multi-scale (resolution) signature matrices

Model

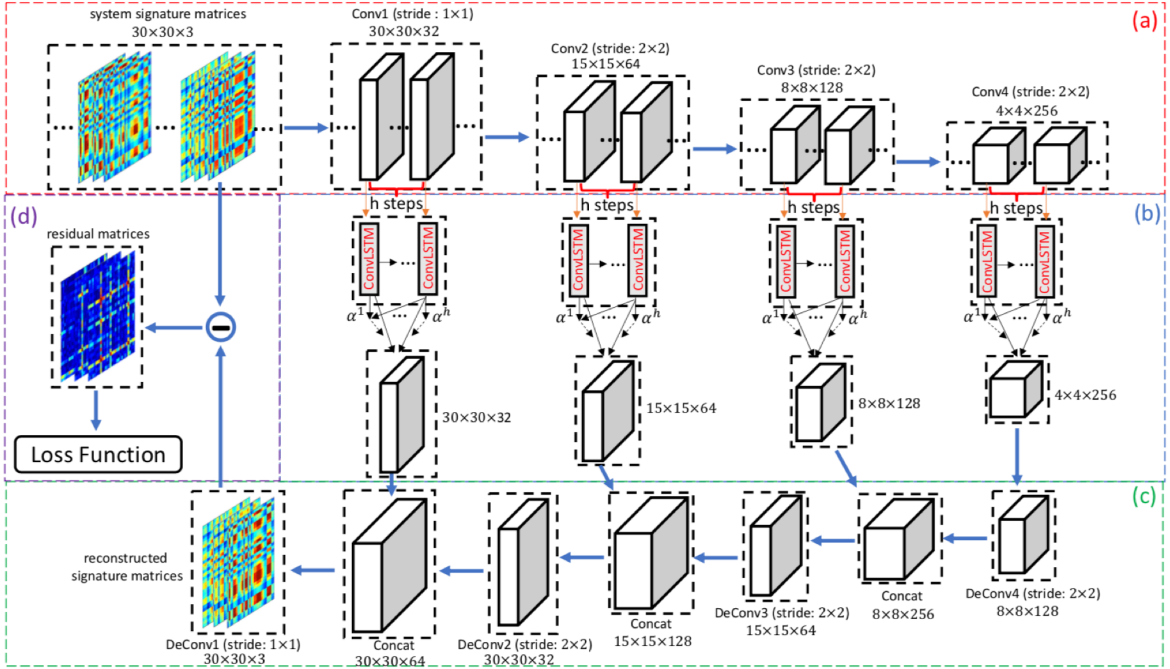


Figure 2: Framework of the proposed model: (a) Signature matrices encoding via fully convolutional neural networks. (b) Temporal patterns modeling by attention based convolutional LSTM networks. (c) Signature matrices decoding via deconvolutional neural networks. (d) Loss function.

- ❖ Signature matrix encoding: 4 layer CNNs
- ❖ Temporal dependency modeling: convLSTM
- ❖ Signature matrix decoding: 4 layer CNNs
- ❖ Connect to convLSTM in each conv layer for model enhancement
- ❖ Anomaly score: number of broken elements in residual matrix (by cutoff threshold)
- ❖ MSCRNN: multi-scale (resolution) convolutional recurrent auto-encoder

Experiment

Dataset

- ✓ **Synthetic data**: 30 time series, 20000 points, train: 0 - 8000, validate: 8001 - 10000 test: 10001 - 20000, 5 anomalies, 3 root causes for each anomaly
- ✓ **Real world Power Plant data**: 36 time series, 23040 points, train: 0 – 10080, validate: 10081 – 18720, test: 18721 – 23040, 5 anomalies, 3 root causes for each anomaly

Baseline

- ✓ **Classification model**: One Class-SVM(OC-SVM)[1],
- ✓ **Density estimation model**: Deep Autoencoding Gaussian Mixture Model(DAGMM)[2]
- ✓ **Prediction model**: History Average(HA), Auto-Regression Moving Average(ARMA), LSTM Encoder-Decoder(LSTM-ED)[3]
- ✓ **Model variant**: ConvLSTM layers removed ($CNN^{ED(3,4)}_{ConvLSTM}$, $CNN^{ED(4)}_{ConvLSTM}$), attention module removed ($CNN^{ED}_{ConvLSTM}$)

Metric

- ✓ **Recall**, **Precision, F1 Score**
- ✓ Experiment on both synthetic data and real world data are repeated 5 times, average scores are reported

[1] One-class SVMs for document classification, JMLR 2001
[2] Deep autoencoding gaussian mixture model for unsupervised anomaly detection, ICLR 2018
[3] A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction, IJCAI 2017

Experiment

- Anomaly Detection: Performance

Table 2: Anomaly detection results on two datasets.

Method	Synthetic Data			Power Plant Data		
	Pre	Rec	F_1	Pre	Rec	F_1
OC-SVM	0.14	0.44	0.22	0.11	0.28	0.16
DAGMM	0.33	0.20	0.25	0.26	0.20	0.23
HA	0.71	0.52	0.60	0.48	0.52	0.50
ARMA	0.91	0.52	0.66	0.58	0.60	0.59
LSTM-ED	<u>1.00</u>	<u>0.56</u>	<u>0.72</u>	<u>0.75</u>	<u>0.68</u>	<u>0.71</u>
$CNN^{ED(4)}_{ConvLSTM}$	0.37	0.24	0.29	0.67	0.56	0.61
$CNN^{ED(3,4)}_{ConvLSTM}$	0.63	0.56	0.59	0.80	0.72	0.76
$CNN^{ED}_{ConvLSTM}$	0.80	0.76	0.78	0.85	0.72	0.78
MSCRED	**1.00**	**0.80**	**0.89**	**0.85**	**0.80**	**0.82**
Gain (%)	–	30.0	23.8	13.3	19.4	15.5

- ❖ Temporal prediction models perform better than classification model and density based models. Both synthetic and real world datasets have time dependency property
- ❖ LSTM-ED has better performance than ARMA, indicating deep learning based model achieves better generalization ability than traditional temporal dependency models
- ❖ Our proposed MSCRNN performs best on all metrics of two datasets, demonstrating the effectiveness of MSCRNN
- ❖ With the increment of ConvLSTM layers, the performance of MSCRED improves
- ❖ The attention module further improves anomaly detection performance

Experiment

- Anomaly Detection: Case Study

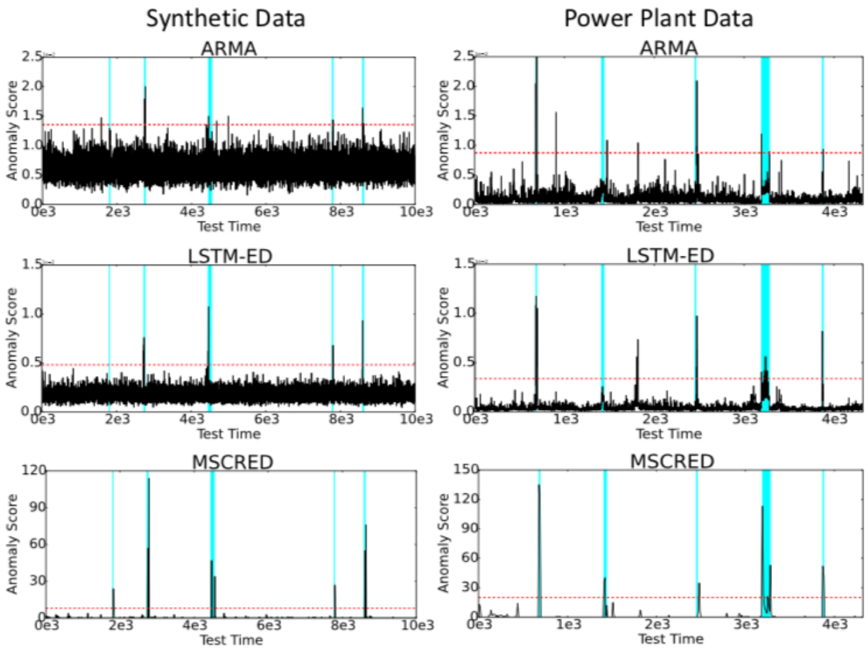


Figure 3: Case study of anomaly detection. The shaded regions represent anomaly periods. The red dash line is the cutting threshold of anomaly.

- ❖ The anomaly score of ARMA is not stable and the results contain many false positives and false negatives
- ❖ The anomaly score of LSTM-ED is smoother than ARMA while still contains several false positives and false negatives
- ❖ MSCRED can detect all anomalies without any false positive and false negative

Experiment

- Anomaly Diagnosis: Performance

✓ Root Cause Identification

- ❖ Ranking sensors based on anomaly score, find causal sensors
- ❖ MSCRNN performs better than the best baseline LSTM-ED for both datasets

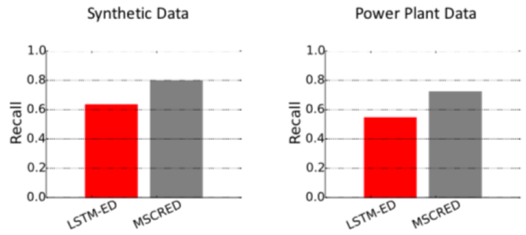


Figure 5: Performance of root cause identification.

✓ Anomaly Scale Analysis

- ❖ MSCRNN(S) detects three types of anomalies
- ❖ MSCRNN(M) detects medium and long anomalies
- ❖ MSCRNN(L) detects long anomaly
- ❖ Interpret anomaly types (duration) by joint considering three detection results

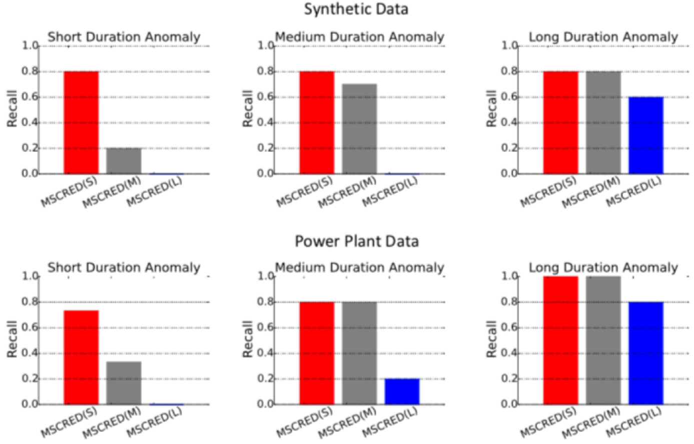


Figure 6: Performance of three channel anomaly scores of MSCRED over different types of anomalies.

MSCRNN(S): anomaly score computed on residual matrix of small-scale widow
MSCRNN(M): anomaly score computed on residual matrix of medium-scale widow
MSCRNN(L): anomaly score computed on residual matrix of large-scale widow

Experiment

○ Anomaly Diagnosis: Case Study

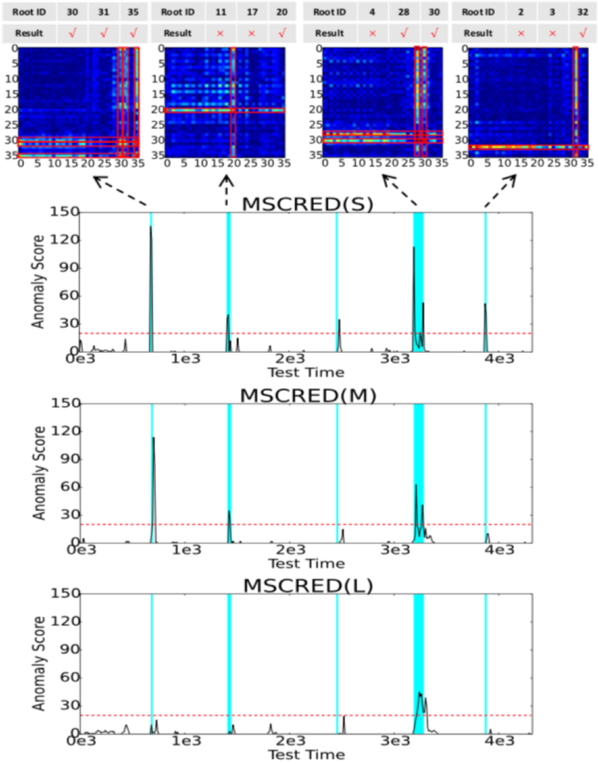


Figure 7: Case study of anomaly diagnosis.

- ❖ MSCRED(S) detects all of 5 anomalies including 3 short, 1 medium and 1 long duration anomalies. MSCRED(M) misses two short duration anomalies and MSCRED(L) only detects long duration anomaly
- ❖ We can accurately pinpoint more than half of the anomaly root causes (rows/columns highlighted by red rectangles) in this case

○ Robustness to Noise

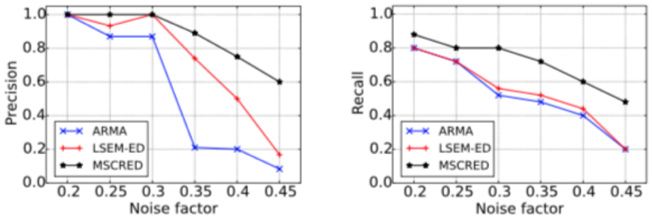


Figure 8: Impact of data noise on anomaly detection.

- ❖ MSCRED consistently outperforms ARMA and LSTM-ED when the scale of noise varies from 0.2 to 0.45. Compared with ARMA and LSTM- ED, MSCRED is more robust to the input noise

Conclusion

- One Innovative Model
  - ✓ Multi-scale (resolution) signature matrices for the whole system
  - ✓ System signature encoding via CNN
  - ✓ Temporal dependency modeling via ConvLSTM
  - ✓ System signature decoding via CNN
- Two Useful Applications
  - ✓ Anomaly detection
  - ✓ Anomaly diagnosis: root cause identification, anomaly scale interpretation
- Experiment Demonstration
  - ✓ Both synthetic data and real data
  - ✓ Four category baselines
  - ✓ Best performance for all metrics in both datasets
  - ✓ Robustness to noise

Q&A