2018 IEEE 16th Int. Conf. on Dependable, Autonomic & Secure Comp., 16th Int. Conf. on Pervasive Intelligence &
Comp., 4th Int. Conf. on Big Data Intelligence & Comp., and 3rd Cyber Sci. & Tech. Cong.

# Applications of Anomaly Detection using Deep Learning on Time Series Data

[1]Van Quan Nguyen, [1]Linh Van Ma, [1]Jin-young Kim, [2]Kwangki Kim, [1*]Jinsul Kim
[1]*School of Electronics and Computer Engineering*
*Chonnam National Univiversity*
Yongbong-ro, Buk-gu, Gwangju 500-757, Korea

[2]*School of IT Convergence*
*Korea Nazarene University*
Cheonan-si, Korea

quanap5@gmail.com, linh.mavan@gmail.com, beyondi@jnu.ac.kr, k2kim@ac.kr, jsworld@chonnam.ac.kr

***Abstract*—In the modern world, time series data has become a critical part of many systems underlying various types that are recorded to reflect the status of objects according to the timeline. There are many kinds of research investigating to automate the process of analyzing time series data. Long Short-Term Memory (LSTM) network have been demonstrated to be a useful tool for learning sequence data. In this paper, we explore LSTM based approach to analyzing temporal data for abnormal detection. Stacked Long Short-Term Memory (LSTM) network is utilized as a predictor which is trained on normal data to learn the higher level temporal features, then such predictor is used to predict future values. An error-distribution estimation model is built to calculate the anomaly in the score of the observation. Anomalies are detected using a window-based method based on anomaly scores. To prove the promise applicable potential of our approach, we conducted the experiment on some domains (industry system, health monitor system, social based event detection system) come up with time series data including power consumption, ECG signal, and social data respectively.**

***Keywords—Deep Learning, Recurrent Neural Network (RNN), Long Short Term Memory (LSTM), Time Series Data, Anomaly Detection***

## I. INTRODUCTION

In the real-world, mechanical devices such engines, vehicles, even body part of human are typically instrumented by the various physical sensor to recorded the behavior and health of objectives. For monitoring status during operation, we would like to be able to discriminate between the normal and anomalous status of a considered system [1]. For example, we analyze the signal from sensor built-in smart factory to recognize that what is going wrong, need to be replaced. Based on the physiological signal, the expert can determine the status of health of any person. However, in many situations, detecting anomaly from normal signal become challenging since the definition of abnormal or normal may frequently change. Generally, anomaly detection is the identity of data points, pattern, observations or events that do not conform to the desired pattern of a given dataset. The usage anomaly detection system is very helpful in behavior analysis or support for other kinds of analysis like detection, identification, and prediction of the occurrence of these anomalies.

The remainder of the paper is organized as follows: Section 2 briefly introduces background about LSTM and related works. Section 3 focus on design LSTM as a predictor for anomaly detection problem and its applications. Then, that different data from several domains are applied to investigate performance is the main content of Section 4. Last, Section 5 is for the conclusion and more discussion.

## II. RELATED WORKS

Anomalies are typically defined in term of deviation from some expected behavior or considered as patterns in data that not conform to a well-defined notion of normal data. For data over time, [2] compute a vector of features which may include lag correlation, the strength of seasonality, spectral entropy on each time series, measure the characteristic of the series. And then the bivariate outlier detection methods used on highest density region and α-hulls are applied to the first two principal component of the principal component analysis. While [3] proposed Relaxed of linear programming Support Vector Data Description (RLPSVDD) to solve an anomalous problem of cloud services. The burst algorithm is introduced in [4] to detect the disaster-related social data also is an application of anomaly detection on time series data.

Physiological signals (EEG, ECG) embody human activity, other kinds of data like communication network traffic or sensor data from industrial factories are the typical instances of time series data. Traditional anomaly detection method usually based on statistical measure, in which, we have to identify irregularities in given data to flag data point that deviate from common statistical properties of a distribution, including mean, median, mode, and quantities [1]. Since features come from special unit “memory cell” [5] LSTM neural network is also utilized to overcome the vanish gradient problem that is experienced by Recurrent Neural Network (RNN). LSTM networks are improved version of Recurrent Neural Network (RNN) [6] that have been used for many sequences learning tasks due to capacity of learning long-term dependencies.

[7] proposed an unsupervised approach for detect anomalies at the collective level. This probabilistically aggregates the contribution of the individual anomalies for detecting significantly anomalous groups. Because of collective anomalous score using an unsupervised manner, both unsupervised and supervised approaches can be used for scoring individual anomaly. The proposed model was evaluated on moving crane and fuel consumption dataset.

978-1-5386-7518-2/18/$31.00 ©2018 IEEE
DOI 10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00078

On acoustic novelty detection, [8] presented a novel approach based on non-linear predictive de-noising auto-encoders (DA) with LSTM for isolating abnormal acoustic signals. Long-Short Term Memory (LSTM) recurrent de-noising auto-encoders predict the auditory spectral features of the next short-term frame based on the previous frames. The reconstruction error between the input and the output of the auto-encoder play role as an activation signal to detect novel.

In [5], Malhotra et al. used LSTM network to detect anomaly on time series data. The stacked LSTM network trained on only normal data to generate predictor over some time steps. The distributed probability of error signal was estimated and was used to assess the likelihood of anomaly score.

We can solve anomaly by LSTM based classifier on two classes including normal label and anomalous one. In fact, recording anomalous data is costly or even make situation become dangerous. Clearly, the balance between training classes does not ensure. From the idea of [5] and window based method using adaptive error measure [9] for the particular application, in this work, we use LSTM for building a prediction model for anomaly system is the better choice. This anomaly function can be integrated to supervisory control and data acquisition (SCADA) [10] based factory system in which sensor data are gathered from industrial devices to analyze and discover information as well as visualization or alarm.

## III. Applications of LSTM based Anomaly Detection

### A. LSTM Based Anomaly Detection

We use LSTM-RNN as time series prediction model which feeds input data underlying time series form. In this study, the trained model is used to estimate the distribution of prediction error. The prediction error model verifies the likelihood of anomaly behavior. Our approach overcomes some limitations coming from many situations with not enough anomalous data. Considering a time-series data $X = \{x^{(1)}, x^{(2)}, \ldots, x^{(n)},\}$ with length of L where each sample $x^{(i)} \in R^m$ is an m-dimensional vector at time instance $i$. A prediction model $f$ learns parameters as $\emptyset$ annotation a to predict the next l values from N input samples. The equation of prediction problem is described as $\{\hat{x}^{k+N}, \ldots, \hat{x}^{k+N+l-1}\} = f(\{x^k, \ldots, x^{k+N-1}\}, \emptyset)$ (k is offset). In training phase, the prediction method adapts its parameters $\emptyset$, which become the characteristic of the normal training [5]. After having predictor, prediction error vector can be obtained via some measures [9] such as Relative Error (RE), the Average Relative Error (ARE), etc. Absolution of difference $e^i = |x^i - \hat{x}^i|$ is used to estimate of the distribution of errors fitting a parametric multivariate Gaussian distribution $N = (\mu, \sum)$. For any point $x^{(i)}$ the anomaly scores are expressed as likelihood $p^{(i)}$ of error vector $e^i$ the based on $N = N(\mu, \sum)$. If $p^{(i)} < \tau$ the observation is classified as "anomaly candidate", else "normal candidate" is assigned. The threshold $\tau$ can learned by maximizing $F_{\beta-score}$ (1) or pre-defined.

$$F_{\beta-score} = (1 + \beta)\,\frac{P \times R}{\beta^2 P + R} \qquad (1),$$

where, $P$ is precision, $R$ is recalled on the validate sequences in $v_{N2}$ and $v_A$.

Similar to [5], the normal data are divided into four sets namely $S_N$ (normal training), $v_{N1}$ (normal validation-1), $v_{N2}$ (normal validation-2) and $t_N$ (normal test) while the anomalous time-series can be divided into two sets $v_A$ (anomaly validation) and $t_A$ (anomaly test).

The architecture is composed of three stacked LSTM layer with the number of LSTM cells {64, 256, 100} respectively as shown in Figure 1, that followed by one fully connected layer. We also regularize between each layer with dropout operation 0.2 (20%). The normal training $S_N$ is used to learn the prediction models. The normal validation-1 $v_{N1}$ is used for early stopping during training phase. The error vector calculated on the normal validation-1 time series are used to estimate $\mu$ and $\sum$ of Normal distribution using Maximum Likelihood Estimation (MLE) algorithm. The threshold $\tau$ are chosen with maximum $F_{\beta-score}$. We consider anomaly candidates belong to positive class and normal candidates belong to negative class.

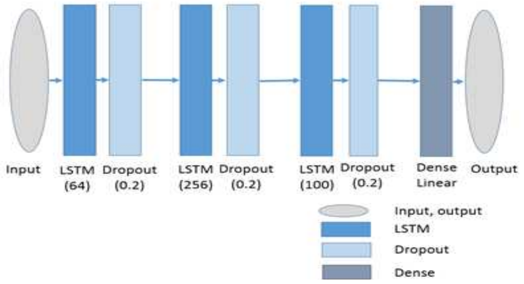


Figure 1. Architecture of stacked LSTM network

**Algorithm 1**: Anomaly Detection Algorithm

```
Input:
Size of slide window W is fixed
Size of time interval (ΔT)
Threshold for Anomaly candidate $\tau$
Threshold for Anomaly detection Th
while streaming do
  given $\hat{x}^i$ is prediction using LTSM model
   $x^i$ is accumulated frequency in interval (ΔT)
  calculate Absolute Error $e^i = |x^i - \hat{x}^i|$
  Check anomaly likelihood of $p^{(i)}$ :
    if anomaly likelihood $p^{(i)} < \tau$
      Assign anomaly candidate $C^i = 1$
    else
      Assign as normal candidate $C^i = 0$
    end

  Check window score for event detection:
    if sum $(C^i) > Th \times$ size of slide window W
       Anomaly detected
    end
end
```

Algorithm 1 is the anomaly detection algorithm, in which we use both prediction model and error-distribution model to detect temporal information of events. Before running real-time anomaly detection, LSTM-RNN based predictor model and error distribution model are trained as described. Informative data are verified by multi-word embedding CNN, these signals are approached with the window-based method to be transformed to time series data (sliding window = time interval (ΔT). Next, the predictor that already learnt normal data behavior could predict future signal using historical signal. To check how incoming signals fit to normal signals, absolute differences between actual signals and predicted signals are computed to estimate anomaly score through the error distribution model.

### B. Implementation Earthquake Detection System using LSTM based anomaly detection and Social Network

To detect the temporal occurrence of a considered event in distributed environment as social network, we implement a Hadoop-based [11] system as Figure 2. In this system, specific topics related informative messages are accumulated under form of time series data. The time series data is analyzed by Long Short-Term Memory (LSTM)-based event detection approach. This approach also performs well in many context applications where balanced data is not always available and real-time disaster event detection is required. We conduct experiments on earthquake dataset to obtain the time response.

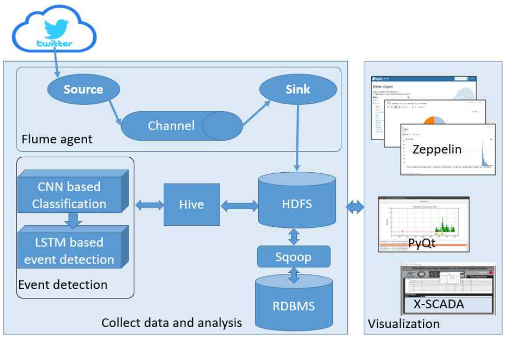


Figure 2. Earthquake detection system using anomaly detection

Social data from the user using the social network is collected via Flume tool [12]. Apache Flume is a tool owning ingestion mechanism for collecting aggregating and transporting large amounts of streaming data from various sources. Hive [13] is used as a data warehouse infrastructure to access data, QL in Hive facilitates in reading, writing and managing large data residing in distributed storage (HDFS). We proposed a Convolution Neural Network (CNN) [14] technique on Natural Langue Processing to determine informative data or sorting before moving to LSTM based anomaly detection. Hadoop streaming is a utility that comes with the Hadoop distribution, so we will use this aid to run executable or script as the mapper or reducer for performing classification and event detection. Sqoop [15] is a tool designed to transfer data between HDFS and relational database. The analyzed social data is then visualized by using Apache Zeppelin dash-board [16], or solution for SCADA [17] based system. Visualizations are very useful and informative for management in local and remote location as well.

## IV. Experiments

We conduct experiments on several real-world datasets including ECG signal dataset, power demand dataset and earthquake related social data. In Figure 3 and Figure 4, the top sub-figure shows the prediction signal (green dash) come from prediction model that trained on normal dataset, corresponding prediction error denoted as solid red line and below sub-figure is anomaly score and learned threshold. This threshold is used to assign if sample is anomaly candidate or not. Anomaly detection candidates are marked with orange color using window score as in Algorithm 1. The performance of anomaly detection is shown in Table I.

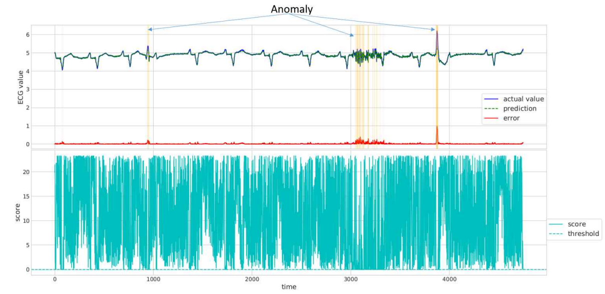


Figure 3. Anomaly detection on ECG signal

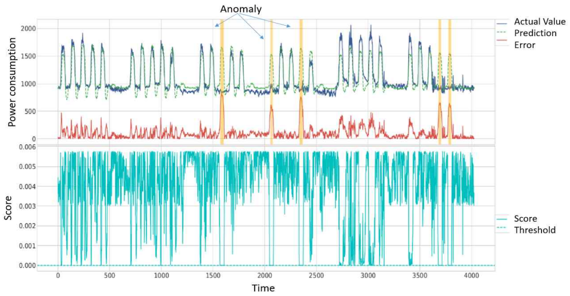
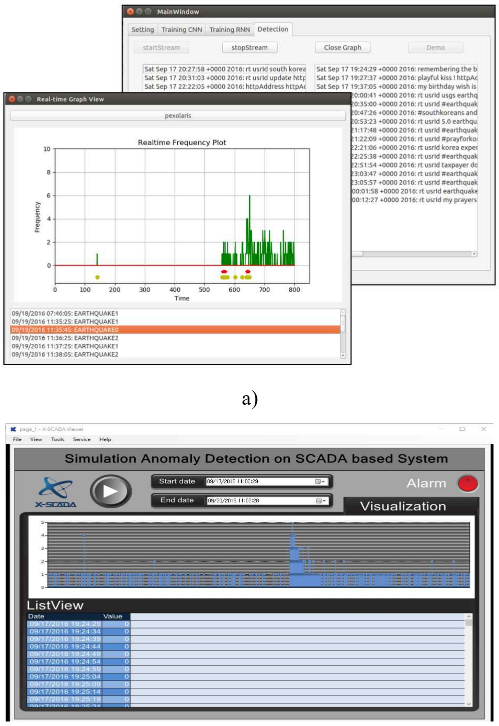


Figure 4. Anomaly detection on Power demand dataset

TABLE I. PERFORMANCE OF LSTM BASED ANOMALY DETECTION

| *Dataset* | *Precision* | *Recall* | *F-score* |
|---|---|---|---|
| ECG | 0.91 | 0.1 | 0.84 |
| Power demand data | 0.92 | 0.14 | 0.87 |



Sat Sep 17 20:27:58 +0000 2016: rt usrId south korea
Sat Sep 17 20:31:03 +0000 2016: rt usrId update http
Sat Sep 17 22:22:05 +0000 2016: httpAddress httpAc

Sat Sep 17 19:24:29 +0000 2016: remembering the b
Sat Sep 17 19:27:37 +0000 2016: playful kiss ! httpAc
Sat Sep 17 19:37:05 +0000 2016: my birthday wish is
20:00:41 +0000 2016: rt usrId usgs earthq
20:35:00 +0000 2016: rt usrId #earthquak
20:47:26 +0000 2016: #southkoreans and
20:53:23 +0000 2016: rt usrId 5.0 earthqu
21:17:48 +0000 2016: rt usrId #earthquak
21:22:09 +0000 2016: rt usrId #prayforko
22:21:06 +0000 2016: rt usrId korea exper
22:25:38 +0000 2016: rt usrId #earthquak
22:51:54 +0000 2016: rt usrId taxpayer dc
23:03:47 +0000 2016: rt usrId #earthquak
23:05:57 +0000 2016: rt usrId #earthquak
00:01:58 +0000 2016: rt usrId earthquake
00:12:27 +0000 2016: rt usrId my prayers

09/18/2016 07:46:05: EARTHQUAKE1
09/19/2016 11:35:25: EARTHQUAKE1
09/19/2016 11:35:45: EARTHQUAKE0
09/19/2016 11:36:25: EARTHQUAKE2
09/19/2016 11:37:25: EARTHQUAKE1
09/19/2016 11:38:05: EARTHQUAKE2

a)

Simulation Anomaly Detection on SCADA based System

Start date 09/17/2016 11:02:29
End date 09/20/2016 11:02:28

| Date | Value |
|---|---|
| 09/17/2016 19:24:29 | 0 |
| 09/17/2016 19:24:34 | 0 |
| 09/17/2016 19:24:39 | 0 |
| 09/17/2016 19:24:44 | 0 |
| 09/17/2016 19:24:49 | 0 |
| 09/17/2016 19:24:54 | 0 |
| 09/17/2016 19:24:59 | 0 |
| 09/17/2016 19:25:04 | 0 |
| 09/17/2016 19:25:09 | 0 |
| 09/17/2016 19:25:14 | 0 |
| 09/17/2016 19:25:19 | 0 |

b)

Figure 5. Visualization using: a) PyQt solution; b) SCADA solution

Figure 5 illustrates an application of anomaly detection in case of event detection on earthquake-related social data. Such figure contains the GUI of an implementation earthquake detection system using PyQt and visualization using X-SCADA. The average response time from our system is within one minute compare the actual earthquake as shown in Table II.

TABLE II. TIME RESPONSE OF EARTHQUAKE DETECTION SYSTEM

| **Actual earthquake**<br>*(https://earthquake.usgs.gov)* | | **LSTM based system** | |
|---|---|---|---|
| **Location** | **Time** | **Response** | **Delay** |
| M 4.6 - 10km SSW of Kyonju, South Korea 35.761°N 129.163°E | 2016-09-19 11:33:58 (UTC) | 11:35:25 | 87 sec |
| M 4.9 - 14km SW of Gyeongju, South Korea 35.743°N 129.106°E | 2016-09-12 10:44:33 (UTC) | 10:45:25 | 52sec |
| M 5.6 - 14km W of Nereju, Romania 45.714°N 26.528°E | 2016-12-27 23:20:55 (UTC) | 23:21:20 | 25sec |
| M 5.6 - 27km SW of Hawthorne, Nevada 38.376°N 118.899°W | 2016-12-28 08:18:00 (UTC) | 08:19:12 | 72sec |
| M 5.9 - 13km NE of Daigo, Japan 36.860°N 140.442°E | 2016-12-28 12:38:49 (UTC) | 12:39:19 | 30sec |
| M 5.3 - 18km SE of Volcano, Hawaii 19.330°N 155.121°W | 2017-06-08 17:01:19 (UTC) | 17:02:19 | 60sec |
| Average time response | | ~53sec | |

## V. CONCLUSION

Anomaly detection is an important problem that has been researched within diverse scope and application domains. In this paper, we discuss the anomalies of time series data. Especially, we explore LSTM based predictor operation on time series data to learn temporal signal feature for detecting anomaly pattern. This approach is very potential for applying in different time series data such as ECG, the sensor signal, etc. Moreover, we deployed the earthquake event detection based on big data framework for anomaly detection, the time response of detection is just within one to several minutes. This result is acceptable and useful for the meteorological department to forecast as well as publish the warning to residents when a disaster occurs.

## ACKNOWLEDGMENT

This work (Grants No. C0513295) was supported by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2017. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology (MEST) (Grant No. NRF-2017R1D1A1B03034429 and 2017R1D1A3B03034951). Also, this research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2018-2016-0-00314) supervised by the IITP(Institute for Information & communications Technology Promotion).

## REFERENCES

[1] V. Chandola, A. Banerjee and V. Kumar, “Anomaly detection: A survey,” ACM computing surveys (CSUR), vol.41, no.3, pp. 15, (2009).

[2] R. J. Hyndman, E. Wang, and N. Laptev, “Large-scale unusual time series detection,” in Data Mining Workshop (ICDMW), 2015 IEEE International Conference, pp. 1616-1619, (2015)

[3] C. Huang, G. Min, Y. Wu, Y. Ying, K. Pei, and Z. Xiang, "Time Series Anomaly Detection for Trustworthy Services in Cloud Computing Systems,” IEEE Transactions on Big Data, (2017).

[4] V. Q. Nguyen, H. J. Yang, K. B Kim, and A.R. Oh. “Real-Time Earthquake Detection Using Convolutional Neural Network and Social Data,” in Multimedia Big Data (BigMM), 2017 IEEE Third International Conference on, IEEE, pp. 154-157, (2017).

[5] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, “Long short term memory networks for anomaly detection in time series,” Proceedings (2015), pp. 89, (2015).

[6] S. Hochreiter and J. Urgen Schmidhuber, “Long Short-Term Memory,” Neural Computation, vol. 9, no. 8, pp. 1735-1780, (1997).

[7] T. Olsson, A. Holst, “A probabilistic approach to aggregating anomalies for unsupervised anomaly detection with industrial applications,” In: FLAIRS Conference, pp. 434–439, (2015).

[8] E. Marchi, F. Vesperini, F. Weninger, F. Eyben, S. Squartini, B. Schuller, “Non-linear prediction with lstm recurrent neural networks for acoustic novelty detection,” in 2015 International Joint Conference on Neural Networks (IJCNN), IEEE, pp. 1–7, (2015).

[9] L. Bontemps, V. L. Cao, J. McDermott and N.-A. Le-Khac, “Collective Anomaly Detection Based on Long Short-Term Memory Recurrent Neural Networks,” Springer International Publishing, pp. 141–152, (2016).

[10] S. D. Antón, D. Fraunholz, C. Lipps, F. Pohl, M. Zimmermann and H. D. Schotten, “Two decades of SCADA exploitation: A brief history,” in Application, Information and Network Security (AINS), 2017 IEEE Conference, IEEE, pp. 98-104, (2017).

[11] http://hadoop.apache.org

[12] http://flume.apache.org

[13] https://hive.apache.org

[14] V. Q. Nguyen, A. T. Nguyen, H. J. Yang, “Multi-word Embeddings CNN for Identifying Informative Messages” in the 5th International Conference on Big Data Applications and Services (5th BIGDAS) (2017).

[15] http://sqoop.apache.org

[16] https://zeppelin.apache.org

[17] http://www.xisom.com