2016 8th International Conference on Computational Intelligence and Communication Networks

Improvement of Network Intrusion Detection Accuracy by using Restricted Boltzmann Machine

Sanghyun Seo
Department of Computer Engineering
Dongguk University
Seoul, Korea
shseo@dongguk.edu

Seongchul Park
Department of Computer Engineering
Dongguk University
Seoul, Korea
scpark@dongguk.edu

Juntae Kim
Department of Computer Engineering
Dongguk University
Seoul, Korea
jkim@dongguk.edu

Abstract— **In various data of network intrusion detection used for classification algorithm's learning, a great deal of noise and outlier data are mixed. In case of a learning performed by using data of high impurities, no matter how the performance of classification algorithm is outstanding, any network intrusion detection model of high performance becomes hard to anticipate. To increase the accuracy of network intrusion detection, not only the performance of classification algorithm should be increased but also the management on noises and outliers in the data used for the classification algorithm's learning. Restricted Boltzmann Machine (RBM) is a type of unsupervised learning that doesn't use class labels. RBM is a probabilistic generative model that composes new data on input data based on the trained probability. The new data composed through RBM show that the noises and outliers are removed from the input data. When the newly composed data are applied to the network intrusion detection model, negative effects from the noise and outlier data to the learning are eliminated. In this study, noises and outliers in KDD Cup 1999 Data are removed by applying the data to RBM and composing a new data. Then, use results between the existing data and the data from which noises and outliers are removed are compared. In conclusion, this study demonstrates the performance improvement of network intrusion detection resulted by removing noises and outliers included in the data through RBM.**

Keywords : Network Intrusion Detection System; Deep learning; RBM(Restricted Boltzmann Machine);

I. INTRODUCTION

Intrusion prevention systems (IPS) are classified into Host-based Intrusion Detection System (HIDS) that monitors the system in the operation system depending on the installed location and Network-based Intrusion Detection System (NIDS) that analyzes network packers at a specific point in the network. Network intrusion detection methods are classified into Misuse detection method and Abnormality detection method.

In general, the misuse detection method is applied to NIDS. This method detects intrusions though pattern matching on normal approaches and abnormal approaches by extracting specific signatures from attacks. The attack detection method using signatures learns patterns of previous intrusions or attacks so that the error rate of this method is high. This method is weak against newly attempted attack types or even attack methods that have been slightly revised [5] [8].

Contrariwise, the abnormality detection method detects intrusions by defining packets out of the normal range as abnormal based on characteristics of normal data. The abnormality detection method configures normal and average states as its determination criteria. Any detected packet changes rapidly compared to the configured criteria is classified as abnormal [2] [13]. Regarding the abnormality detection method, Fiore. U. et al. propose DRBM (Discriminative Restricted Boltzmann Machine) that adds detection features in existing RBM models [6].

Like this, divergence network detection method of NIDS is trained by network packet data. But the data used for learning contains attributes not significantly influential to detect intrusions and a great deal of noise and outlier data. Such meaningless attribute values and noise and outlier data are major causes that reduces performances of classification and detection techniques. As a result, for improving performance of NIDS, it is important to reduce noise and outlier data used learning for network detection method.

II. RELATED RESEARCHES

In removing the noise and outlier and extracting feature from data, Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are techniques mostly used to preprocess data. PCA finds the vector with largest data distribution and the most influential component in the data by vertically projecting the data. When PCA is used, any high level data can be reduced to the minimal required level. Through this, the calculation amount of classification model is reduced and impacts of components including noises are removed so that performance improvement of a classification model can be anticipated. LDA maximizes characteristics of attribute values by maximizing gap between data classes and minimizing gaps in the class. Therefore, it helps the

978-1-5090-1144-5/16 $31.00 © 2016 IEEE
DOI 10.1109/CICN.2016.87

classification model to identify a glass label so that performance improvement of the intrusion detection technique can be anticipated [2] [15].

RBM is a type of artificial neural networks presented by Hinton. G. E. [10]. This is an energy based probabilistic generative model receiving attentions as a major algorithm for Deep learning. RBM determines state of each nodes by converting the energy to the probability. The weight and bias between each node are trained to reduce the difference between the input data and the reconstructed data by the probability. When new data input to learning completed RBM, the energy and probability are calculated based on the weight and bias and reconstructed data are produced. The data reconstructed by RBM are based on the probability of trained model so that noise and outlier data are reconstructed with a low probability. Therefore, when data are reconstructed through the model of learning completed RBM, noises and outliers are removed from the data. If an existing classification algorithm learns data in which noises and outliers are removed through RBM, performance improvement of network intrusion detection can be expected [7][9][10][11].

This study compares intrusion detection rates between the NIDS using only a classification model and the NIDS trained with data in which noises and outliers are removed through RBM. Chapter 2 describes PCA, LDA and RBM as a way to reduce the noise and outlier data. Chapter 3 presents the theoretical background of RBM and an actual network intrusion detection model using RBM. Chapter 4 conducts an experiment with KDD Cup 1999 data and presents its result. Chapter 5 presents the conclusion and a direction for future studies.

III. RESTRICTED BOLTZMANN MACHINE

Boltzmann Machine (BM) is a stochastic recurrent neural network in the form that all nodes in visual and hidden layers are fully connected. When data input in the visual layer, the value of hidden layer is defined with the probability based on Boltzmann distribution. The followings are energy formula of BM and Fig 1 is the BM model [1].

$$\mathrm{E} = -\Big(\sum_{i<j} w_{ij} s_i s_j + \sum_i \theta_i s_i\Big) \qquad (1)$$

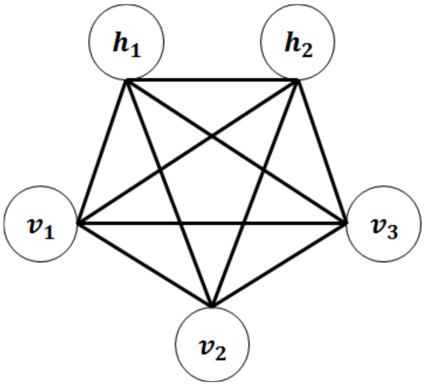

Fig. 1 Boltzmann Machine

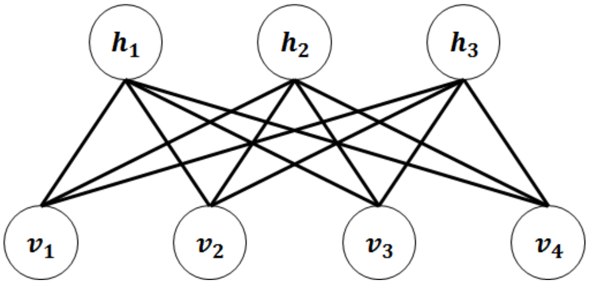

Fig. 2 Restricted Boltzmann Machine

Where w_{ij} is the strength of connection between units i and j, s_i is 1 if unit i is on and 0 otherwise, and θ_i is a threshold[1].

BM activates state of each nodes by employing the energy model. As the energy of node is higher, the probability to exist in a corresponding state is lower. As the energy of node is lower, the probability to exist in a corresponding state is higher. As Fig 1 shows, all nodes are connected in BM so that dependency between each nodes occurs. And, since the number of nodes increases, the calculation to acquire the probability to determine each node's state becomes complicated.

Unlike BM having all nodes connected, RBM refers the model in which each node is only connected to the ones in different layers and connections in the same layer are removed [10] [11]. As Fig 2 shows, connections between nodes in the same layer are restricted. A node in the visual layer is connected to all nodes in the hidden layer, and a node in the hidden layer is connected to all nodes in the visual layer. Like BM, the energy based model is employed and each node is in a binary state of 0 or 1. In RBM, the energy of a node is calculated through connections to all nodes in the other layer. The following figures show the energy formula of RBM and the state probability $p(v)$ on the input data v.

$$\mathrm{E(v,h)} = -\sum_{i\in visible} a_i v_i - \sum_{j\in hidden} b_j h_j - \sum_{ij} v_i h_j w_{ij} \qquad (2)$$

Where v_i, h_j are the binary states of visible unit i and hidden unit j, a_i, b_j are their biases and w_{ij} is the weight between them[11].

$$\mathrm{p(v,h)} = \frac{1}{Z} e^{-E(v,j)} \qquad (3)$$

$$\mathrm{Z} = \sum_{v,h} e^{-E(v,h)} \qquad (4)$$

$$p(v) = \frac{1}{Z} \sum_h e^{-E(v,h)} \qquad (5)$$

Learning for RBM is performed by finding the stable energy distribution best representing the distribution of training data. Log- likelihood method is used to minimize the energy to training data v and maximize probability $p(v)$. When the probability is maximized by using Gradient Ascent, following results are acquired [11].

$$\frac{\delta \log p(v)}{\delta w_{ij}} = < v_i h_j >_{data} - < v_i h_j >_{model} \qquad (6)$$

$$\Delta w_{ij} = \epsilon < v_i h_j >_{data} - < v_i h_j >_{model} \qquad (7)$$

Where ϵ: learning rate[11].

The first entry of formula (6) called "positive phase" refers the expected value of all h to input data v. This enables an easy calculation by sing input data v. However, the second entry called "negative phase" refers expected value of the entire model to all node (v, h) so that an actual calculation is not possible. Hinton G. E. has proposed Contrastive Divergence (CD) to calculate approximated values even without completely performing the calculation in the second entry [9]. CD calculates the approximated actual distribution by using the distribution produced through Gibbs sampling by n iterations. It has been demonstrated that only 1 Gibbs sampling provides an almost identical result to an actual distribution [10] [11].

In case of RMB's learning, shorter differences between input data and reconstructed data makes better learning on parameters connecting each nodes. When new data are input into the trained RBM, state of each nodes is determined by the probability calculated by weight and bias between each nodes. Like this, the noise and outlier data have difficulty in being activated due to the low probability. Consequently, noises and outliers are considered to be removed in the reconstructed data through RBM.

IV. NETWORK INTRUSION DETECTION SYSTEM BY USING RESTRICTED BOLTZMANN MACHINE

A state of each nodes in RBM is determined by the probability acquired by the calculation of weight and bias between each nodes. In RBM, a state of nodes is a binary value of 0 or 1 so that the data for learning RBM should be converted to binary value. RBM learns with the binary converted data and becomes a trained model when the initially configured the number of learning iterations is over.

Fig 3 shows as actual NIDS designed by using RBM trained by such process. First, in the learning area, labels are separated from labeled data for learning and the data are converted to binary. Then, the leaning for RBM is process with the converted data. After this process, the separated labels are attached again to corresponding reconstructed data. Noises and outliers are removed and previous labels are included again in these data. By using these reconstructed data, the classification algorithm to detect intrusions is trained. In the detection area, actual packet data are immediately converted to binaries after entrance. By testing with the classification algorithm for intrusion detection, normal packets and intrusion packets are separated from these data.

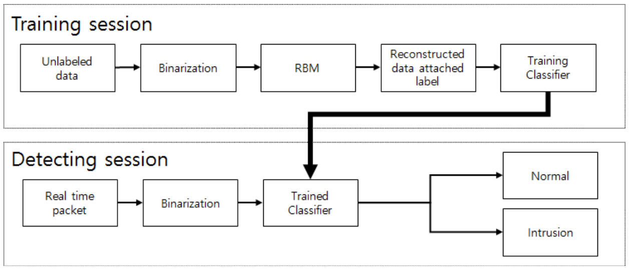

Fig. 3 Proposed Network Intrusion Detection System

The presented NIDS is differed from existing NIDS by adding the process removing noise and outlier data through RBM in the learning area. Also the conversion process of packet data to binary values is added in the detection area.

V. EXPERIMENT AND ASSESSMENT

The experiment is performed through KDD Cup 1999 data [17]. KDD Cup 1999 data are composed of 4 large types of attack classes and the normal class referring normal accesses. DoS is the packet excessively requesting to disable the service and Probe is the packet collecting port and/or other system data before actual attack. R2L refers the packet to acquire an external access authority by an unauthorized user. U2R is the packet that attempts to acquire a root authority by an unauthorized user. Finally, Normal refers the packet that is not attack data, but is data accessing normally. Table 1 shows the outlined characteristics of KDD Cup 1999 data [14].

TABLE I. Class labels and the number of samples which appears in "10% KDD" Data

Class	Attack type	Number of samples
DOS (Denial of Service)	back	2,203
	land	21
	neptune	107,201
	pod	264
	smurf	280,790
	teardrop	979
PROBE	satan	1,589
	ipsweep	1,247
	nmap	231
	portsweep	1,040
NORMAL	normal	97,277
R2L (Remote to Local)	Guess_passwd	53
	ftp_write	8
	imap	12
	phf	4
	multihop	7
	warezmaster	20
	warezclient	1,020
	spy	2

U2R (User to Root)	Buffer overflow	30
	loadmodule	9
	perl	3
	rootkit	10
SUM		494,020

Logistic regression is the classification technique used in this experimentation. The technique is designed to classify any class labeled as normal to Normal and 22 types of attack data to intrusions. The experimentation includes the comparison of classification performances between the classification techniques learning directly using binary converted data and leaning using data noise and outlier removed through RBM. Also, technique performances between different RBMs are compared by adjusting the learning rate and batch size and then, the training data size. The Fig 4 show Pseudo-likelihood values representing differences between energies of input and reconstructed data and performances of trained classification techniques.

TABLE II. Training time according to batch size and learning rate

Training time(sec)		Learning rate	
		0.01	0.1
Batch size	10	958 sec	939 sec
	100	559 sec	565 sec

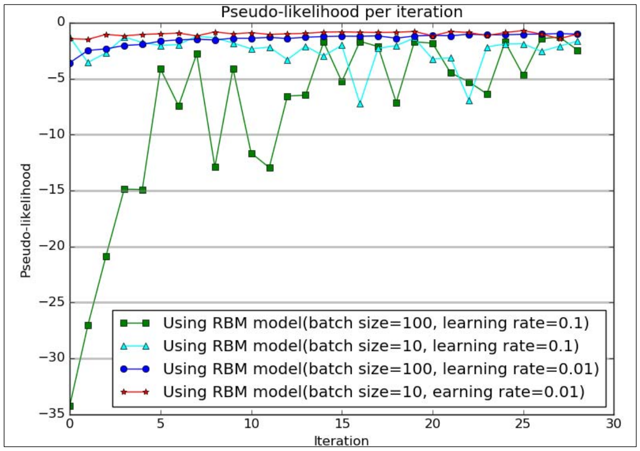

Fig. 4 Pseudo-likelihood according to batch size and learning rate

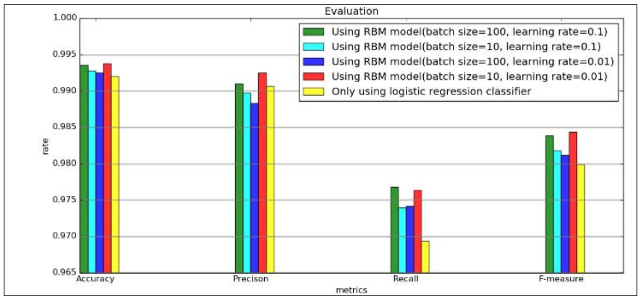

Fig. 5 Accuracy, Precision, Recall, F-measure according to batch size and learning rate

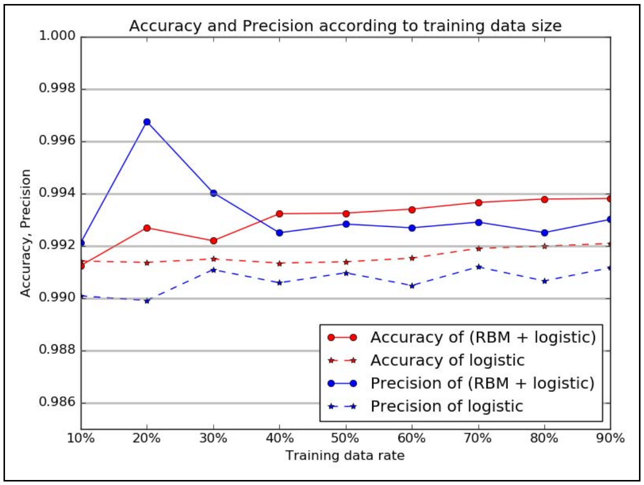

Fig. 6 Accuracy and Precision according to training data size

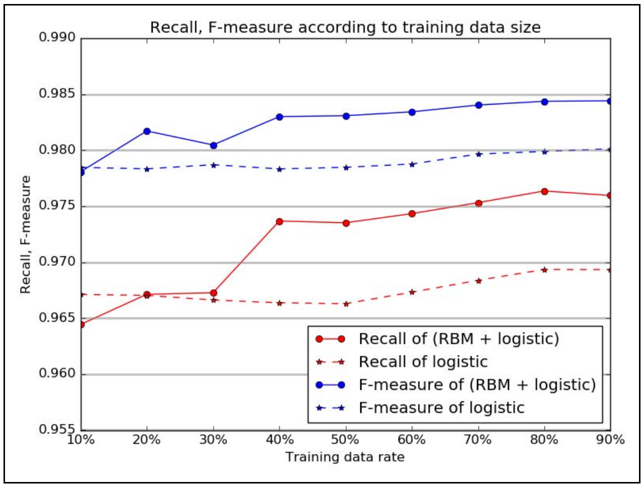

Fig. 7 Recall and F-measure according to training data size

Fig 4 shows the record of each iteration on the energy difference between the input data and reconstructed data presented with pseudo-likelihood by adjusting the learning rate and the batch size. The batch sizes are compared by setting 100 and 10 as recommended by Hinton. G. E. based on empirical knowledge [11]. According to the change in pseudo-likelihood, when the learning rate and the batch size are high, the amplitude is relatively higher. If the change in pseudo-likelihood is high, setting of the number of iteration becomes sensitive so that precautions are required. Especially, in case of learning rate of 0.01, the change in pseudo-likelihood is stable. Like this, the change in pseudo-likelihood is usually influenced by the learning rate than the batch size. Contrastively, Table 2 identifies the batch size has more critical impact than that of the learning rate to the training time of RBM.

Fig 5 shows performance differences between the existing classification model and classification models using RBMs differed in the learning rate and the batch size. In overall, classification models trained with data having noise and outlier removed through RBMs record higher performances. The RBM set with the batch size of 100 and the learning rate of 0.1 should be noticed. This RBM records the most excessive change in the pseudo-likelihood value and, finally, the lowest pseudo-likelihood value, yet, the second highest classification performance. It is demonstrated that such RBM is enabled to record better performances in actual classifications despite the large difference in energy between the input data and the reconstructed data since the data are reconstructed by the probability.

The case of setting the batch size as 10 and the learning rate as 0.01 shows the top performance along with higher performance in all account of Accuracy, Precision, Recall and F-measure than the results of using the simple classification model. Especially, the extensive performance improvement on Recall refers the increase of rate to classify actual normal packets to normal packets. The misuse detection method frequently misses detections of the class labels not included in the signature list. This shows that the recall improvement refers the more efficient learning for a classification model due to removed noises and outliers from existing data by the RBM.

Next, the second experimentation is performed by only adjusting the size of training data on the model set with the batch data of 10 and the learning rate of 0.01 showing the top performance in the above trial. Fig 5 and Fig 6 show a change in Accuracy, Precision, Recall and F-measure according to the training data size. In overall, classification models using RBM show higher performances than that of simple classification. Interestingly, the performances of classification models using RBM relatively soar in the 40% size of training data section. Even though the simple classification is trained by 90% size of training data, the performances of classification models using RBM are higher.

VI. CONCLUSION

RBM is a type of artificial neural networks performing probabilistically unsupervised learning. Noise and outlier data of relatively lower frequencies are activated in lower probabilities. Therefore, being input new data into a trained RBM model, noises and outliers are activated in low probabilities. This study proposes a training method for the classification models to detect network intrusions using the data reconstructed based on such characteristics of RBM.

Through experimentations, the classification model trained with the reconstructed data through RBM is verified to record the higher performance in network intrusion detections than that of the general classification model. Especially, the classification performance is improved more through proper setting of the data size used to train RBM and parameters such as the batch size and the learning rate. Like this, this study shows possibilities of noise and outlier removals in data by using RBM mostly used in the deep learning area as well as possibilities to improve performances of classification models by applying to general classification techniques.

However, parameters of the batch size, the learning rate, the number iteration in previous studies are empirical and variable factors so that follow-up researches are required. Also, performance improvements on network intrusion detection systems through comparisons and/or combinations of PCA or LDA, and RBM can be anticipated.

REFERENCES

[1] Ackley, D. H., Hinton. G. E, and Terrence J. Sejnowski., "A learning algorithm for Boltzmann machines.", *Cognitive science* 9(1), 1985, pp. 147-169.

[2] A.K. Ghosh, A. Schwartzbard, “A study in Using Neural Networks for Anomaly and Misuse Detection.” USENIX Security, 1999.

[3] Bouzida, Y., Cuppens, F., Cuppens-Boulahia, N., & Gombault, S., “Efficient intrusion detection using principal component analysis” *3ème Conférence sur la Sécurité et Architectures Réseaux (SAR), La Londe, France*. 2004, pp. 381-395.

[4] C. Kruegel, D. Mutz, W. Robertson, F. Valeur, . "Bayesian event classification for intrusion detection.", *Computer Security Applications Conference, 2003. Proceedings. 19th Annual*. IEEE, 2003, pp. 14-23.

[5] Denning, D. E., "An intrusion-detection model.", *IEEE Transactions on software engineering, (2)*, 1987, pp. 222-232

[6] Fiore, U., Palmieri, F., Castiglione, A., & De Santis, A., "Network anomaly detection with the restricted Boltzmann machine.", *Neurocomputing* 122, 2013, pp.13-23.

[7] Fischer, Asja., Christian Igel., "An introduction to restricted Boltzmann machines.", *Iberoamerican Congress on Pattern Recognition,* 2012, pp.14-36

[8] Han, H., Lu, X. L., Lu, J., Bo, C., & Yong, R. L., "Data mining aided signature discovery in network-based intrusion detection system.", *ACM SIGOPS Operating Systems Review* 36(4), 2002, pp. 7-13.

[9] Hinton, G. E., "Training products of experts by minimizing contrastive divergence.", *Neural computation* 14(8), 2002, pp.1771-1800.

[10] Hinton, G. E., Osindero, S., & Teh, Y. W., "A fast learning algorithm for deep belief nets.", *Neural computation,* 18(7), 2006, pp.1527-1554..

[11] Hinton, G. E., "A practical guide to training restricted Boltzmann machines.", *Momentum,* 9(1), 2010, pp.3-20.

[12] Mukkamala, S., Janoski, G., & Sung, A., "Intrusion detection using neural networks and support vector machines.", *Neural Networks,* 2002, pp.1702-1707.

[13] Lazarevic, A., Ertöz, L., Kumar, V., Ozgur, A., & Srivastava, J., “A Comparative Study of Anomaly Detection”, Schemes in Network Intrusion Detection. In *SDM, 2003,* pp. 25-36.

[14] Olusola, A. A., Oladele, A. S., & Abosede, D. O., “Analysis of KDD’99 Intrusion detection dataset for selection of relevance features.”, *Proceedings of the World Congress on Engineering and Computer Science* Vol. 1, 2010, pp.20-22.

[15] S. Chavan, K. Shah, N. Dave, S. Mukherjee, A. Abraham, “A study in Using Neural Networks for Anomaly and Misuse Detection”, Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference on. Vol. 1. IEEE, 2004, pp.70-74.

[16] Tan, Z., Jamdagni, A., He, X., & Nanda, P., "Network Intrusion Detection based on LDA for payload feature selection." *2010 IEEE Globecom Workshops*. IEEE, 2010, pp.1545-1549.

[17] Stolfo, S. J. "KDD cup 1999 dataset." UCI KDD repository. http://kdd. ics. uci. Edu, 1999.