IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

Intrusion Detection System Using K-Means and Edited Nearest Neighbour Algorithm.

Mr. Abdul Khadar A[1], Modem Tharun Kumar[2], Sharath K N[3], Sukesh V N[4], Tejaswini K N[5]
Assistant Professor, Department of Information Science and Engineering[1]
Students, Department of Information Science and Engineering[2,3,4,5]
S.J.C. Institute of Technology, Chikkaballapur, Karnataka, India

Abstract: *In imbalanced network traffic, malicious cyber-attacks can often hide in large amounts of normal data. It exhibits a high degree of stealth and obfuscation in cyberspace, making it difficult for Network Intrusion Detection System (NIDS) to ensure the accuracy and timeliness of detection. This paper researches machine learning and deep learning for intrusion detection in imbalanced network traffic. It proposes a novel Difficult Set Sampling Technique (DSSTE) algorithm to tackle the class imbalance problem. First, use the Edited Nearest Neighbor (ENN) algorithm to divide the imbalanced training set into the difficult set and the easy set. Next, use the K- Means algorithm to compress the majority samples in the difficult set to reduce the majority. Zoom in and out the minority samples' continuous attributes in the difficult set synthesize new samples to increase the minority number. Finally, the easy set, the compressed set of majority in the difficult, and the minority in the difficult set are combined with its augmentation samples to make up a new training set. The algorithm reduces the imbalance of the original training set and provides targeted data augment for the minority class that needs to learn. It enables the classifier to learn the differences in the training stage better and improve classification performance. To verify the proposed method, we conduct experiments on the classic intrusion dataset NSL-KDD. We use classical classification models: random forest(RF), Support Vector Machine (SVM), XGBoost, Long and Short- term Memory (LSTM), Adaboost, AlexNet, Mini- VGGNet.*

Keywords: IDS, Imbalanced Network traffic, Machine Learning, Deep Learning

REFERENCES

[1]. D. E. Denning, ''An intrusion-detection model,'' IEEE Trans. Softw. Eng., vol. SE-13, no. 2, pp. 222– 232, Feb. 1987.
[2]. N. B. Amor, S. Benferhat, and Z. Elouedi, ''Naive Bayes vs decision trees in intrusion detection systems,'' in Proc. ACM Symp. Appl. Comput. (SAC), 2004, pp. 420–424.
[3]. M. Panda and M. R. Patra, ''Network intrusion detection using Naive Bayes,'' Int. J. Comput. Sci. Netw. Secur., vol. 7, no. 12, pp. 258–263, 2007.
[4]. M. A. M. Hasan, M. Nasser, B. Pal, and S. Ahmad, ''Support vector machine and random forest modeling for intrusion detection system (IDS),'' J. Intell. Learn. Syst. Appl., vol. 6, no. 1, pp. 45–52, 2014.
[5]. N. Japkowicz, ''The class imbalance problem: Significance and strategies,'' in Proc. Int. Conf. Artif. Intell., vol. 56, 2000, pp. 111–117.
[6]. Y. LeCun, Y. Bengio, and G. Hinton, ''Deep learning,'' Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[7]. Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and
[8]. M. S. Lew, ''Deep learning for visual understanding: A review,'' Neurocomputing, vol. 187, pp. 27–48, Apr. 2016.
[9]. T. Young, D. Hazarika, S. Poria, and E. Cambria, ''Recent trends in deep learning based natural language processing [review article],'' IEEE Comput. Intell. Mag., vol. 13, no. 3, pp. 55–75, Aug. 2018.
[10]. N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, ''A deep learning approach to network intrusion detection,'' IEEE Trans. Emerg. Topics Comput. Intell., vol. 2, no. 1, pp. 41–50, Feb. 2018.
[11]. D. A. Cieslak, N. V. Chawla, and A. Striegel, ''Combating imbalance in network intrusion datasets,'' in Proc. IEEE Int. Conf. Granular Comput., May 2006, pp. 732–737.
[12]. M. Zamani and M. Movahedi, ''Machine learning techniques for intrusion detection,'' 2013, arXiv:1312.2177.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

[Online]. Available: http://arxiv. org/abs/1312.2177

[13]. M. S. Pervez and D. M. Farid, ‘‘Feature selection and intrusion classification in NSL-KDD cup 99 dataset employing SVMs,’’ in Proc. 8th Int. Conf. Softw., Knowl., Inf. Manage. Appl. (SKIMA), Dec. 2014,pp.1–6