Robust anomaly detection methods for contamination network data

Nguyen Manh Tuan[1], Nguyen Hai Hao[1],
Dang Le Dinh Trang[2], Nguyen Van Tuan[2], Cao Van Loi[2*]

[1] Cyberspace Operation Command, Hanoi, Vietnam;
[2] Faculty of Information Technology, Le Quy Don Technical University, Hanoi, Vietnam.
[*]Corresponding author: loi.cao@lqdtu.edu.vn
Received 15 March 2022; Revised 29 April 2022; Accepted 06 May 2022; Published 19 May 2022.
DOI: https://doi.org/10.54939/1859-1043.j.mst.79.2022.41-51

ABSTRACT

Recently, latent representation models, such as Shrink Autoencoder (SAE), have been demonstrated as robust feature representations for one-class learning-based network anomaly detection. In these studies, benchmark network datasets that are processed in laboratory environments to make them completely clean are often employed for constructing and evaluating such models. In real-world scenarios, however, we can not guarantee 100% to collect pure normal data for constructing latent representation models. Therefore, this work aims to investigate the characteristics of the latent representation of SAE in learning normal data under some contamination scenarios. This attempts to find out wherever the latent feature space of SAE is robust to contamination or not, and which contamination scenarios it prefers. We design a set of experiments using normal data contaminated with different anomaly types and different proportions of anomalies for the investigation. Other latent representation methods such as Denoising Autoencoder (DAE) and Principal component analysis (PCA) are also used for comparison with the performance of SAE. The experimental results on four CTU13 scenarios show that the latent representation of SAE often out-performs and are less sensitive to contamination than the others.

Keywords: Anomaly detection; Latent representation; One-class classification; Contamination.

1. INTRODUCTION

For decades, a number of machine learning and deep learning techniques have been introduced for solving problems in network anomaly detection [1-3]. One of the biggest obstacles in this domain is to collect anomalous data and later label them with specific types of network attacks. Attack activities have continuously evolved as well as many zero-day vulnerabilities have been increasingly exploited. Thus, labeling network data as normal or specific attack types is a challenging task. For these reasons, semi-supervised learning methods, especially one-class classification (OCC), tend to be very common for the network anomaly detection problem [4, 5].

Recently, the latent representation of autoencoders (AEs) has been employed to enhance the performance of anomaly detection models [4-8]. For OCC-based anomaly detection, Shrink autoencoder (SAE) is one of the typical methods for constructing a "good" feature space for subsequent one-class classifiers [4]. SAE with a new regularizer learns to represent data in a more meaningful feature space. In this feature space, normal data is forced to distribute in a tiny region around the origin while the rest of the space is reserved for anomalies that appear in the future. Once the training is completed, the SAE decoder is discarded while its encoder is employed to produce the feature representation for subsequent classifiers. However, the network data collected from real-world environments is often diverse, and may contain outliers and unexpected anomalies. This

could be due to human limitations in collecting, processing and extracting data, while sophisticated attacks are on the rise. For example, hackers usually try to counterfeit some characteristics of the normal event to hide malicious activities. For these reasons, humans can not guarantee to completely filter out outliers and unexpected anomalies from normal network data.

To deal with the mentioned above issues, one-class learning models should have the capability to learn the normal behavior from network data with contamination such as anomalies. As mentioned before, latent representation models, such as SAE have been demonstrated as a robust feature representation model for network anomaly detection. Therefore, this study will investigate the characteristics of the latent representation of SAE in learning normal network data under some contamination conditions. This aims to find out wherever the latent feature space of SAE is robust to contamination or not. Thus, we will design a number of experiments with normal data contaminated with different anomaly types, such as Gaussian noise and real-world anomalies with different levels. Other latent representation methods such as Denoising Autoencoder (DAE) and Principal component analysis (PCA) are used for comparison with the performance of SAE.

The rest of the paper is organized as follows. Section 2 presents some recent studies on latent representation methods for network anomaly detection. Section 3 describes our approach to investigate the latent representation of SAE in comparison to those of DAE and PCA. This is followed by Experiments, Results and Discussions in section 4 and 5. The conclusion and future direction is presented in section 6.

2. RELATED WORKS

In [6], Sarah et al. proposed a hybrid model to solve the high-dimensional problem of anomaly detection. They use an unsupervised Deep Belief Network (DBN) to extract generic underlying feature, and then use these features to train an one-class SVM. The DBN is trained as a dimensionality reduction algorithm that turns data into a lower-dimensional set of features and creates a non-linear manifold. This helps improve the scalability of a one-class SVM by reducing the amount of redundant features. The hybrid model DBN-SVM attains the accuracy of a state-of-the-art autoencoder while reducing training and testing time, also avoids the complexities of non-linear kernel machines. Recently the authors of [4] proposed Shrink Autoencoders (SAE) and Dirac delta Variational Autoencoder (DVAE), which generate the data representation by gathering normal data behaviors. These regularized AEs were designed to overcome the problem of identifying anomalies in high-dimensional network data. The Shrink Autoencoder adds a new regularizer to the loss function while DVAE attempts to encode data to be distributed as a standard Gaussian in the latent space. The purpose of these regularizers AEs is to direct normal data to a small region near the origin of the latent space, and saves the remainder of the space for anomalies in the future. The experimental results were produced from five simple one-class classifications (OCC) algorithms: LOF, CEN, KDE, MDIS and OCSVM, using the latent representations of SAE and DVAE on 14 problems. Both models bring nice distribution when forcing normal data into a very tight area centered at the origin.

Application for network security where legitimate data may contain contamination of

cyber-attacks, there are many researchers have attempted to extract potential anomalies (or noise) hidden in training data or seek for robust feature representation to the contamination. Pang et al. [9] proposed a ranking model-based framework, called RAMODO, to learn a low-dimensional feature representation of normality that is robust to contamination. The authors used a meta triplet sample and an outlier score-based ranking loss to iteratively extract potential outliers from normality in an unsupervised learning manner. However, the normality is allowed to distribute in some regions in the feature space with arbitrary shapes. Thus, the resulting models can be susceptible to hyper-parameters [4]. Recently, [7] represented the effects of noise with different degrees on anomaly detection models, including the hybrid AE-SOM model and the PCA model. The authors create four training datasets with different IoT malware ratios to create unbalanced and balanced sets. The experimental results highly recommend that AEs can be used to represent unbalanced data, while PCA is more preferred used for balanced data cases. In [10], Zhou et al. proposed a method to leverage an autoencoder to encode the input data and utilize three factors: hidden representation, reconstruction residual vector and reconstruction error, as the new representation for the input data. To cope with the limitation of data insufficiency for labeled anomalies, the proposed model contains the feature encoder and the anomaly score generator, which are jointly optimized during the training process with a specially designed deviation loss. Guansong Pang et al. [11] have the same interest in Deep Weakly-supervised Anomaly Detection. Rather than optimizing the feature representation, Pang proposed an anomaly detection formulation for the scenario with limited labeled abnormal data and abundant information on unlabeled data. They convert the anomaly detection task into an ordinal regression problem of pairwise relation, which fully utilizes the limited number of labeled anomaly instances to form instance pairs for further anomaly detection.

3. CONTAMINATED INVESTIGATION

This section will introduce the method to explore the behavior of the latent representation of SAE and the subsequent one-class classifiers in hybrid models under contamination scenarios. Firstly, we will make some assumptions about the possible influences of anomalies on the latent representation methods and following classifiers. Following this, we introduce the methods to contaminate normal data with anomalies and employed the contaminated data to evaluate hybrid anomaly detection models.

Based on the principles of SAE, DAE and PCA, we can make some initial assumptions about the influences of anomalies hidden in training data on their feature representations. When PCA is used for feature reduction, it will project data onto a new feature space using only the first few principal components that reserves the most data variation. If anomalies hidden in normal data deviate so much from normal data, the principal components representing the variation can be included in the new coordinate system. Thus, PCA can project data into a new feature space in which data points (normal or anomalies) are mostly separated from each other. On the other hand, Autoencoders try to represent the distribution of a certain data class rather than project the data points far away from each other. For instance, SAE attempts to represent the normal class in tiny region around the origin of its latent feature space [4]. It reserves the rest of the feature space for anomalies appearing in the future. Similarly, DAE tends to

push normal data points close together, but in an arbitrary shape [4]. When normal data is contaminated with anomalies at a certain level, these anomalies can be treated as the second class in training data. This will make SAE and DAE difficult to learn multi-distributions simultaneously.

We also intentionally select hyper-parameters of the subsequent one-class learning methods to clearly show the influences of the contamination on the performance of hybrid methods. In this work, we employ two well-known methods such as one-class SVM and LOF, for constructing hybrid models. In OCSVM, nu determines an upper bound of the fraction of training errors and a lower bound of the fraction of support vectors. Thus, nu indicates the lower bound of the number of data points that will be used as support vectors. These points tend to be points that deviate significantly from the rest of normal data, such as outliers and anomalies. When the number of anomalies hidden in normal data increases, it is possible that the number of support vectors being anomalies increases. If the value of nu is smaller than the contamination rate, all support vectors could be anomalies. In LOF, the value k is the number of neighbors selected to calculate the local density for classifying new data points. When LOF is used as one-class classification, normal data is used for estimating the local density of new data points. The local density of a given data point above a predetermined threshold indicates an anomaly. When training data including anomalies, LOF-based classifiers could wrongly classify new data points if they fall into the region of the anomalies.

From the above analysis, we investigate the performance of the latent representation of SAE as well as that of the hybrid SAE and OCCs under contamination scenarios. Specifically, we design a number of experiments by contaminating the normal training data with different types of anomalies and different contamination rates. We also compare the performance of SAE with other latent representation methods such as DAE and PCA. This aims to show the behavior of SAE under a number of contamination scenarios, and which one-class classifiers are good for working on the latent representation of SAE under the scenarios.

4. EXPERIMENTS

In this section, we describes four botnet datasets and the methods to contaminate normal training data for experiments. Following this, we define terms for hybrid models and presents hyper-parameters in details.

4.1. Botnet datasets

A botnet traffic data, namely CTU13, is employed for experiments. The dataset was captured in the CTU University, Czech Republic, in 2011. The dataset consists of a large capture of real botnet traffic mixed with normal traffic and background traffic. We choose four scenarios from the CTU13 dataset, namely CTU13-8, CTU13-9, CTU13-10 and CTU13-13. Each of these datasets is divided into 40% for training (discarding anomalies and keeping normal traffic) and 60% for evaluating (contains normal traffic and botnet). The 14 features in CTU13 except source/destination IP addresses are selected. The three categorical features such as protocol, sTos and dTos, are encoded using One-hot Encoding, resulting in higher-dimensional versions of these scenarios. The details of the scenarios is shown in table 1.

Table 1. *Four CTU13 scenarios used for experiments.*

Scenario	Normal train	Normal test	Anomaly test	No. of original dimension	No. of processed dimension
CTU13-08	29128	43694	3677	16	40
CTU13-09	11986	17981	110993	16	41
CTU13-10	6338	9509	63812	16	38
CTU13-13	12775	19164	24002	16	40

4.2. Contamination scenarios

The normal training data extracted from each of four datasets (CTU13-08, CTU13-09, CTU13-10 and CTU13-13) is contaminated with anomalies at different levels such as 0%, 1%, 5% and 10%. We use two kinds of anomaly data to contaminate the normal training data as follows.

- The first is real botnet data taken from each CTU13 scenario. As described earlier, we randomly sample 40% of each CTU13 dataset for training. In this part, the normal data is chosen as the normal training set while a certain proportion of botnet traffic is randomly sampled for contaminating the normal training set. The proportion can be 0%, 1%, 5% and 10%.

- Instead of using real botnet traffic, artificial anomaly data can be randomly generated based on Gaussian distribution with mean (m) of zero and standard deviation (std) of 0.2. The number of features of the artificial anomaly is the same as that of the dataset. The artificial anomalies are then mixed with the normal training data to create contaminated training data as the first case above. This aims to simulate unseen anomaly data because in case we can not define specific types of anomalies.

Therefore, we design two experiments using the two contamination data above for investigating the hybrid models introduced in section 3. The "Are Under the Curves" (AUC) is used to measure the performance of these hybird models. The results and analysis are presented in section 5.

4.3. Parameter settings

As mentioned, we investigate the behavior of the latent representation of SAE in comparison to other latent representation models such as DAE anb PCA. The combination of these feature representation models with OCC methods can create a set of hybrid anomaly detectio models. We use the terms of SAE-OCSVM, DAE-OCSVM, PCA-OCSVM, SAE-LOF, DAE-LOF and PCA-LOF to refer to the hybrids between the latent representation methods (i.e. SAE, DAE and PCA) and one-class classification methods (i.e. OCSVM and LOF). The hyper-parameters of subsequent classifiers are also intentionally chosen such as $nu = 0.1$ and 0.5 for OCSVM (namely as OCSVM-0.5 and OCSVM-0.1) and $k = 0.1$ for LOF. The hyper-parameter settings of SAE, DAE, LOF and OCSVM are presented in table 2. Regards Principle Component Analysis model, we select the first seven principal components.

Table 2. *Hyper-parameter settings for SAE, DAE, LOF and OCSVM.*

Parameter	SAE	DAE	LOF	OCSVM
Hidden layer	7	7	-	-
Trigger function	tanh	tanh	-	-
Epoch	1000	1000	-	-
Batch size	100	100	-	-
Shrink lambda	10	-	-	-
Neighbor (*k*)	-	-	1%	-
nu	-	-	-	0.5, 0.1
Kernel function	-	-	-	rbf

5. RESULTS AND ANALYSIS

The experimental results are presented in tables 3 and 4 as well as figure 1. In these tables, the cells with darker colors represent lower performances (AUC values). In the figure, we visualize the data in the latent feature space of SAE in comparison to those of DAE and PCA when using artificial anomalies.

5.1. The first experiment

In this experiment, the normal training data is contaminated with a small subset of real anomalies. In table 3, its columns indicate the performance of the hybrid latent respresentation methods (SAE, DAE and PCA) and OCCs (OCSVM-0.5, OCSVM-0.1 and LOF) models, and its rows represent the performance of these hybrid models on the four CTU13 scenarios. For each CTU13 dataset, we report the results on four different contamination rates.

It can be seen from table 3 that the AUC values produced from all hybrid models are influenced by contamination in the training data. The AUC values tend to decrease as the contamination rate increases. To explain for this, we will analyses deeply two parts of these hybrid models such as latent representation blocks and subsequent one-class classifiers.

Table 3. *Performance of hybrid models when training data contaminated with real anomalies.*

Contanimation rate	OCSVM nu=0.5			OCSVM nu=0.1			LOF		
	SAE	DAE	PCA	SAE	DAE	PCA	SAE	DAE	PCA
CTU 13-08									
0%	0.987	0.952	0.956	0.985	0.597	0.803	0.987	0.897	0.878
1%	0.968	0.876	0.962	0.953	0.792	0.804	0.972	0.896	0.891
5%	0.939	0.955	0.956	0.872	0.704	0.807	0.887	0.896	0.910
10%	0.840	0.959	0.956	0.716	0.666	0.792	0.800	0.857	0.867
CTU 13-09									
0%	0.950	0.676	0.792	0.945	0.719	0.783	0.954	0.890	0.941
1%	0.905	0.564	0.740	0.902	0.599	0.678	0.942	0.912	0.918

5%	0.850	0.732	0.821	0.841	0.744	0.743	0.927	0.900	0.924
10%	0.841	0.596	0.845	0.789	0.641	0.675	0.910	0.893	0.935
CTU13-10									
0%	0.999	0.985	0.980	0.999	0.962	0.405	0.997	0.721	0.914
1%	0.985	0.995	0.997	0.984	0.985	0.986	0.954	0.770	0.896
5%	0.951	0.987	0.979	0.879	0.646	0.613	0.814	0.343	0.920
10%	0.915	0.968	0.946	0.679	0.552	0.444	0.588	0.313	0.302
CTU 13-13									
0%	0.965	0.636	0.925	0.963	0.365	0.795	0.974	0.859	0.961
1%	0.937	0.591	0.921	0.904	0.362	0.804	0.941	0.930	0.961
5%	0.926	0.786	0.923	0.829	0.299	0.630	0.905	0.923	0.960
10%	0.800	0.459	0.918	0.704	0.327	0.547	0.874	0.847	0.958

For the subsequent classifiers, the table shows that the hybrid models with OCSVM-0.5 often perform better than those with OCSVM-0.1 as the contaminate rate increasing. This results strongly support our analysis to the influence of the contamination data on the performance of OCSVM as presented. The reason is that when the contamination rate increases, the percentage of the support vectors of OCSVM-0.1 being anomalies can increase and can be higher than that of OCSVM-0.5. The more support vectors are anomalies, the worse the performance OCSVM produces. Therefore, the hybrid models with OCSVM-0.5 often out-perform the hybrid models with OCSVM-0.1 on almost cases. Moreover, the hybrid models with LOF shows a considerable decrease in their performance as the contamination rate increasing. This means that hybrid models with LOF $k = 0.1$ are considerably effected by the contaminated data.

For the latent representation models, table 3 clearly illustrates that the hybrid models with SAE often out-perform the hybrid models with other feature representation models like DAE and PCA. In addition, there is a gradual decrease in the AUC on the SAE columns as the contamination rate increasing. This means that SAE is still influenced by the anomalies, but not much and in predictable manner. On the other hand, the DAE and PCA columns show many darker colored cells without any clear trends as the contamination rate increasing. This could be explained as the following. PCA tends to focus on projecting input data into a new feature space in which data points are separated from each other as much as possible rather than focusing on representing normal data. Thus, the contamination rate has shown influences on the performance of the hybrid models with PCA but does not follow a clear trend. Moreover, the performance of the hybrid models with DAE illustrates an unpredictable trend because the latent feature space of DAE is not designed for the task of learning normal data. Therefore, it tends to perform less effectively than SAE even at the contamination rate of 0.0%. When the contamination rate increases, it becomes more difficult for DAE to learn normal data. Thus, the hybrid models with DAE can sometimes produce good AUC values or very badly AUC values.

The above analysis can confirm that the contamination in training data has considerable influence on the performance of the hybrid models. When evaluating on the

first contamination scenarios, the hybrid models with OCSVM-0.5 and LOF can produce good AUC values. For feature representation models, SAE is very robust to the contamination because its AUCs are the highest on almost cases and change gradually as the contamination rate increasing.

5.2. The second experiment

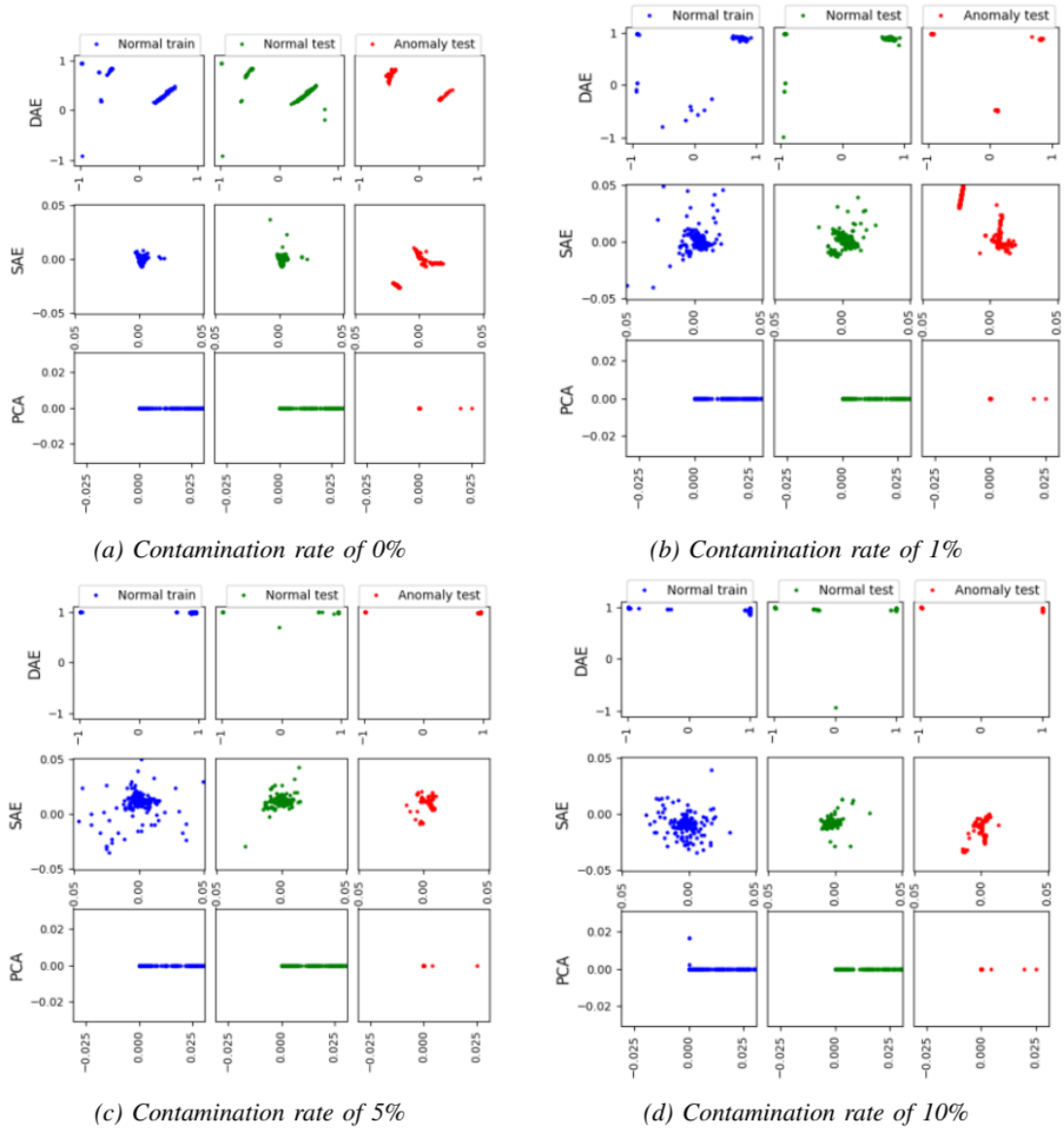
The normal training data is contaminated with artificial anomalies generated from Gaussian distribution as described in section 2. The results are illustrated in table 4 and figure 2. These table and figure have the same format as those in the first experiment.

Table 4. *Performance of hybrid models when training data contaminated with artificial anomalies.*

Contanimation rate	OCSVM nu=0.5			OCSVM nu=0.1			LOF		
	SAE	**DAE**	**PCA**	**SAE**	**DAE**	**PCA**	**SAE**	**DAE**	**PCA**
CTU 13-08									
0%	0.982	0.906	0.956	0.968	0.815	0.802	0.983	0.894	0.879
1%	0.978	0.964	0.956	0.976	0.869	0.800	0.981	0.903	0.883
5%	0.976	0.945	0.954	0.961	0.804	0.798	0.962	0.922	0.892
10%	0.948	0.942	0.955	0.853	0.628	0.799	0.956	0.888	0.896
CTU 13-09									
0%	0.899	0.531	0.755	0.900	0.573	0.724	0.941	0.895	0.939
1%	0.925	0.626	0.788	0.916	0.666	0.663	0.938	0.898	0.925
5%	0.880	0.628	0.720	0.874	0.581	0.530	0.936	0.920	0.923
10%	0.916	0.806	0.612	0.916	0.756	0.527	0.942	0.915	0.919
CTU13-10									
0%	1.000	0.984	0.979	0.999	0.473	0.342	0.999	0.750	0.884
1%	0.996	0.988	0.993	0.995	0.912	0.958	0.979	0.721	0.904
5%	0.996	0.984	0.942	0.995	0.147	0.067	0.971	0.348	0.900
10%	0.991	0.934	0.896	0.989	0.130	0.053	0.976	0.296	0.312
CTU 13-13									
0%	0.961	0.630	0.927	0.962	0.328	0.747	0.975	0.884	0.959
1%	0.967	0.671	0.904	0.965	0.327	0.420	0.972	0.922	0.955
5%	0.907	0.530	0.913	0.815	0.371	0.442	0.966	0.907	0.963
10%	0.930	0.320	0.911	0.752	0.273	0.378	0.967	0.904	0.967

In overall, the result from table 4 shows a very similar patterns to those from table 3 in the first experiment. When exam on the subsequent one-class classifiers, OSVM-0.5 and LOF often out-perform OCSVM-0.1 on almost cases. For latent representation models, SAE also demonstrates a better performance in comparison to DAE and PCA. For example, the AUC values produced by the hybrid models using SAE are often much higher than those produced by the hybrid models using DAE and PCA. The AUC values on the SAE columns also change more slightly than the AUC values on other columns as

the contamination rate increasing. However, when the contamination rate decreasing the AUC values from the second experiment seem to change more smoothly than in the first experiment. This could be explained that the real anomalies may be sampled from some particular attacks while artificial anomalies generated by Gaussian distribution can play as general anomalies. when contaminating data with real anomalies, the performance of the hybrid models may change more rapidly than using artificial anomalies.



(a) Contamination rate of 0%

(b) Contamination rate of 1%

(c) Contamination rate of 5%

(d) Contamination rate of 10%

Figure 1. *Latent data of CTU13-08 created by SAE, DAE and PCA when using four contamination levels.*

Figure 1 illustrates the latent data of CTU13-08 in the latent feature spaces of SAE, DAE and PCA trained using four different contamination rates. It can be seen that data points in the feature spaces of SAE and DAE tend to be more and more separated as the contamination rate increasing. This is because the more number of anomalies is hidden in training data, the more difficult SAE and DAE can learn to represent normal data in

small regions. On the other hand, the data patterns created from PCA tend to be more stable as the number of anomalies increasing. This result can be used to explain why the AUC values in the PCA columns in table 2 tend to be less effected by the contamination rate than other columns.

6. CONCLUSIONS

In this paper, we have proposed a method to explore the behavior of the latent representation of SAE in conjunction with one-class learning methods for network anomaly detection under contamination data scenarios. This attempts to point out wherever the latent feature space of SAE is robust to contamination or not, and which contamination scenarios it prefers. To extensively investigate SAE, we design a set of experiments with different contamination types and rates for evaluating SAE in comparison to other methods such as DAE and PCA. The normal data from four CTU13 scenarios are contaminated with anomalies to make it more likely to be similar to real-world data.

The experimental results demonstrate that the latent representation of SAE performs well, and is often less sensitive to the contaminated data than DAE and PCA. The results also suggest that hyper-parameters of the subsequent one-class learning methods could be used to reduce the influence of the contamination data on the performance of hybrid anomaly detection models. Further work will focus on how to estimate the hyper-parameters to reduce the influence of contamination data, and also conduct experiments on recent datasets.

REFERENCES

[1]. A. Zimek, E. Schubert, and H.-P. Kriegel, "*A survey on unsupervised outlier detection in high-dimensional numerical data,*" Statistical Analysis and Data Mining, vol. 5, no. 5, pp. 363–387, 2012.

[2]. G. Pang, L. Cao, and C. Aggarwal, "*Deep learning for anomaly detection: Challenges, methods, and opportunities,*" in Proceedings of the 14th ACM International Conference on Web Search and Data Mining, pp. 1127–1130, 2021.

[3]. G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, "*Deep learning for anomaly detection: A review,*" ACM Computing Surveys (CSUR), vol. 54, no. 2, pp. 1–38, 2021.

[4]. V. L. Cao, M. Nicolau, and J. McDermott, "*Learning neural representations for network anomaly detection,*" IEEE Transactions on Cybernetics, no. 99, pp. 1–14, 2018.

[5]. V. L. Cao, M. Nicolau, and J. McDermott, "*A hybrid autoencoder and density estimation model for anomaly detection,*" in Parallel Problem Solving from Nature, pp. 717–726, Springer, 2016.

[6]. S. M. Erfani, S. Rajasegarar, S. Karunasekera, and C. Leckie, "*High-dimensional and large-scale anomaly detection using a linear one-class svm with deep learning,*" in Pattern Recognition 58, p. 121–134, 2016.

[7]. H. N. Nguyen, V. C. Nguyen, N. N. Tran, and V. L. Cao, "*Feature representation of autoencoders for unsupervised iot malware detection*," in International Conference on Future Data and Security Engineering, pp. 272–290, Springer, 2021.

[8]. A. S. Iliyasu, U. A. Abdurrahman, and L. Zheng, "*Few-shot network intrusion detection using discriminative representation learning with supervised autoencoder*," Applied Sciences, vol. 12, no. 5, p. 2351, 2022.

[9]. G. Pang, L. Cao, L. Chen, and H. Liu, "*Learning representations of ultrahigh-dimensional data for random distance-based outlier detection*," in Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, pp. 2041–2050, 2018.

[10]. Y. Zhou, X. Song, Y. Zhang, F. Liu, C. Zhu, and L. Liu, "*Feature encoding with autoencoders for weakly-supervised anomaly detection*," IEEE, 2021.

[11]. P. Guansong, S. Chunhua, J. Huidong, and v. d. H. Anton, "*Deep weakly-supervised anomaly detection*," arXIV Computing Surveys (CSUR), vol. 54, no. 2, pp. 1–38, 2020.

TÓM TẮT

Các phương pháp phát hiện bất thường cho dữ liệu chứa nhiễu

Gần đây, các mô hình biểu diễn không gian ẩn, chẳng hạn như Shrink Autoencoder (SAE), đã thể hiện hiệu năng mạnh mẽ trong việc nâng cao hiệu suất của kỹ thuật phân loại một lớp trong phát hiện bất thường mạng. Tuy nhiên, tập dữ liệu bình thường được dùng để huấn luyện các mô hình đang được giả định là hoàn toàn sạch, không chứa nhiễu và dữ liệu bất thường nào, điều này là khó khả thi trong thực tế. Do đó, bài náo này nghiên cứu khả năng biểu diễn ẩn của SAE trong việc trích xuất dữ liệu có chứa nhiễu. Mục đích là để tìm ra với mức độ nhiễu nào biểu diễn ẩn của SAE có khả năng bị ảnh hưởng mạnh. Bài báo thiết kế một số thí nghiệm với các mức độ nhiễu khác nhau cùng với một số phương pháp sinh nhiễu khác nhau. Các phương pháp biểu diễn đặc trưng khác như Denoising Autoencoder (DAE) và Phân tích thành phần chính (PCA) cũng được sử dụng để so sánh với hiệu suất của SAE. Kết quả thử nghiệm trên bốn kịch bản CTU13 cho thấy rằng, biểu diễn tiềm ẩn của SAE thường hoạt động tốt hơn và ít bị ảnh hưởng bởi nhiễu hơn so với các mô hình biểu diễn đặc trưng khác.

Từ khoá: Phát hiện bất thường; Biểu diễn ẩn; Học một lớp; Sự nhiễm bẩn.