

# Experiments on city train vibration anomaly detection using deep learning approaches

Taehee Kim<sup>1</sup>, Cheolwoo Ro<sup>2</sup>, Kiho Suh<sup>3</sup>

<sup>1</sup>School of Culture and Contents, Youngsan University, Korea

<sup>2</sup>Department Division of Computer Software Engineering, Silla University, Korea

<sup>3</sup>Artificial Intelligence Research Institute, Youngsan University, Korea

---

## Article Info

### Article history:

Received Jan 18, 2020

Revised Mar 22, 2020

Accepted Apr 6, 2020

### Keywords:

Accelerometer

Anomaly detection

Deep learning

Time-series data

Train vibration data

---

## ABSTRACT

Anomaly detection is widely in demand in the field where automated detection of anomalous conditions in many observation tasks. While conventional data science approaches have shown interesting results, deep learning approaches to anomaly detection problems reveal new perspectives of possibilities especially where massive amount of data need to be handled. We develop anomaly detection applications on city train vibration data using deep learning approaches. We carried out preliminary research on anomaly detection in general and applied our real world data to existing solutions. In this paper, we provide a survey on anomaly detection and analyse our results of experiments using deep learning approaches.

Copyright © 2020 Institute of Advanced Engineering and Science.  
All rights reserved.

---

## Corresponding Author:

Cheolwoo Ro,

Division of Computer Software Engineering,

Silla University, Busan 46958, South Korea.

Email: cwro@silla.ac.kr

---

## 1. INTRODUCTION

Anomaly detection takes place in many different kinds of data types spanning from symbolic data to numeric data, thus tackles many different technological problems. Since many anomaly detection tasks require handling of massive amount of data, deep learning approaches are being actively attempted and suggest positive results. However, anomaly detection for continuous time-series data, such as vibration signal, is less studied compared to anomaly detection of other non-temporal and symbolic types. Moreover, online real-time anomaly detection is technologically much demanding for real world applications, which would provide much benefit when solved.

We carry out a development of anomaly detection solution on city train operation by observing vibration data collected using an accelerometer that is mounted on a train. Our goal is to experiment feeding our data to typical applications as a feasibility study. In this paper, we first look at existing anomaly detection works using deep learning approaches. We then describe our experiments on the application of our data to two of those existing systems that are known to be appropriate for time-serial data. The result shows positive feasibility for anomaly detection on time-series vibration data using those deep learning approaches. However, in actual practical circumstances we often encounter situations where there are lack of or no anomalous data that makes robust classification very difficult or impossible. Without solving this problem, the capacity of anomaly detection using deep learning approaches will be very limited. We suggest that one class classifier should provide a strong clue for tackling this problem of lack of anomalous data, which we suggest as future research direction.

## 2. SURVEY ON ANOMALY DETECTION

In a large dataset, some rare cases of data that stand out among others, that is known to be anomaly conditions, are often required to be picked up. Anomaly could broadly fall into such categories as point, contextual, and collective anomalies with respect to the characteristics of data [1-6]. Point anomalies works on a set of classifiable entities so that anomalous values could be linearly separated from others. Contextual anomalies refer to conditional anomalies where an individual data instance is anomalous within a context, such as in a case where abrupt temperature change in a monthly room temperature pattern. Collective anomalies refer to a case when a set of related data instances is anomalous compared to the rest of the data such as in a case where missing heartbeat found among a series of other ones.

On the other hand, anomaly detection processing methods can be categorized into: Clustering-based technology [7]; Statistical techniques [8]; Information theory-based technology [9]; Spectral techniques; Robust PCA-based method [10], etc. Directly related to the topic of this paper, Sequential anomaly detection looks at data that are sequential, e.g., system call intrusion, climate data, DNA sequence, etc. Majority of conventional approaches are of the type of off-line while in the real world on-line application is often desired.

Recent applications of deep learning technologies to anomaly detection shows in good performance in some domains and the scope of application is rapidly expanding. Many anomaly detection applications for sequential data, such as time-series vibration data, are implemented using Long Short Term Memory (LSTM) networks (e.g. [11, 12]).

Anomaly detection from a time-series and streaming type of data can be applied to such tasks as dealing with sensor data analysis, diagnosing machineries, medical applications, network intrusion, documents, finance, etc [13]. In time-series data, anomaly detection tasks firmly rely on the characteristic of ‘the continuity in time.’ However, in reality, one will often find lack of anomalous data. For example, in such a case in a factory machine, it is very difficult to obtain anomalous data that would only available from a very limited rare case of malfunction of machines. Therefore, many try to solve this problem by modelling normal observation and finding out anomalies by figuring out distance from the normal observation.

Generative Adversarial Network(GAN) is recently often used for anomaly detection. Schlegl, et al. reported unsupervised anomaly detection using GAN on medical images [14]. They first train discriminator that will learn to discriminate real data from fake data. This trained discriminator is then used to find out anomalous data. Once learned, the discriminator will obtain manifold in the latent space that had been formed from given source data. In order to detect anomaly conditions, any incoming data that is sufficiently off the learned manifold can be regarded as anomalous data. GAN is also applied to noisy image training data in the case of hand written character recognition [15]. GAN anomaly detection is also applied to time-series data (eg. [16, 17]). We first use GAN for our first experiment to examine feasibility for our train data to be used for anomaly detection.

## 3. CITY TRAIN VIBRATION DATA

For our experiments, we use vibration data from an accelerometer that is mounted on a city train of Busan city, Republic of Korea. Experiment data were collected in an ASCII data form during a normal trip of a train on a railway. The data format is shown in Figure 1.

	A	B	C	D	E	F	G	H	I
	time	id	destination	last station	next station	section distance	train speed	Longitudinal vibration	Lateral vibration
1	시간	편성	행선지	출발역	도착역	역간거리	속도	상하진동	좌우진동
2	2017-03-02AM 6:50:39	10	ANPYUNG	youngsan	Dongbusa	3	9.34375	1.188872	0.107574
3	2017-03-02AM 6:50:39	10	ANPYUNG	youngsan	Dongbusa	3	9.34375	1.187877	0.107574
4	2017-03-02AM 6:50:39	10	ANPYUNG	youngsan	Dongbusa	3	9.34375	1.193848	0.10658
5	2017-03-02AM 6:50:39	10	ANPYUNG	youngsan	Dongbusa	3	9.34375	1.193848	0.110554
6	2017-03-02AM 6:50:39	10	ANPYUNG	youngsan	Dongbusa	3	9.34375	1.188872	0.11254
7	2017-03-02AM 6:50:39	10	ANPYUNG	youngsan	Dongbusa	3	9.34375	1.189867	0.10956
8	2017-03-02AM 6:50:39	10	ANPYUNG	youngsan	Dongbusa	3	9.34375	1.195838	0.111547
9	2017-03-02AM 6:50:39	10	ANPYUNG	youngsan	Dongbusa	3	9.34375	1.191857	0.111547
10	2017-03-02AM 6:50:39	10	ANPYUNG	youngsan	Dongbusa	3	9.34375	1.195838	0.108567
11	2017-03-02AM 6:50:39	10	ANPYUNG	youngsan	Dongbusa	3	9.34375	1.192853	0.099627
12	2017-03-02AM 6:50:39	10	ANPYUNG	youngsan	Dongbusa	3	9.34375	1.194843	0.10956
13	2017-03-02AM 6:50:39	10	ANPYUNG	youngsan	Dongbusa	3	9.34375	1.196833	0.118488
14	2017-03-02AM 6:50:39	10	ANPYUNG	youngsan	Dongbusa	3	9.34375	1.203787	0.11552
15	2017-03-02AM 6:50:39	10	ANPYUNG	youngsan	Dongbusa	3	9.34375	1.200814	0.112454

Figure 1. Raw vibration data format

The sampling rate of the data is approximately 33.3Hz. We use the longitudinal vibration data and lateral vibration data for our experiment. Since the train travels in various speeds from a station to another station, we took a section of data that are sampled during a constant speed of approximately 70Km/h which is a typical top speed of train between stations.

Figure 2 left shows the format of longitudinal(X-column) and lateral(Y-Column) vibration data extracted from the raw data table to feed into our network as input data. From the raw data, Figure 2 right shows longitudinal and lateral vibration waveform along time of trip, in which the vibrations are clearly shown.

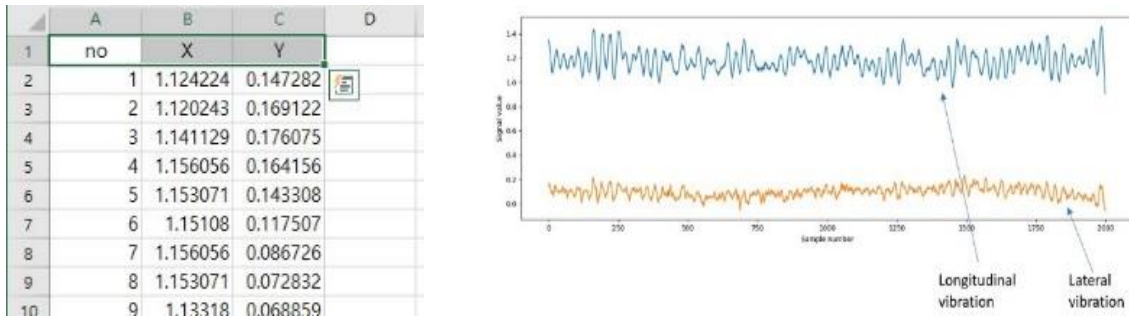


Figure 2. longitudinal (X) and lateral (Y) vibration data as input data (left), waveform (right)

We carried out spectral density analysis on the train vibration data as shown in Figure 3. The waveform shown on the top shows vertical vibration middle, horizontal vibration and the last, root-mean-square of vertical and horizontal vibrations. The three bottom graphs show the power spectrum density of the waveforms in an order from left to right. They show that large portions of frequency components of the vibrations are located below approximately 20Hz. This frequency characteristic will probably change according to train speed while still reflecting some meaningful information about the train operation including the condition of the railroad, leaving out such train-speed-dependent analysis left for our future work.

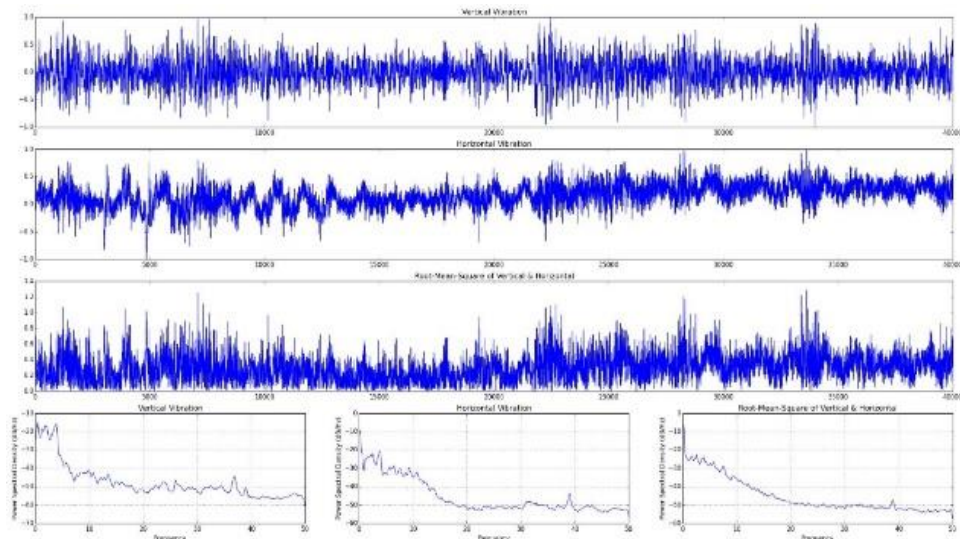


Figure 3. Power spectral density analysis of vibration signals

#### 4. TRAINING DATA USING GAN

We first carried out an experiment to examine whether a deep learning system could learn our train data waveform pattern. Once a system learns typical vibration waveform patterns then anomaly detection could be realized by calculating distance from learned form and the incoming signal.

We chose infoGAN by Chen et al. for our first experiment [18]. InfoGAN is a generative adversarial network that learns disentangled representations in a completely unsupervised manner. Thus such classification ability could be used to learn typical pattern of time-series vibration data. We adopted an InfoGAN implementation [19] for time-series data and feed our data to execute learning process. The result is shown in Figure 4, where the cross entropy value(loss/ce) drops significantly that indicates that learning process converges to a certain direction.

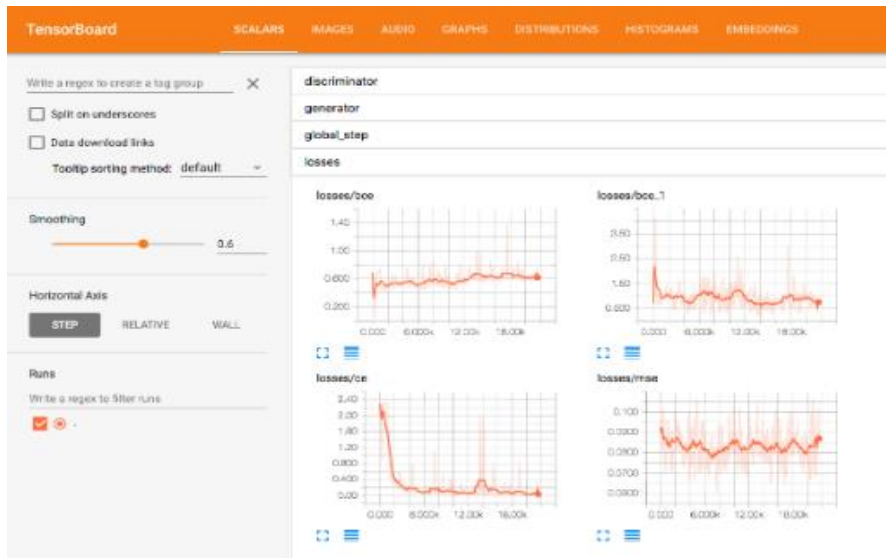


Figure 4. Result of learning process of vibration data using time-series implementation of InfoGAN

### 5. TRAINING DATA USING LSTM

In this section, we describe our experiment on our city train vibration data anomaly detection based on an LSTM implementation anomaly detection [11, 20]. Anomaly detection using machine learning can be divided into two kinds of approaches: 1) both normal data and anomalous are available; 2) only normal data are available. The case 2) about is difficult since the network never saw anomalous data. However, a network that could learn to predict next coming sequence like LSTM could be used to tackle this problem. An LSTM network learns off-line the pattern of the normal healthy data and be able to predict next coming sequence given the previous data. Then the system compares the incoming real data with the predicted data to figure out the difference. If the difference is higher than a certain threshold, the system reports the incoming data as anomaly. Since the vibration data that we have available are only normal data, therefore ‘leaning to predict and comparing’ as explained above would be an appropriate approach. An LSTM is considered to be a suitable method for this purpose. This approach could be more articulated to be able to work online for real-time applications. We describe our feasibility study for train vibration data below using an LSTM model. The structure of layered LSTM that we adopted for our experiment is shown in Figure 5 [11]. This model is based on typical LSTM structure and we used a layered LSTM structure.

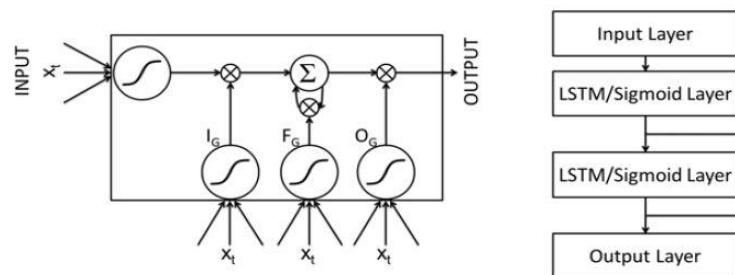


Figure 5. Typical LSTM structure(left) and Layered LSTM structure (right) [11]

The way the LSTM cells are connected is in a form of fully connected stacked LSTM as shown in Figure 6 along input and output sequences. In order for the LSTM network to learn the sequence of the vibration data, the model used parameters shown in Table 1.

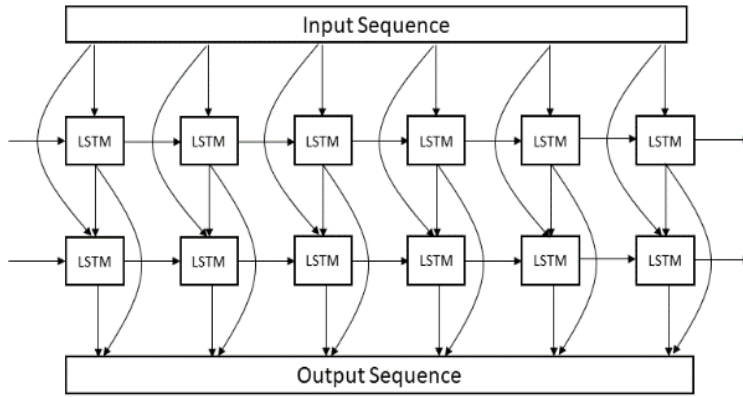


Table 1. Parameters used for LSTM model training

Parameter	Value
Number of input features	128
Number of hidden layers per layer	128
Number of layers	2
Number of epoch	400
Initial learning rate	0.0002
Batch size	64
Sequence length	50
dropout	0.2

Figure 6. Fully connected stacked LSTM model

Using the parameters above, we trained the LSTM model. Figure 7. shows the signal that our LSTM model generates along the learning span of 400 epochs. According to the waveform prediction results shown in Figure 7, along the progress of learning, our LSTM model learns to generate sophisticated waveforms gradually resembling the train data. However, even in the generated waveform in the last graph resembles the signal in terms of global periodic variation, it misses out details of textures. We leave further tuning of the model and parameters for future work when sensing device is re-engineered and much more data are available.

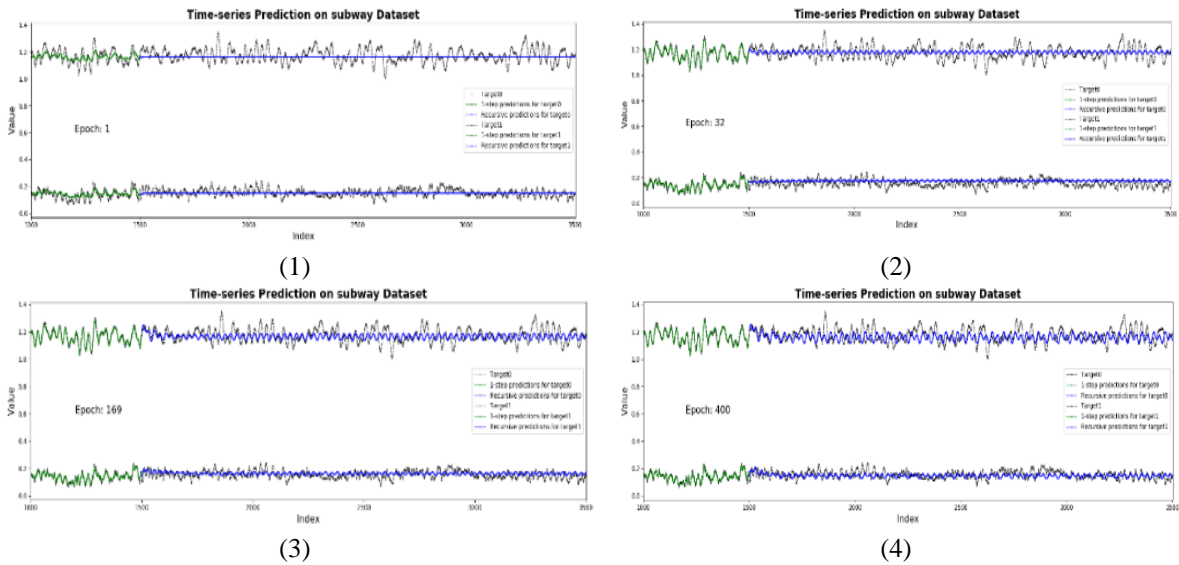


Figure 7. The waveforms that the LSTM model predicts and generates (from top to bottom along epochs)

We next carry out comparison between real input signal and the predicted signal. The measure that indicates the amount of difference between input signal and predicted (or desired) signal is referred to as ‘Anomaly Score.’ In our experiment, we take the Anomaly score by Multivariate Gaussian Distribution shown below [11].

$$Anomaly\ score = (e - \mu)\Sigma^{-1}(e - \mu)^T \tag{1}$$

The comparison result is shown in Figure 8. The overlapped signal area on the top of in the figure shows the signals overlapped, the input original normal signal and the generated predicted signal. The red graph in the bottom of the graphs show the results of computed anomaly score. The smaller amplitude indicates the higher precision the prediction is. There are many very high peaks in the anomaly score and this means the model's poor performance such that the predicted signal keeps significant distance from the original signal. This means that the system would most likely report such peaks as anomaly conditions. We will work on improving the performance when we obtain more vibration data in the future.

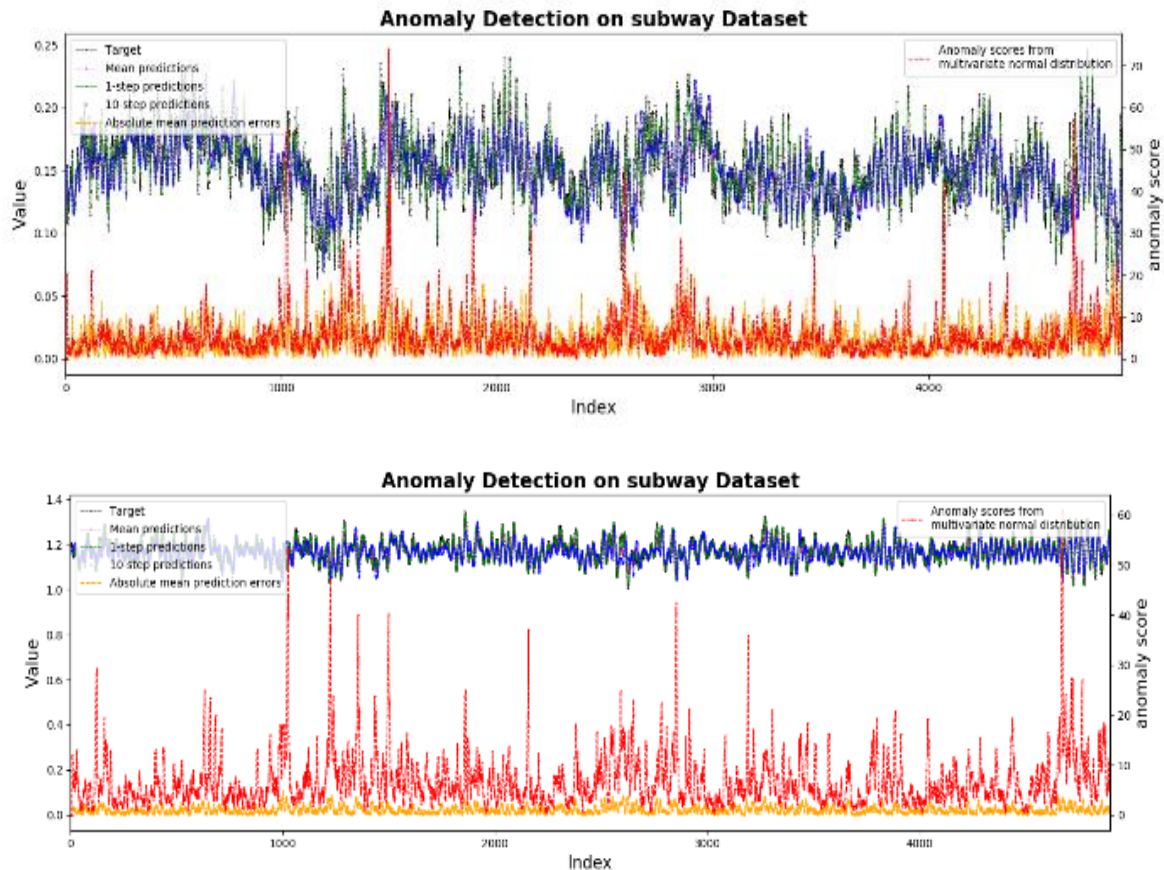


Figure 8. The overlap of input and output data (blue part) and the following anomaly score (red part), (top: vertical vibration, bottom: horizontal vibration)

So far, we described our experiments using an LSTM model that predicts next steps of signals given the current signals. We then showed our results of computation of anomaly scores. This method could be effective when one lacks anomalous data or none.

However, we criticise that this method must be limited in situations where natural signal data are unique or show a narrow scope of variations. This means that if the vibration characteristic would vary depending on locations of railroad and the kinds of trains, such method will be too limited. Thus, we introduce a classification and clustering method that would overcome the problem of varying signal characteristics. The procedure of an anomaly detection by classification and clustering is as follows:

- a) Taking samples of signal segments using overlapped windows by half of each segment to prevent feature loss;
- b) Carrying out classification and clustering by unsupervised learning on the sampled segments, then the resulting sets of segments will be classified sets in terms of known characteristics of features;
- c) Given the learned clusters, actual incoming data is fed by taking samples of overlapped windowed signals then locate those samples in the cluster space for comparison with the learned clusters;
- d) If the incoming data fall onto an existing cluster space then the signal is classified as normal, otherwise the incoming signal is classified as anomaly.

Classification and clustering method described above is expected to be an effective anomaly detection in the real application to the train operation. We implemented clustering of the train vibration data mentioned above using the code provided in [21]. Total 40,000 numbers of quantized samples of vertical and horizontal vibrations are processed. In the step of data pre-processing we first normalize the data and takes segments of 100 data samples which thus makes 4,000 numbers of segments. We chose only 100 segments from this pool for simplicity. Similarities are computed between every pair of data segments. We tried four different kinds of similarity computation methods as follows:

- a) Euclidean Distance: Similarities by geometrical distance that is widely used for standard data;
- b) Root Mean Square Error(RMSE): Standard deviation generalized distance;
- c) Dynamic Time Warping(DTW): Calculating optimal matching pairs of sequences in time that allows warping in time space;
- d) Longest Common Subsequence(LCS): Time-series similarity computation by finding the longest progression, which is robust to noise.

Our experiment is limited to make six clusters given 100 segments of the input signal. The experiment showed that LCS method provides the best result. Roughly clustered data can be observed in the Principal Component Analysis(PCA) calculated by the LCS method. Figure 9 shows the codes and result of PCA on the clustering using the LCS method.

We extract intersecting features given mean, standard deviation, and slope based on the result computed by LCS method on the signal samples. Figure 10. shows the result. The bottom left and bottom centre that is the horizontal vibration(sensor2) shows clusters are roughly formed.

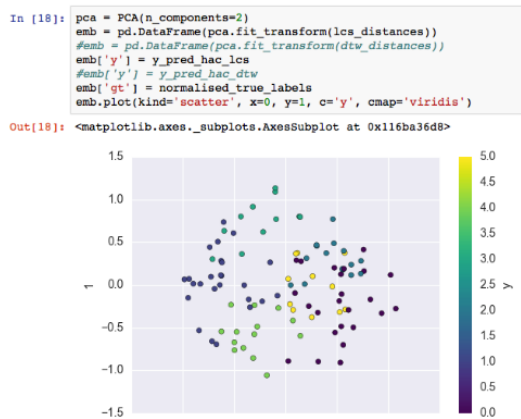


Figure 9. Examples of sampling of windowed segmentation of source train vibration data

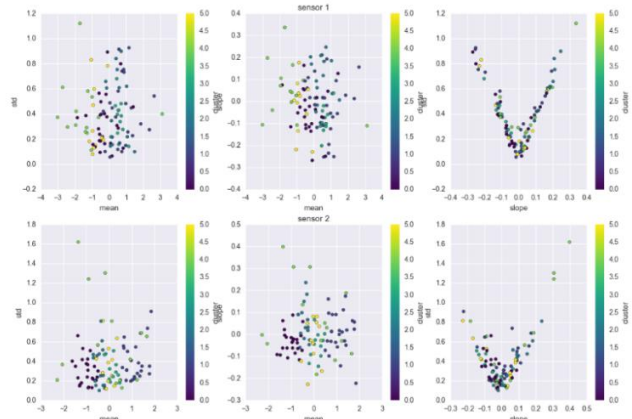


Figure 10. Cluster graphs obtained by calculation across the reference of mean, standard deviation, and slop

The result so far shows positive feasibility of anomaly detection of our train vibration data by classification and clustering. We need to increase the size of data for more generalized results. The number of clusters also need to be changed during clustering experiments to find an optimal number of clusters. We chose 100 data samples as the size of the sampling window. This data sample number needs to be adjusted for optimal classification and clustering. Different data sets and different window size might also influence anomaly detection algorithm to be chosen.

## 6. FUTURE RESEARCH AND CONCLUSION

Our future works need be carried out in two directions: 1) real-time adaptive anomaly detection; and 2) anomaly detection using one-class classifier. First, real-time adaptive anomaly detection is crucial because our trains will experience various railway conditions [22]. The physical construction of trains might also be variable. Thus, it is not surprising that even normal signals could vary depending on locations, time, and trains. Therefore, taking a development direction in the adaptive, real-time, and online learning is desired. Second, since it is difficult to obtain anomalous data in our case, developing a robust method that uses only normal data to learn effectively and distinguish any unknown anomaly signal from normal ones. In this perspective, research on one-class classifier will be very helpful. Sabokrou, et al. reported a successful end-to-end network for learning one-class classifier. By preserving both generator and discriminator, they were

able to build a robust classifier learning network. Other approaches to one-class classifier includes [23-26]. Other areas of artificial intelligence need to be considered to tackle anomaly detection problems, for example the feature extraction in Natural Language Processing [27].

We described in this paper our experiments on anomaly detection using our time-series data collected from city train vibration during normal operations. We first examined feasibility for training of data using GAN and obtained positive results. We then applied our data to a layered LSTM. The system learns to predict vibration signals although the anomaly scores were not satisfactory. Our development of new device to collect vibration data is underway. Once the device is available we will carry out an upgraded development using new data sets. We have shown in this paper classification and clustering experiments on our vibration data. We obtained a satisfactory degree of trust of feasibility so that future refinement and introducing alternative methods such as one-class classifier would be expected to be practically feasible in real-time anomaly detection on time-series vibration data.

## ACKNOWLEDGEMENT

This work was supported by Youngsan University Research Fund of 2019.

## REFERENCE

- [1] R. Chalapathy and S. Chawla, "Deep Learning for Anomaly Detection: A Survey," arXiv:1901.03407v2, 2019.
- [2] H. Wang, M.J. Bah, and M. Hammad, "Progress in Outlier Detection Techniques: A Survey," *IEEE Access*, vol. 7, pp. 107964-108000, August, 2019.
- [3] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection," *ACM Computing Surveys*, vol. 41, no. 3, pp. 1-58, Jul. 2009.
- [4] M. Siti and K. Fitra, "Analysis of DBSCAN and K-means Algorithm for Evaluating Outlier on RFM Model of Customer Behavior," *Telecommunication, Computing, Electronics and Control*, vol. 17, no. 1, pp. 110-117, February 2019.
- [5] C.G., Madhu and R.G.M. Sreerama, "Framework for Contextual Outlier Identification using Multivariate Analysis approach and Unsupervised Learning," *International Journal of Electrical and Computer Engineering*, vol. 8, no. 2, pp. 1092-1101, April 2018.
- [6] R. Maryem, et al, "Forecasting Financial Budget Time Series: AIRMA Random Walk vs. LSTM Neural Network," *International Journal of Artificial Intelligence*, vol 8, no. 4, pp. 317-327, 2019.
- [7] E. Eskin et al., "A Geometric Framework for Unsupervised Anomaly Detection: Detecting Intrusions in Unlabeled Data", 2002.
- [8] K. Yamanishi, "On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms," *KDD*, 2000.
- [9] W. Lee, et al, "Information-Theoretic Measures for Anomaly Detection," *IEEE Symposium on Security*, 2001.
- [10] R. Shyu, S. Chen, K. Sarinnapakorn, L. Chang, "A novel anomaly detection scheme based on principal component classifier," In *Proceedings of the IEEE Foundations and New Directions of Data Mining Workshop*, 2003.
- [11] P. Malhotra, L Vig, G. Shroff, P. Agarwal, "Long Short Term Memory Networks for Anomaly Detection in Time Series", *Proc. Of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pp. 89-94, 2015.
- [12] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, G. Shroff, "LSTM-based Encoder-Decoder for Multi-sensor Anomaly Detection", *Anomaly Detection Workshop, ICML*, 2016.
- [13] J. Gamboa, "Deep Learning for Time-Series Analysis," *arXiv preprint arXiv:1701.01887v1*, 2017.
- [14] T. Schlegl, et al, "Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery," *Proceedings of IPMI*, 2017.
- [15] M. Kinura, T. Yanagihara, "Anomaly Detection Using GANs for Visual Inspection in Noisy Training Data," *arXiv preprint arXiv:1807.01136v2* (2018).
- [16] S. Hyland, C. Esteban, G. Rätsch, "Real-Valued (Medical) Time Series Generation with Recurrent Conditional GANs," *arXiv preprint arXiv:1706.02633v2* (2017).
- [17] D. Li, D. Chen, J. Goh, S. Ng, "Anomaly Detection with Generative Adversarial Networks for Multivariate Time Series," *7<sup>th</sup> International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications on the ACM Knowledge Discovery and Data Mining conference*, *arXiv preprint arXiv:1809.04758v2*, 2018.
- [18] X. Chen, Y. Duan, R. Houthoofd, J. Schulman, H Sutskever, P. Abbeel, "InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets," *arXiv preprint arXiv:1606.03657v1*, 2016.
- [19] [https://github.com/buriburisuri/timeseries\\_gan](https://github.com/buriburisuri/timeseries_gan)
- [20] <https://github.com/chickenbestlover/RNN-Time-series-Anomaly-Detection>
- [21] <https://github.com/effa/time-series-clustering>
- [22] S. Ahmad, S. Purdy, "Real-Time Anomaly Detection for Streaming Analytics," *arXiv preprint arXiv:1607.02480v1*, 2016.
- [23] S. Pidhorskyi, R. Almoosen, D. Adjero, G. Doretto, "Generative Probabilistic Novelty Detection with Adversarial Autoencoders," *arXiv preprint arXiv:1807.02588v2*, 2018.



- [24] S.R. Arashloo, J. Kittler, "One-Class Kernel Spectral Regression for Outlier Detection," *arXiv preprint arXiv:1807.01085v5*, 2018.
- [25] P. Perera and V. Patel, "Learning Deep Features for One-Class Classification," *arXiv preprint arXiv:1801.0535v2*, 2019.
- [26] L. Ruff, et al., "Deep One-Class Classification," In *International Conference on Machine Learning (ICML)*, 2018.
- [27] X. Wang, and H.C. Kim, "New Feature Selection Method for Text Categorization," *Journal of Information and Communication Convergence Engineering*, vol. 15, no. 1, pp. 53-61, Mar 2017.

## BIBLIOGRAPHIES OF AUTHORS



**Taehee Kim.** Taehee Kim received his Ph.D. degree in 1996 at the University of Edinburgh in Artificial Intelligence. He joined Korea Institute of Science and Technology Information for three years before he took a professorship at Youngsan University until now. Prof. Kim has recently worked as a senior researcher at the Artificial Intelligence Research Institute in Pangyo, Korea for about a year in 2017 where he developed deep learning systems for human body motion learning. His primary interest is in learning motions given condition in such as in dancing to music.



**Cheolwoo Ro.** Cheolwoo Ro is working as Professor in Computer Software Engineering Division of Silla University, KOREA. He worked and developed TDX and CDMA as a senior researcher for 10 years at ETRI. His area of research and specialisation are modeling and performance analysis of system, SDN, system prediction using machine learning, and cloud computing.



**Kiho Suh.** Kiho Suh is working as research engineer in Intelligence Design based in South Korea. His area of research includes generative models, reinforcement learning, and graph neural networks. His previous experience includes Artificial Intelligence Research Institute, IBM, and LG.