LSTM-based Encoder-Decoder for Multi-sensor Anomaly Detection

Pankaj Malhotra, Anusha Ramakrishnan, Gaurangi Anand, Lovekesh Vig,
Puneet Agarwal, and Gautam Shroff

ICML, 2016

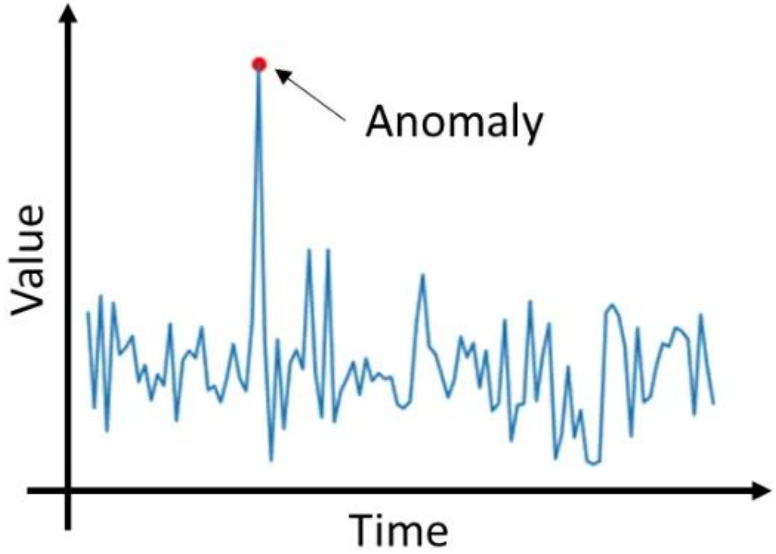
2020. 1. 22
Seonyoung Kim

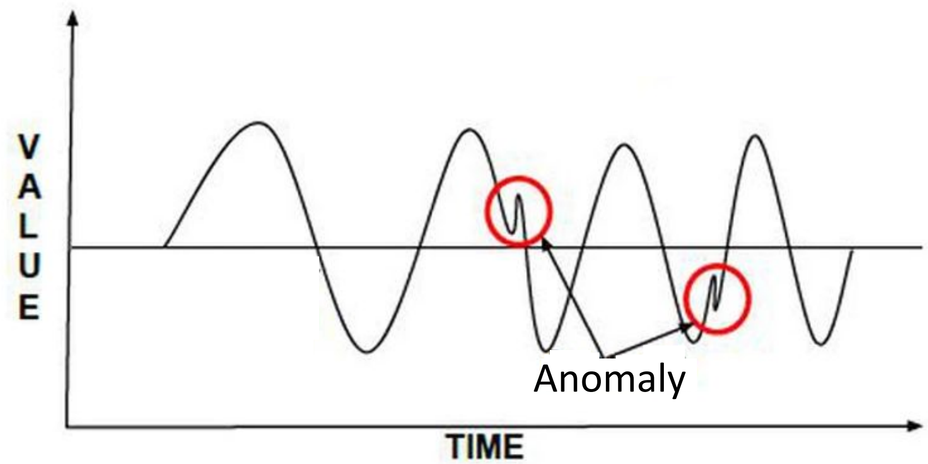
CONTENTS

Introduction

- Mechanical devices are instrumented with numerous sensors to capture the behavior and health of the machine

- Anomaly Detection
  - the process of identifying unexpected items or events in data sets, which differ from the normal data

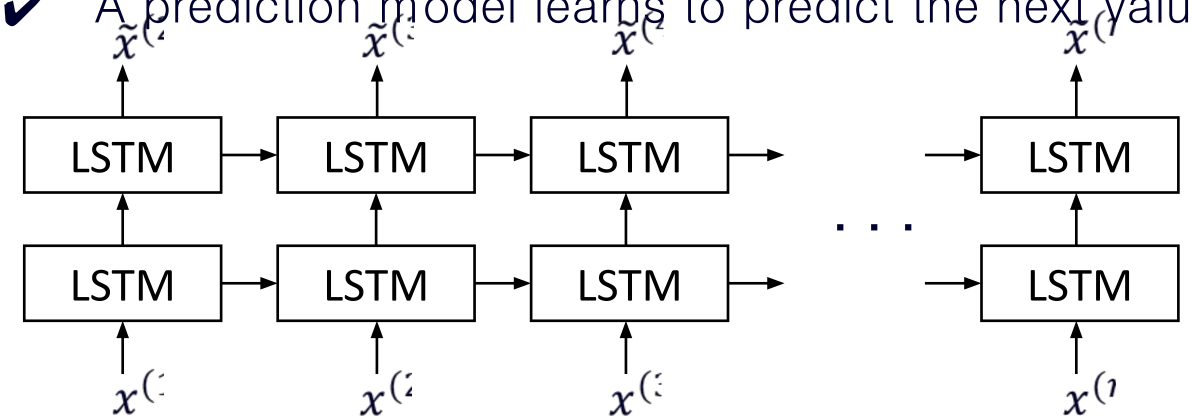




Related Work and Motivations

- Traditional techniques monitoring anomalies use statistical measures to detect changes in the underlying distribution
  - Exponentially weighted moving average (EWMA), SVR

- Long Short Term Memory Networks for Anomaly Detection in Time Series [4]
  - Stacked LSTM based prediction model
    - ✔ A prediction model learns to predict the next values



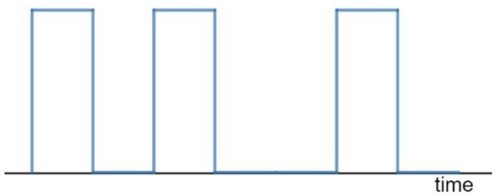
$A\ time\ series\ X = \{x^{(1)}, x^{(2)}, \ldots, x^{(L)}\}\ of length\ L$

$x^{(i)} \in R^m :\ an\ m - dimensinal\ vector\ of\ readings\ for\ m\ variables\ at\ time\ instance\ t_i$

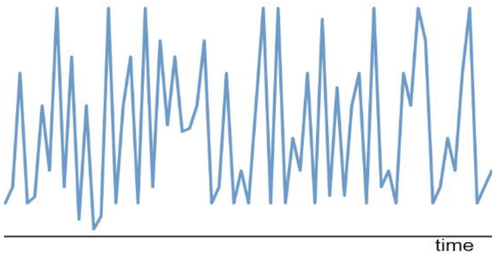
$\tilde{x}^{(i)}\ :\ the\ predicted\ value\ of x^{(i)}$

Related Work and Motivations

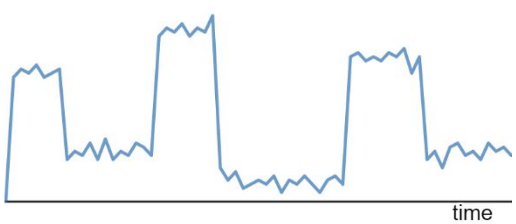
- Predictable & Unpredictable time-series
  - control variable : variable which are not captured by sensor and it can be simple or changes frequently in certain range
  - result = predictable time series or unpredictable time-series



Readings for a control variable with two states



Readings for a control variable that changes frequently



Predictable time-series

?

Unpredictable time-series

- Detecting anomalies with control variable that changes frequently becomes challenging using traditional approaches based on prediction models

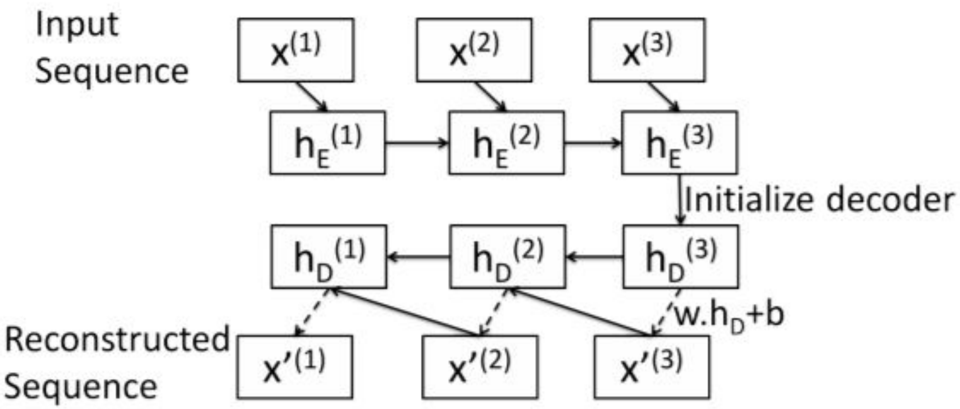
Proposed method: EncDec-AD

- Solve the problem of the prediction model for anomaly detection
  - Predictable & Unpredictable problem
  - Uses autoencoder-based method to detect the anomalies

- Overview
  - Train LSTM Encoder-Decoder model
    1. **the LSTM Encoder learns a fixed length vector representation of the normal time-series data**
    2. **the LSTM Decoder uses this representation to reconstruct the time-series**
  - Compute likelihood of anomaly
    - ✔ With trained model, compute likelihood of anomaly of time-series data

Proposed method: EncDec–AD

▪ LSTM Encoder–Decoder as reconstruction model



✓ $a\ time\ series\ X = \{x^{(1)}, x^{(2)}, \dots, x^{(L)}\}\ of\ length\ L$

✓ $x^{(i)} \in R^m$: $an\ m-dimensinal\ vector\ of\ readings\ for\ m\ variables\ at\ time\ instance\ t_i$

✓ $h_E^{(i)} \in R^c$: $the\ hidden\ state\ of\ encoder\ at\ time\ t_i\ for\ each\ i \in \{1,2,\dots,L\},$
$where\ c\ is\ the\ number\ of\ LSTM\ units\ in\ the\ hidden\ layer\ of\ the\ encoder$

✓ $h_D^{(i)} \in R^c$: $the\ hidden\ state\ of\ decoder\ at\ time\ t_i\ for\ each\ i \in \{1,2,\dots,L\},$
$where\ c\ is\ the\ number\ of\ LSTM\ units\ in\ the\ hidden\ layer\ of\ the\ decoder$

✓ w : $weight\ matrix$

✓ b : $bias$

- The model is trained to minimize the objective function:

$$\sum_{X \in s_N} \sum_{i=1}^{L} \left\| x^{(i)} - x'^{(i)} \right\|^2$$

✓ s_N : $set\ of\ normal\ training\ sequences$

Proposed method: EncDec-AD

- Computing likelihood of anomaly(1/3)
  - Divide the normal time-series into four sets of time-series
    - ✓ s_N : $set\ of\ normal\ training\ sequences$
    - ✓ v_{N1} : $set\ of\ normal\ validation\ sequences\_1$
    - ✓ v_{N2} : $set\ of\ normal\ validation\ sequences\_2$
    - ✓ t_N : $set\ of\ normal\ test\ sequences$

  - Divide the anomalous time-series into two set of time-series
    - ✓ v_A : $set\ of\ anomalous\ validation\ sequences$
    - ✓ t_A : $set\ of\ anomalous\ test\ sequences$

Proposed method: EncDec-AD

- Computing likelihood of anomaly(2/3)
  - The set of sequence s_N is used to learn the LSTM encoder-decoder reconstruction model
  - The reconstruction error vector at time t_i
    - ✓ $The\ error\ vector\ e^{(i)} = \left|x^{(i)} - x'^{(i)}\right|$
  - The error vectors for the points in the sequences in v_{N1} are used to estimate a Normal distribution $N(\mu, \Sigma)$ using Maximum Likelihood Estimation
    - ✓ Given $v_{N1}\ set\ \{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$
    - ✓ $\mu = \frac{1}{m}\sum_{i=1}^{m} x^{(i)}$
    - ✓ $\Sigma = \frac{1}{m}\sum_{i=1}^{m}(x^{(i)} - \mu)(x^{(i)} - \mu)^T$

Proposed method: EncDec-AD

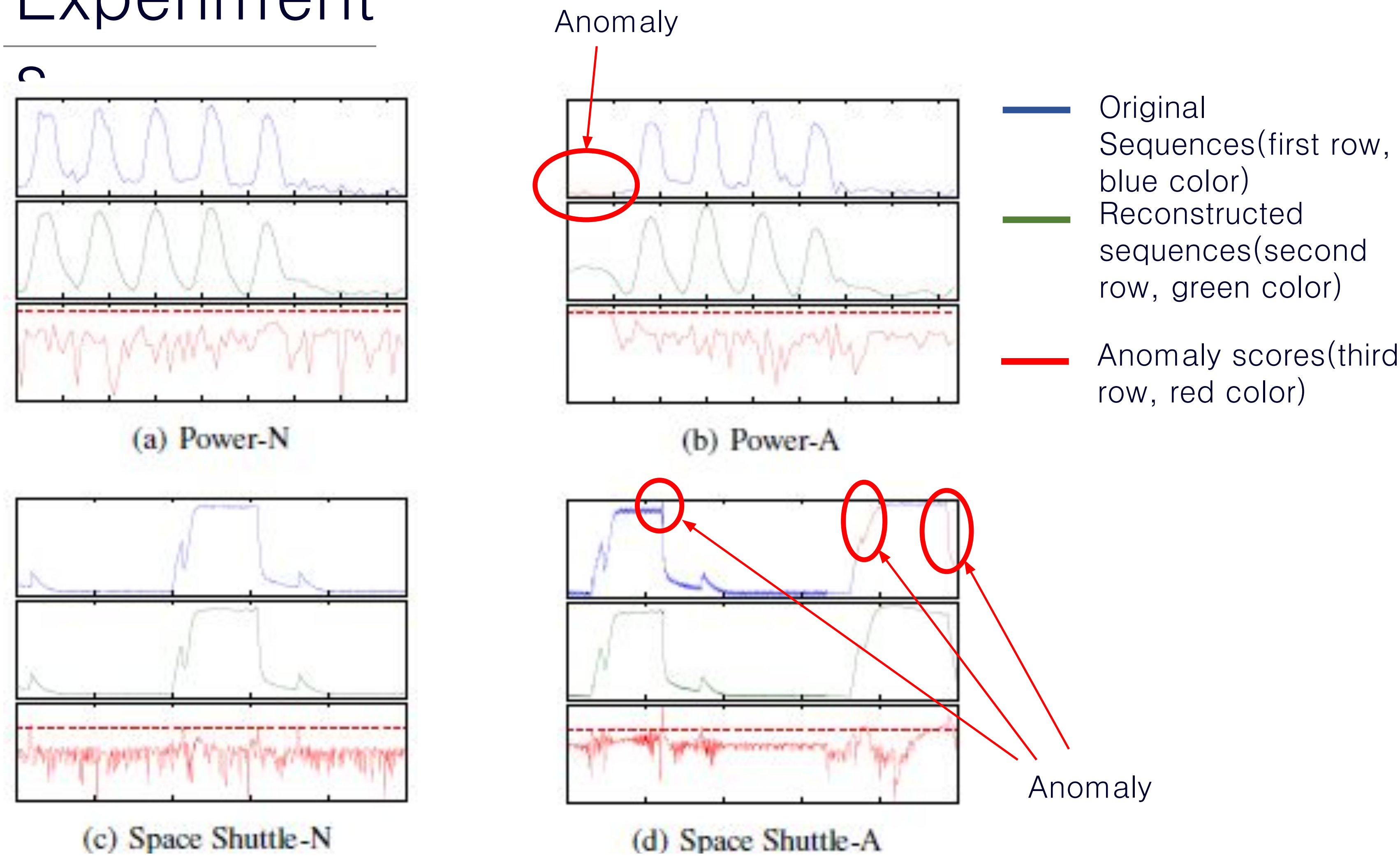
- Computing likelihood of anomaly(3/3)
  - The anomaly score
    - ✓ For any point $x^{(i)}$
    - ✓ $a^{(i)} = \left(e^{(i)} - \mu\right)^T \Sigma^{-1} \left(e^{(i)} - \mu\right)$
    - ✓ If $a^{(i)} > \tau$, a point in a sequence can be predicted → "anomalous"
    - ✓ Otherwise, a point in a sequence can be predicted → "normal"

  - The v_{N2} and v_A are used to learn τ by maximizing $F_\beta - score$
    - ✓ $F_\beta = (1 + \beta^2) \times P \times \frac{R}{\beta^2 P + R}$
    - ✓ $P : precision$
    - ✓ $R : recall$

Experiment

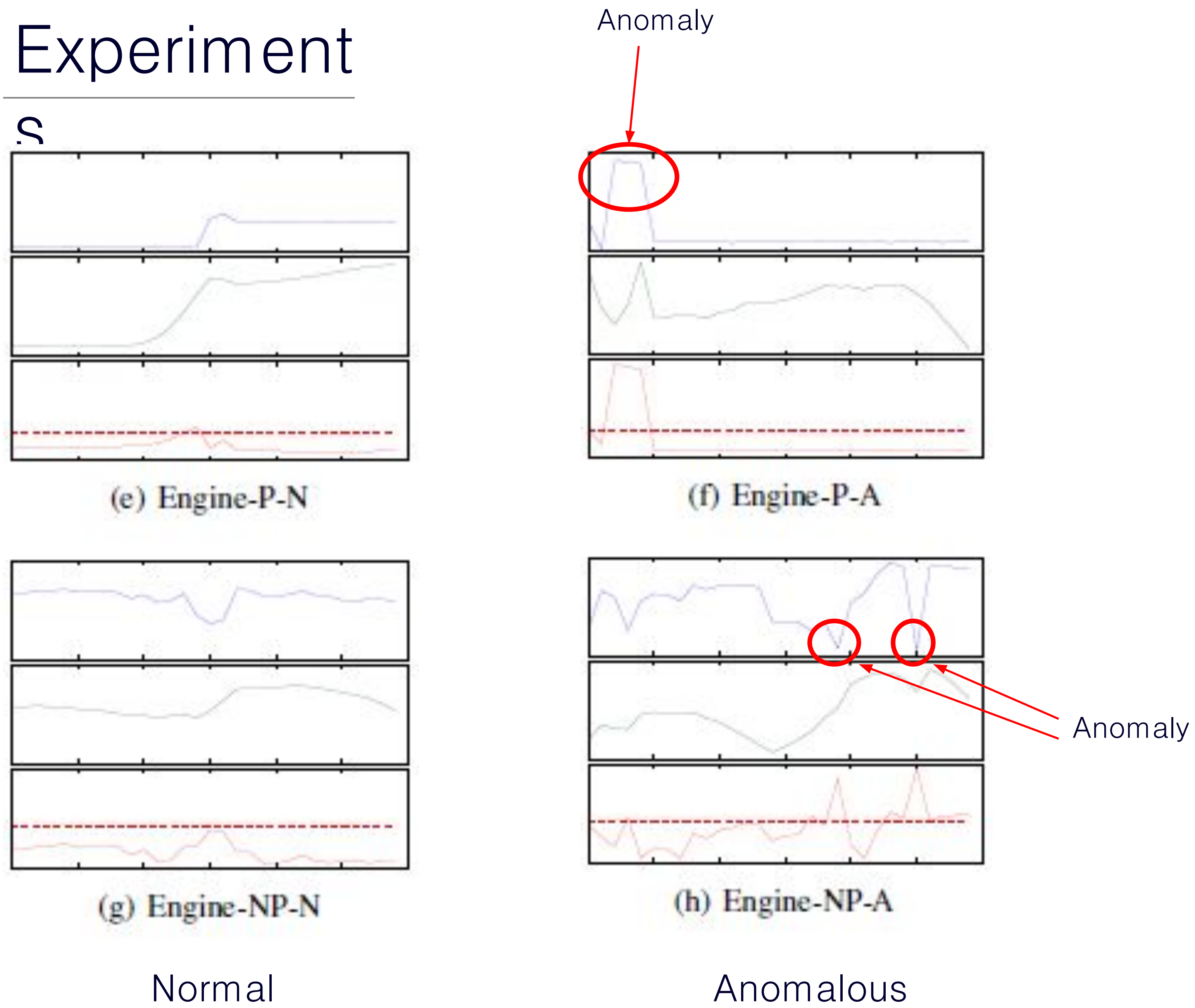
- Dataset
  - Power demand
    - ✔ One univariate time-series with 35,040 readings for power demand recorded over a period of one year
    - ✔ Length of sequence = 84
  - Space shuttle
    - ✔ Periodic sequences with 1000 points per cycle, and 15 such cycles
    - ✔ Length of sequence = 1500
  - Engine data
    - ✔ Reading for 12 sensors such as coolant temperature, torque, accelerator etc.
    - ✔ Consider two different applications of the engine : Engine-P, Engine-NP
    - ✔ Engine-P has a discrete external control with two states : 'high' and 'low' → "predictable"
    - ✔ Engine-NP has any value within a certain range and changes very frequently → "unpredictable"
    - ✔ Length of sequence = 30
  - ECG
    - ✔ Quasi-periodic time-series
    - ✔ Contains one anomaly corresponding to a pre-ventricular contraction
    - ✔ Length of sequence = 208

Experiments

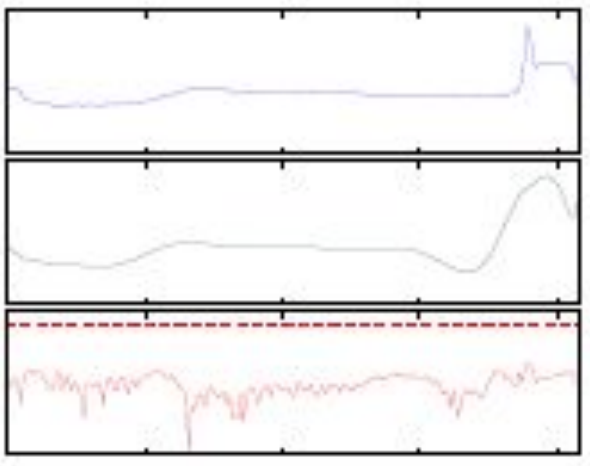


Normal Anomalous

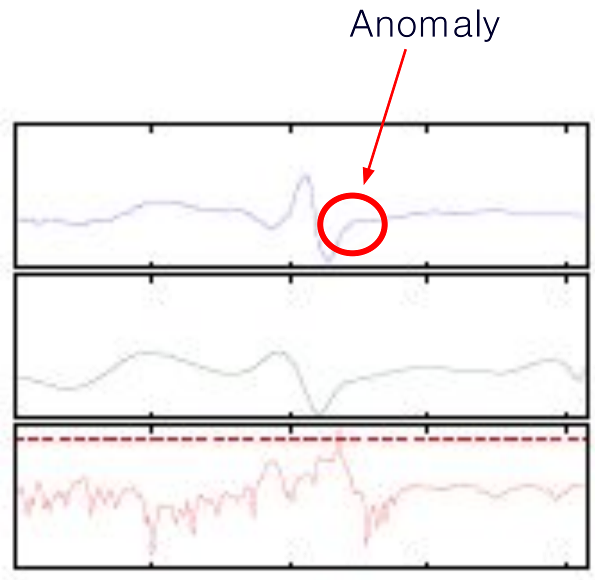
Experiments



Experiment
s



(i) ECG-N



(j) ECG-A

Observation

Datasets	L	c	β	P	R	F_β-score	TPR/FPR
Power Demand	84	40	0.1	0.92	0.04	0.77	33.0
Space Shuttle	500	50	0.05	0.83	0.08	0.81	4.9
Engine-P	30	40	0.05	0.94	0.02	0.82	13.8
Engine-NP	30	90	0.05	1.0	0.01	0.83	∞
ECG	208	45	0.05	1.0	0.005	0.65	∞

- ✓ L : *Length of sequence*
- ✓ c : *the number of LSTM units in the hidden layer of encoder and decoder*
- ✓ β : *the value of* β *in* $F_\beta - score$
- ✓ P : *Precision*
- ✓ TPR : *Ture Positive Rates*
- ✓ FPR : *False Positive Rates*

- The positive likelihood ratio(TPR/FPR) > 1 for all the datasets
  - The probability of reporting an anomaly in anomalous region is much higher than the probability of reporting an anomaly in normal region.

- For periodic time-series, we experiment with varying window lengths
  - Being able to detect anomalies in all scenarios

- Compare to LSTM-AD[3], this method gives better results for Engine-NP where the sequence are not predictable

Conclusion

- This paper proposes EncDec-AD for anomaly detection in multi-sensor time-series
- EncDec-AD first learns a LSTM-based Encoder-Decoder model to reconstruct values of normal time-series
- Then, EncDec-AD computes anomaly score of each time step by calculating the likelihood value with reconstruction error vectors
- Experimental results show that EncDec-AD detects anomalies in not only predictable time-series but also unpredictable time-series data

References

[1] SUTSKEVER, I.; VINYALS, O.; LE, Q. V. Sequence to sequence learning with neural networks. *Advances in NIPS*, 2014.

[2] CHO, Kyunghyun, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[3] HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long short-term memory. *Neural computation*, 1997, 9.8: 1735-1780.

[4] Malhotra, Pankaj, Vig, Lovekesh, Shroff, Gautam, and Agarwal, Puneet. Long short term memory networks for anomaly detection in time series. In ESANN, *23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2015.

Thank you