ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20221227094126

Preliminary communication
Received: 2022-12-27, Accepted: 2022-12-31

A Study on Verification of CCTV Image Data through Unsupervised Learning Model of Deep Learning

Yangsun Lee

Abstract: Abnormal behavior is called an abnormal behavior that deviates from the same normal standard as the average. The installation of public CCTVs to prevent crimes is increasing, but the crime rate is rather increasing recently. In line with this situation, artificial intelligence research using deep learning that automatically finds abnormal behavior in CCTV is increasing. Deep learning is a type of artificial intelligence designed based on artificial neural networks, and the quality of learning data is important for high accuracy in the development of artificial intelligence through deep learning. This paper verifies whether learning data for abnormal behavior detection is suitable as learning data which is being constructed using an MPED-RNN model for binary classification to determine whether there is an abnormal behavior by frame using skeleton data of a person based on an autoencoder. As a result of the experiment, the unsupervised learning-based MPED-RNN model used in this paper is not suitable for verifying images with a similar number of frames with and without abnormal behavior, such as the corresponding data, and it is judged that appropriate results can be derived only when verified with a supervised learning-based model.

Keywords: abnormal behavior; abnormal behavior detection; artificial neural network; deep learning; MPED-RNN; unsupervised learning model

1 INTRODUCTION

Abnormal behavior is called an abnormal behavior that deviates from the same normal standard as the average. The installation of public CCTVs (CCTV - close-circuit television) to prevent crimes is increasing, but the crime rate is rather increasing recently. In line with this situation, artificial intelligence research using deep learning that automatically finds abnormal behavior in CCTV is increasing. Deep learning is a type of artificial intelligence designed based on artificial neural networks, and the quality of learning data is important for high accuracy in the development of artificial intelligence through deep learning [4, 5, 12, 15].

This paper verifies whether learning data for abnormal behavior detection is suitable as learning data which is being constructed using an MPED-RNN model for binary classification to determine whether there is an abnormal behavior by frame using skeleton data of a person based on an autoencoder. As a result of the experiment, the unsupervised learning-based MPED-RNN model used in this paper is not suitable for verifying images with a similar number of frames with and without abnormal behavior, such as the corresponding data, and it is judged that appropriate results can be derived only when verified with a supervised learning-based model [8, 10, 11].

2 RELATED STUDIES

2.1 Autoencoder

The autoencoder is an unsupervised deep learning, consisting of two structures: encoder and decoder. The autoencoder learns through the process of minimizing the difference between the original data and the restored data after encoding the input data and restoring it again through the decoder [1, 2, 13]. Fig. 1 shows the structure of the autoencoder.

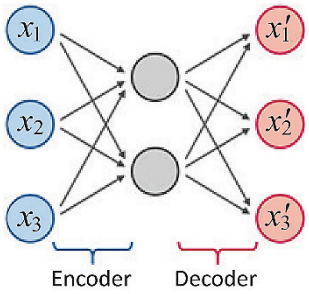

Figure 1 Autoencoder

2.2 RNN (Recurrent Neural Network)

RNN is a neural network that continuously uses the information of the previous step while repeating itself. We use historical information as a loop structure to improve the performance of neural networks on current inputs [6, 7, 9]. Fig. 2 shows the RNN structure.

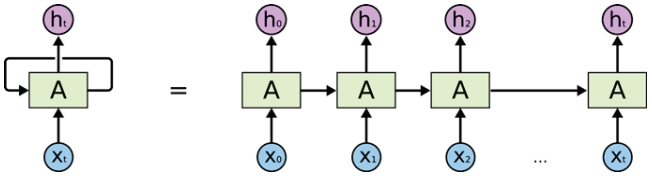

Figure 2 RNN structure

2.3 MPED-RNN Model

The MPED-RNN model is an autoencoder-based anomaly detection model with skeleton data input. The encoder-decoder has a repeated structure and features a temporal and spatial pattern of the skeleton trajectory [14, 16, 19, 20, 21]. Fig. 3 shows the structure of the MPED-RNN model.

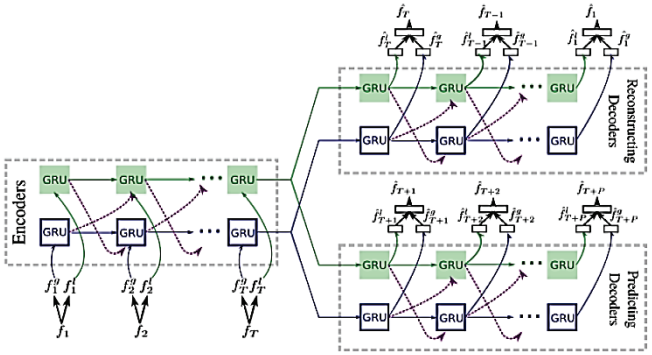

Figure 3 MPED-RNN model structure

The MPED-RNN Model learns by dividing the model's input, skeleton data, into global body movement which is information about large movements with little shape, size, and deformation, and local body posture which is information about fine movements such as internal deformation of skeleton movements. When an irregular pattern occurs during learning, the frame in which the pattern occurs is classified as an abnormal behavior.

2.4 AUROC (Area under the ROC Curve)

In the MPED-RNN model, the default output evaluation index is AUROC. AUROC represents the area under the ROC curve, a graph that corresponds to the vertical and horizontal axes of the True Positive Rate (TPR), which is the ratio that accurately predicted the normal, and false positive rate (FPR), which is the ratio that incorrectly predicted the normal. Fig. 4 shows the ROC curve. [3, 6, 7].

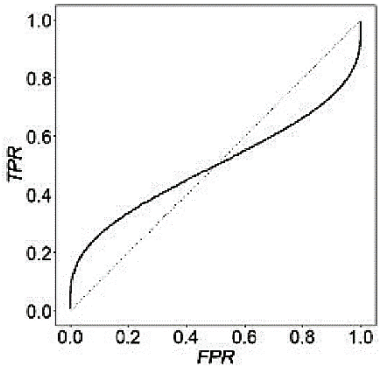

Figure 4 ROC curve

2.5 Unsupervised Learning

As a kind of machine learning, it falls into the category of problems that determine how data is composed. Unlike supervised learning or reinforcement learning, this method is not given a target value for the input [2, 11, 18, 22, 24]. Fig. 5 shows the unsupervised learning process.

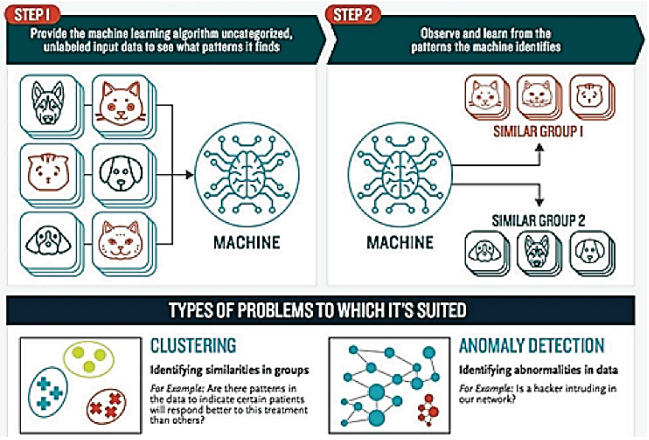

Figure 5 Unsupervised learning process

3 VERIFICATION OF LEARNING DATA WITH MPED-RNN

MPED-RNN, an anomaly detection model, uses skeleton data for each person in the video as learning data, and evaluation is conducted using skeleton data and a frame-level mask that expresses which frame the anomaly behavior occurred. Fig. 6 is a data verification system for learning conducted in this paper.

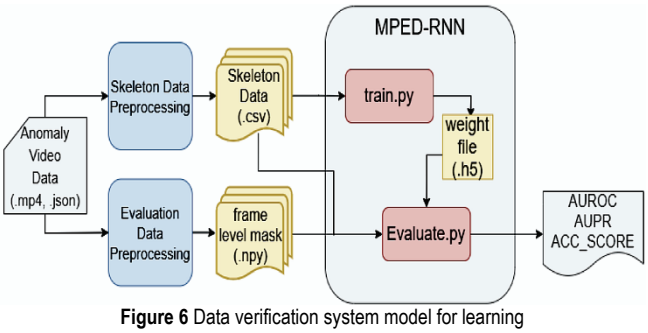

Figure 6 Data verification system model for learning

3.1 Preprocessing of Skeleton Data

In order to preprocess image data among learning data as skeleton information used as learning data in the MPED-RNN model, skeleton data was first extracted from the image. The extracted skeleton data is a JSON (JavaScript Object Notation) file with a frame number, a person number, and a person's joint coordinates as shown in Fig. 7. Fig. 7 shows the extracted skeleton data.


```
"frame_index": "0",
"objects": [],
"objects_changed": "no",
"persons": [
    {
        "index": "0",
        "keypoints": [
            "0.000000,0.000000",
            "877.502808,380.112000",
            "948.150146,380.220673",
            "974.620850,503.878204",
```


Figure 7 Extracted skeleton data

The input data of the MPED-RNN model is a csv file representing the trajectory of skeleton data for each person. Therefore, the extracted skeleton data were used to divide the files by person, and the frame in which the person appeared in each file and the coordinates of the 17 joints observed in the frame were stored in the form of a csv file. Fig. 8 shows a preprocessed skeleton file.

framenumber	point1_x	point1_y	point2_x	point2_y	point3_x	point3_y
2	1278	277	1366	318	1366	318
3	1272	277	1366	318	1366	318
4	1266	274	1366	318	1366	318
5	1260	271	1366	319	1366	319
6	1260	271	1363	318	1363	318
...	...	...	...	...	...	...
3597	1166	763	1186	798	1186	798
3598	1166	763	1184	798	1184	798
3599	1166	763	1186	798	1186	798
3600	1166	763	1186	798	1186	798
3601	1166	763	1186	798	1186	798

Figure 8 Preprocessed skeleton data

3.2 Preprocessing Evaluation Data

The frame_level_mask file used to evaluate abnormal behavior classification in the MPED-RNN model is a binary file that expresses 0 and 1 with and without abnormal behavior by frame. In order to produce a frame_level_mask of learning data, a start_frame_index in which abnormal behavior begins and an ends_frame_index in which the abnormal behavior ends was extracted from the annotation file provided with the learning image data to produce a binary file with the information. Fig. 9 shows a data annotation file for learning, and Fig. 10 shows a generated frame-level mask.


```
"block_detail": "A21",
"start_time": "00:00:59.400",
"main_object": "Ob0",
"block_type": "action",
"end_time": "00:01:53.157",
"block_index": "2",
"start_frame_index": "1800",
"end_frame_index": "3429",
"num_persons": "1"
```


Figure 9 Data annotation file for learning

3.3 Learning

Learning was conducted using the generated skeleton data as an input to an unsupervised learning model. 454 images out of a total of 572 images were used as learning data. Fig. 11 shows part of the learning data.

Fig. 12 shows the learning settings. Epoch proceeded to 20. One epoch refers to the forward pass/backward pass process for the entire data in an artificial neural network. In other words, in the model, a total of 20 learning are conducted on the entire data.

In the learning process, if the video is put in, it is separated for each frame and clustering is performed by grouping frames with similar skeleton values. If most frames have similar skeleton values, but there are frames with different skeleton values than other frames, we classify the frames as abnormal behavior and proceed with learning.

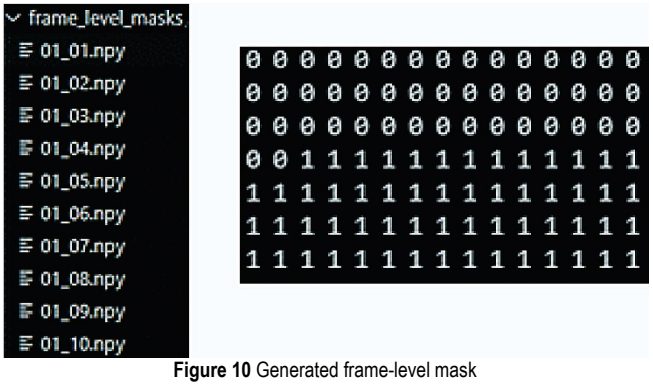

Figure 10 Generated frame-level mask

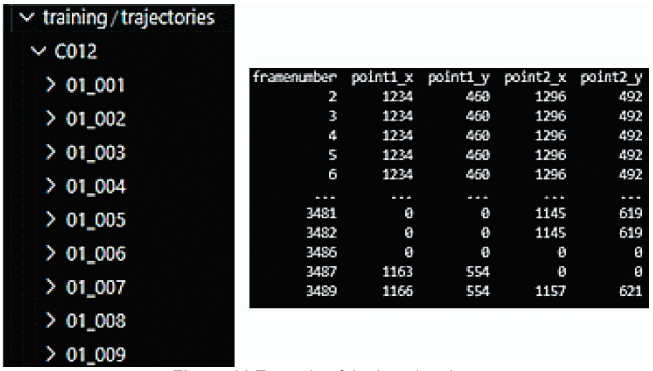

Figure 11 Example of the learning data

```
input_length 12
global_input_dim 4
local_input_dim 34
reconstruction_length 12
prediction_length 6
global_hidden_dims 8
local_hidden_dims 16
extra_hidden_dims
output_activation linear
reconstruct_reverse True
reconstruct_original_data True
multiple_outputs True
multiple_outputs_before_concatenation True
cell_type gru
optimiser adam
learning_rate 0.001
loss mse
```

Figure 12 Learning settings

3.4 Evaluation

Since the learning model of this paper is unsupervised learning, the intermediate result before AUROC output is predicted by frame. TPR(true positive rate) and FPR(false positive rate) were calculated based on an arbitrary classification point with the predicted value, and AUROC, the lower area of the green ROC curve, was output.

AUPR, the lower area of the precision-replay graph, was output with precision, which is the actual normal ratio among frames predicted to be normal, and reproduction, which is the

normal ratio among frames predicted to be normal. Fig. 13 shows an example of a predicted value for each frame and an output result.

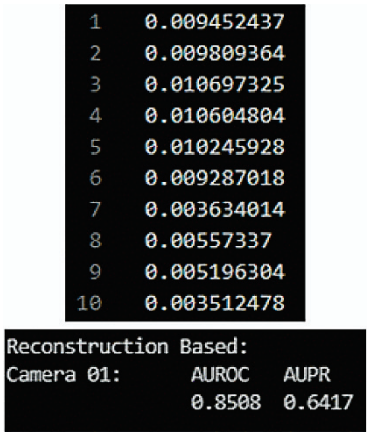

Figure 13 Predicted value for each frame and an output result

The model's evaluation method uses learned weights to quantify abnormal behavior for each frame, and then classify abnormal behavior using clustered values based on arbitrarily determined values. It is a method of extracting accuracy by comparing the classified binary file with the frame level mask, which is an answer binary file input by the user.

4 EXPERIMENTAL RESULTS AND ANALYSIS

The verification of learning data was conducted in a Geforce RTX 2080 environment with about 11 GB of memory. Fig. 14 shows a data source image for learning.

Figure 14 Data source image for learning

4.1 AUROC Results

Fig. 15 shows the verification results of learning data. The model's input data uses 17 joint coordinates, but 13 of the joint coordinates extracted from learning data were available. Therefore, as if the four uninputed joint coordinates were not observed, it is the result of entering the joint coordinates at the location most similar to the result of the learning by entering '0'.

Looking at the AUROC value, it was about 0.6. The value of AUROC in a binary classifier is from 0.5 to 1.0, and the binary classifier must have at least an AUROC value of 0.8 in order for it to be useful.

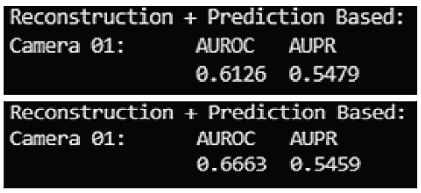

Figure 15 Verification results of learning data

4.2 Added Evaluation Index

In addition to AUROC and AUPR, which are essentially provided evaluation indicators in the MPED-RNN model, the numerical values of classification points were changed to find the optimal classification points for the specific section with the highest accuracy, and output the maximum accuracy. As can be seen from Fig. 16, it can be seen that the AUROC of the learning data is about 0.66 and the optimal classification point accuracy is about 0.65. The learning data shows lower accuracy than the HR-Avenue data used as the performance evaluation of the model.

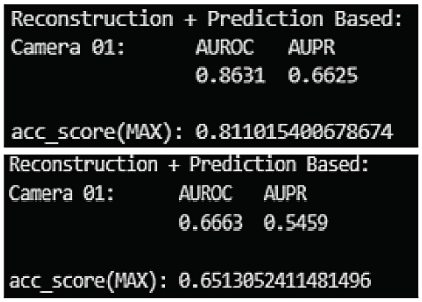

Figure 16 HR-Avenue and E2ON data verification results

However, this is not the low quality of the image, but in the case of the MPED-RNN model, which finds irregular patterns in the image and classifies them as abnormal behavior, since it is based on an autoencoder that performs unsupervised learning, if the number of frames in which abnormal behavior occurs in the learning data is similar to the number of frames in which abnormal behavior does not occur, the accuracy is lowered, and the accuracy of the learning data is lowered.

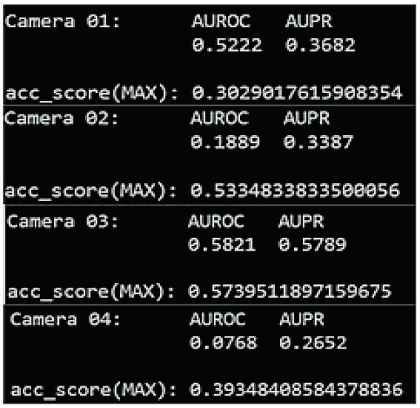

Figure 17 Abnormal behavior evaluation results

4.3 Abnormal Behavior Evaluation Results

Fig. 17 shows the evaluation results of child abuse, home invasion, theft, and vehicle theft learning data, respectively.

It was confirmed that the evaluation results were very low in the case of residential intrusion and vehicle theft abnormal behavior with more than 50% of the total number of frames of the learning video. Therefore, it is judged that the learning data is not suitable because the MPED-RNN model shows very low accuracy.

5 CONCLUSION AND FUTURE RESEARCH

This paper verified whether learning data for abnormal behavior detection is suitable as learning data through the MPED-RNN model. Due to the nature of the data, the accuracy was not high in the unsupervised learning-based MPED-RNN model, but it is judged as valid learning data in supervised learning-based models because the frame of precursor and abnormal behavior is clear and skeleton data extraction is accurate.

Currently, artificial intelligence technology is a technology that attracts attention among the 4th industrial revolution, and active research is being conducted, and many companies are trying to use it in industrial sites. However, since high-quality learning data for artificial intelligence development is difficult and difficult to build, it is believed that more learning data led by highly reliable national institutions can promote the development of artificial intelligence technology and popularization of artificial intelligence technology.

In the future, we will continue to conduct research on technology that verifies the data with other models based on supervised learning and applies abnormal behavior detection technology to public CCTVs.

Acknowledgements

This research was supported by Seokyeong University in 2022.

6 REFERENCES

[1] Bae, H.-J. et al. (2021). LSTM (long short-term memory)-based abnormal behavior recognition using AlphaPose. *KIPS Transaction on Software and Data Engineering, 10*(5), 187-194. https://doi.org/10.3745/KTSDE.2021.10.5.187

[2] Baldi, P. (2012). Autoencoders, unsupervised learning, and deep architectures, workshop on unsupervised and transfer learning. *Proceedings of ICML Workshop on Unsupervised and Transfer Learning, PMLR, 27*, 37-50.

[3] Bradley, A. (1997). The use of the area under the ROC Curve in the evaluation of machine learning algorithms. *Pattern Recognition, 30*(7), 1145-1159. https://doi.org/10.1016/S0031-3203(96)00142-2

[4] Chalapathy, R. & Chawla, S. (2019). Deep learning for anomaly detection: A Survey. arXiv preprint arXiv: 1901.03407. https://doi.org/10.48550/arXiv.1901.03407

[5] Chandola, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys, 41*(3), 1-58. https://doi.org/10.1145/1541880.1541882

[6] Fawcett, T. (2006). Introduction to ROC analysis. *Pattern Recognition Letters, 27*, 861-874. https://doi.org/10.1016/j.patrec.2005.10.010

[7] Feng, K. (2019). Decision making with machine learning and ROC curves. arXiv: 1905.02810, 1-52. https://doi.org/10.2139/ssrn.3382962

[8] Haroon, U. (2022). A novel 3D-convolution neural network for human interaction recognition in videos. *The Journal of KING Computing, 18*(1), 19-28. https://doi.org/10.23019/kingpc.18.1.202202.002

[9] Hong, C. & Choi, S. (2020). ROC curve generalization and AUC. *Journal of the Korean Data & Information Science Society, 31*(4), 477-488. https://doi.org/10.7465/jkdi.2020.31.4.477

[10] Ji, S. (2013). 3D convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(1), 221-231. https://doi.org/10.1109/TPAMI.2012.59

[11] Kiran, B. (2018). An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos. *Journal of Imaging, 4*(36), 1-25. https://doi.org/10.3390/jimaging4020036

[12] Lee, J. (2020). A study on the implementation of intelligent abnormal behavior monitoring system using deep learning. *PhD Thesis*, Hanse Univ.

[13] Malhotra, P. (2016). LSTM-based encoder-decoder for multi-sensor anomaly detection. *ICML 2016 Anomaly Detection Workshop*, arXiv: 1607.00148. https://doi.org/10.48550/arXiv.1607.00148

[14] Morais, R. (2019). Learning regularity in skeleton trajectories for anomaly detection in videos. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11996-12004. https://doi.org/10.1109/CVPR.2019.01227

[15] Park, H. (2013). A study on monitoring system for abnormal behaviors by object's tracking. *Journal of Digital Contents Society*, 14(4), 589-596. https://doi.org/10.9728/dcs.2013.14.4.589

[16] Park, S. Anomaly detection by a surveillance system through the combination of C3D and object-centric motion information, *Journal of KIISE, 48*(1), 91-99.

[17] RNN Structure, https://velog.io/@sksmslhy/RNN

[18] Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., & Langs, G. (2017). Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery. *Information Processing in Medical Imaging, IPMI 2017. Lecture Notes in Computer Science, 10265*. Springer, Cham. https://doi.org/10.1007/978-3-319-59050-9_12

[19] Tran, D. (2015). Learning spatiotemporal features with 3D convolutional networks. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 4489-4497. https://doi.org/10.1109/ICCV.2015.510

[20] Wang, P. (2018). 3D shape segmentation via shape fully convolutional networks, computers & amp. *Graphics, Elsevier BV*, 76, 182-192. https://doi.org/10.1016/j.cag.2018.07.011

[21] Wu, T. & Lee, E. (2019). Human action recognition based on 3D convolutional neural network from hybrid feature. *Journal of KMMS*, 22(12), 1457-1465. https://doi.org/10.9717/kmms.2019.22.12.1457

[22] Xu, H. (2018). Unsupervised anomaly detection via variational auto-encoder for seasonal KPIs in web applications. *Proceedings of the 2018 WWW Conference*, 187-106. https://doi.org/10.1145/3178876.3185996

[23] Osama, M. (2021). Behavior Recognition based on signal processing technology. *International Journal of Hybrid Innovation Technologies, 1*(2), 75-90.

https://doi.org/10.21742/ijhit.2653-309X.2021.1.2.05
[24] Rashid, E. (2016). Software Fault Prediction Using Unsupervised Learning Technique: A Practical Approach. *International Journal of u - and e - Service, Science and Technology, NADIA, 9*(11), 275-288, https://doi.org/10.14257/ijunnesst.2016.9.11.24.

Author's contacts:

Yangsun Lee, Professor
Seokyeong University,
124 Seogyeong-ro Seongbuk-gu,
Seoul, 02173, Korea
yslee@skuniv.ac.kr