비지도 학습 기반 딥러닝 모델을 활용한 가상 네트워크 비정상 행위 탐지에 관한 연구

이청준†, 홍지범§, 최희열†
†한동대학교, §포항공과대학교

chunglee3224@gmail.com, hosewq@postech.ac.kr, heeyoul@gmail.com

Unsupervised learning for anomaly detection in virtual network environment

Chungjun Lee†, Jibeom Hong§, Heeyoul Choi†
†Handong Global Univ., §Pohang Univ. of Science and Technology

Abstract

Software Defined Networking (SDN) and Network Function Virtualization (NFV) have increased flexibility and efficiency in management of service in computer network. However, such development also raised difficulty and complexity of management of computer network and demanded new techniques for an effective management. Anomaly detection is one of such techniques to raise alarm when given service is in abnormal status. There have been many previous works which applied deep learning-based unsupervised anomaly detection in many other domains, but they have not been applied to SFC in virtual network environment. In this paper, we applied deep learning-based auto-encoder, auto-regressor models and thresholding methods to anomaly detection dataset collected from SFC in simulated network environment and achieved detection performance of 78.7% in f1-measure.

I. INTRODUCTION

Softwarization of computer networks led to more flexible utilization of network infrastructures [1]. However, it also led to growth in complexity of management and increase in difficulty for human engineers [2]. Consequently, there have been increasing attention to applying recent development of deep learning models for automatic control and management of softwarized computer network [2]. Especially, anomaly detection is one of important virtual network management techniques which enables quick response to possible quality degradation [2].

There are prior works that addressed anomaly detection problem in different domains using deep learning [3], [4], [5], [6], [7]. The proposed solutions are based on unsupervised learning because of ease for data collection, and difficulty in artificial creation of abnormality. However, the solutions have not been applied to service function chain (SFC) in virtual network environment.

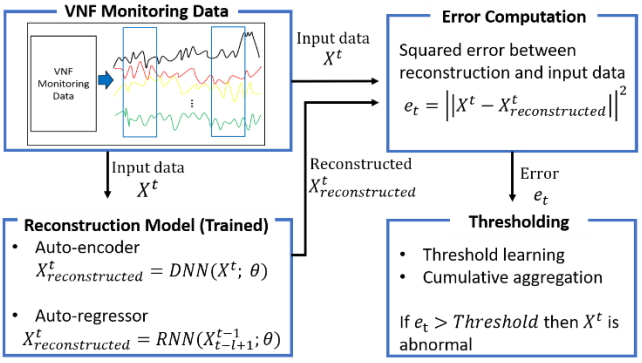

Figure 1: Overview of unsupervised anomaly detection method in virtual network environment.

In this paper, we apply deep learning-based unsupervised learning approach to anomaly detection based on monitoring data of SFC in virtual network environment. Contribution of this paper is to report potential detection performance of unsupervised learning algorithm in such setting. As shown in figure 1, we experiment with two reconstruction models and two thresholding methods to evaluate performances. In result, we observed anomaly detection performance of 78.7% in f1-score without dependence on labelling data.

Ⅱ. BODY

The problem is classifying the status of SFC at a given time. As shown in figure 2, monitoring data is given as X^t where $X^t \in \mathbb{R}^{V \times D}$, and V, D are number of virtual network function (VNF) instances in SFC and number of monitored metrics respectively. t is superscript indicating time index. Constraint is that in training time, we do not have monitoring data corresponding to abnormality. Detection performance is measured by test dataset composed of data and label tuples, (X^t, y^t) where y^t are binary label, and $t \in [1, \dots, N_{testset}]$ which are collected from simulation environment.

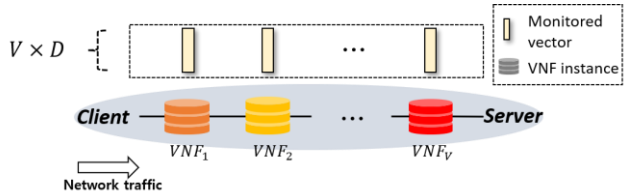

Figure 2: Collection of monitoring data from SFC.

Our unsupervised learning approach follows [5], [6] and [7] where they classify abnormal input data based on deviance from training data. Our proposed method follows 3 steps: (1) learn reconstruction model, (2) compute error and (3) apply threshold. These 3 steps are illustrated in figure 1. We train reconstruction models to accurately reconstruct input data by minimizing the error e_t in equation (1).

$$e_t = \left|\left|X^t - X^t_{reconstructed}\right|\right|^2 \quad (1)$$

The error e_t increases when input data deviates from normal. Reconstruction models are deep learning models like auto-encoder or auto-regressor.

Auto-encoder is deep neural network (DNN) and computes $X^t_{reconstructed}$ based on X^t as in equation (2). In the equation, θ is trainable parameters of the neural network. Auto-encoder learns a function which projects input data X^t into a low dimensional vector and then projects the low dimensional vector to $X^t_{reconstructed}$. On the other hand, auto-regressor is recurrent neural network (RNN), and it learns sequential patterns in training data to compute $X^t_{reconstruected}$ based on previous data sequence as in equation (3). In the equation, l is length of sequence.

$$X^t_{reconstruected} = DNN(X^t; \theta), \quad (2)$$

$$X^t_{reconstruected} = RNN(X^{t-1}_{t-l+1}; \theta). \quad (3)$$

We train the reconstruction models using the following objective function via the gradient descent method. In case of auto-regressor, X^t on left-hand side is replaced by X^{t-1}_{t-l+1} whose subscript indicate start index.

$$J(X^t; \theta) = \left|\left|X^t - X^t_{reconstructed}\right|\right|^2. \quad (4)$$

Thresholding is a process of determining normal or abnormal when an error value is computed. The first method is threshold learning where a set of threshold values are evaluated against small portion of labeled dataset to select the best threshold value [7]. The second method is cumulative aggregation which computes a threshold value using mean and standard deviation of error values from the training data [5]. When the computed error e_t exceeds the threshold, it is classified as an abnormal case.

Anomaly detection data was collected from a testbed that simulates SFC in different traffic scenarios [1]. Each instance of VNF in SFC provides various network management functions to traffic. Specifically, WSD dataset is collected from testbed that simulated web service scenario setting whereas LAD1 and LAD2 datasets are collected from that for login authentication scenario setting. Features consist of 23 OS-monitoring metrics from each VNF node that composes SFC. Each dataset has a different number of VNFs as well as

different types of VNFs, and their statistics are in table 1. For test and valid2 sets, they include cases which simulate occurrence of anomalies. The labelling of the anomalous data is based on service level agreement (SLA) standard. The monitoring data collected at the time when response time (less than 0.5 seconds) or availability (over 99.95% success of requests) of service was not satisfied are labeled as abnormal.

Dataset	WSD	LAD1	LAD2
# of VNFs	5	4	4
Total samples	68,731	121,053	121,053
Anomalies	26,354	19,913	44,513
Training set	27,451	65,529	59,470
Valid1 set	3,050	7,281	5,507
Valid2 set	6,496 (2,108)	9,683 (1,593)	9,680 (3,561)
Testing set	13,742 (5,270)	24,206 (3,982)	24,201 (8,902)

Table 1: Statistics of anomaly detection datasets.

Model	WSD	LAD1	LAD2	Average
Auto-encoder	**0.867**	**0.769**	0.725	**0.787**
Auto-regressor	0.844	0.684	**0.755**	0.761

Table 2: f1-measure of the reconstruction models using the threshold learning method.

Model	WSD	LAD1	LAD2	Average
Auto-encoder	0.848	0.635	0.717	0.733
Auto-regressor	**0.824**	0.645	**0.756**	**0.742**

Table 3: f1-measure of the reconstruction models using the cumulative aggregation method.

For experiment, the goal was to measure the anomaly detection performance of the proposed methods: auto-encoder and auto-regressor models. For thresholding, we experimented threshold learning and cumulative aggregation. Experiment results are shown in table 2 and 3. We observed that the threshold is more optimized using threshold learning where auto-encoder and auto-regressor show 78.7 and 76,1% performance in three datasets on average, while the cumulative aggregation shows 73.3 and 74.2%, respectively.

Ⅲ. CONCLUSION

In this paper, we experimented and showed that unsupervised learning models detects anomalies for SFC in virtual network environment without dependence on label information. Although supervised learning model shows relatively higher performance [1], its development is heavily dependent on label information. Future direction is to apply recent development of deep learning-based anomaly detection models to improve detection performance of unsupervised learning models.

ACKNOWLEDGMENT

This research was supported by the Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (No. 2018-0-00749, Development of virtual network management technology based on artificial intelligence).

REFERENCES

[1] J. Hong, S. Park, J. H. Yoo, and J. W. K. Hong, "Machine learning based sla-aware vnf anomaly detection for virtual network management," in 2020 16th International Conference on Network and Service Management (CNSM), 2020, pp. 1–7.

[2] S. Nedelkoski, J. Cardoso, and O. Kao, "Anomaly detection from system tracing data using multimodal deep learning," in 2019 IEEE 12th International Conference on Cloud Computing (CLOUD), 2019, pp. 179–186.

[3] A. Gulenko, F. Schmidt, A. Acker, M. Wallschlager, O. Kao, and F. Liu, "Detecting anomalous behavior of black-box services modeled with distance-based online clustering," in 2018 IEEE 11th International Conference on Cloud Computing (CLOUD), 2018, pp. 912–915.

[4] F. Schmidt, F. Suri-Payer, A. Gulenko, M. Wallschlager, A. Acker, and O. Kao, "Unsupervised anomaly event detection for cloud monitoring using online arima," in 2018 IEEE/ACM International Conference on Utility and Cloud Computing Companion (UCC Companion), 2018, pp. 71–76.

[5] F. Schmidt, A. Gulenko, M. Wallschlager, A. Acker, V. Hennig, F. Liu, and O. Kao, "Iftm - unsupervised anomaly detection for virtualized network function services," in 2018 IEEE International Conference on Web Services (ICWS), 2018, pp. 187–194.

[6] Malhotra, P., L. Vig, G. Shroff and Puneet Agarwal. "Long Short Term Memory Networks for Anomaly Detection in Time Series." ESANN (2015).

[7] Zhang, C., Song, D., Chen, Y., Feng, X., Lumezanu, C., Cheng, W., Ni, J., Zong, B., Chen, H., & Chawla, N.V. (2019). A Deep Neural Network for Unsupervised Anomaly Detection and Diagnosis in Multivariate Time Series Data. ArXiv, abs/1811.08055.