Ilkogretim Online - Elementary Education Online, 2021; Vol 20 (Issue 3): pp. 4214-4219
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2021.03.435

Intrusion Detection System Using Recursive Feature Elimination And Rnn

Poonam Verma Department of Computer Application Graphic Era Hill University, Dehradun, Uttarakhand India, 248002 pverma@gehu.ac.in

NOOR MOHD Department of Computer Science & Engineering, Graphic Era Deemed to be University, Dehradun, Uttarakhand India, 248002 noormohdcs@gmail.com

Abstract: In recent years, there has been a perceptible rise in the number of attacks that involve breaking into computer networks, which is cause for big factor from both a privacy and a security standpoint. The proliferation of new technologies has led to an increase in the sophistication of cyber-security breaches, to the point that the currently available monitoring tools are unable to adequately handle the problem. In consideration of this, the installation of a network intrusion detection system that is both intelligent and efficient would be absolutely necessary in order to resolve this issue. In this paper, we proposed a model using Recursive Feature Elimination and Ensemble Learning. Features are extracted using Recursive Feature Elimination and RNN. Furthermore, we compare the outcomes of our proposed solution with those of other suggested policies in an effort to identify the method that delivers the most appropriate method for the intrusion detection systems, and we assess the effectiveness of the suggested solution using a number of evaluation matrices.

Keywords: Recurrent Neural Network, Intrusion Detection, Recursive Feature Elimination.

I. INTRODUCTION

The safety of networks and sensitive data is of critical importance to today's expanding economy. When it comes to personal computer network security, just the installation of antivirus and firewall software is done. Maintaining a secure network, however, is no easy undertaking for a company. It demands modern methods of attack and the ability to process massive amounts of data. The safety of networks and sensitive data is a major issue for the expanding economy. In order to protect the network, users typically install security software like antivirus and firewall on a personal computer. However, managing network security is no easy chore for a company. Not only does it necessitate sophisticated new forms of assault, but it can also process massive amounts of data. Possible kinds of intrusion detection include signature-based and profile-based attacks. An attack that is based on a signature can identify all of the predefined attacks. The signature-based files are mapped with the attacks, and the system will only return the appropriate attack type if a match is found. It is important to remember to check for

anomaly-based infiltration, though. Because there is a signature-based file available, the percentage of false positives is relatively low.

Attacks that are profile-based are also known as anomaly-based assaults, and they are distinguished from other types of attacks in that they do not follow a path that has been set. The intrusion detection system (IDS) that is utilised to identify this form of attack needs to be adaptable enough to deal with anonymous scenarios. It has a high percentage of false positives. Intrusion Detection Systems, are an additional layer of defence that can be added to computer networks. The security of a computer system can be put at risk by certain types of malicious behaviour, which can be identified with the use of intrusion detection systems (IDS). These kinds of activities include attacks on networks that are using services that are vulnerable, attempts to escalate privileges, illegal access to sensitive files, and the usage of malicious malware (computer viruses, Trojans, and worms). One of the most important aspects of an effective intrusion detection system (IDS) is its level of precision. In order to obtain the highest possible level of accuracy in intrusion detection, a data collection of high quality is required. The feature selection process is one method that can be utilised to acquire a dataset of superior quality. The selection of features is an essential stage in the majority of classification tasks, since it shortens the amount of time spent learning and increases the accuracy of predictions.

There have been several studies on machine learning that have led to the development of solutions that use machine intelligence to identify intrusions. In the realm of intrusion detection, for example, successful applications include the support vector machine (SVM), artificial neural networks (ANNs), and genetic algorithms (GAs). The straightforward approach to machine learning, on the other hand, has a number of drawbacks, whereas intrusion is growing increasingly intricate and diverse. It is necessary to improve learning approaches, particularly with regard to the automatic extraction and analysis of intrusion features. Deep Learning Network topologies like CNN and RNN have the ability to overcome the difficulties of currently available classifiers. Because of their great accuracy and enhanced performance, Deep Learning Networks may also play a crucial role in Network Intrusion Systems. As a result of its useful features and proven track record in object classification, RNN has been adopted for our use. The ability to recall past events was a major factor in selecting RNN for this task, since it may help immensely when determining the fallout from a wide variety of attacks.

II. RELATED STUDY

On the subject of intrusion detection, there have been a significant number of previous studies. In most cases, these studies consist of two stages: the preprocessing stage and the classification stage. The preprocessing method, which is sometimes referred to as the feature selection process, is an essential step in the intrusion detection procedure. It is able to isolate the most important aspects of the raw data, which can have a significant bearing on the findings. Additionally, it can lessen the amount of space required for data storage while simultaneously enhancing the effectiveness of model training and the precision of classifiers. Despite the fact that the raw data has a high dimension, it

performs effectively in algorithms with a modest input scale. The selection of features can be broken down into two different methods: filter and wrapper, depending on whether or not the feature is dependent on the classifier. The classifier method is not dependent on the filter approach. It chooses features based on the statistical characteristics of all the raw data that are available.

An effective filter approach for the classification of network traffic was proposed by Shi et al. [1]. They extracted multifractal features by using Wavelet Leaders Multifractal Formalism, and then they removed unnecessary and superfluous features by using PCA-based FS approach. The findings revealed a considerable improvement in accuracy when compared to previous ML-based methods. The backpropagation algorithm also chooses features to use in the model based on the classifier's predictive performance. This means that the wrapper technique would require more time spent training, but would yield better results overall. However, it is hard to guarantee characteristics of the various classifiers [2] since the overall classification accuracy of the extracted features chosen using wrappers is largely reliant on the individual classifier. Following data preprocessing, a classifier is needed for traditional intrusion detection. Network traffic categorization has been applied to a wide variety of machine learning methods, including but not limited to: support vector machines [3], decision trees [4], artificial neural networks [5], Naive Bayesians [6], fuzzy logic [7], generic algorithms [8], and K-nearest neighbour classifiers [9]. Staudemeyer [10] claims that successful deployment of a long short-term memory (LSTM) recurrent neural network to intrusion monitoring has been achieved. The Long Short-Term Memory (LSTM) algorithm can be taught to use a historical perspective to make connections. The recommended classifier was successful in recognising denial of service (DOS) assaults and network probes, both of which comprised unique time series of events, and both of which could be separated from one another.

III. PROPOSED METHOD

The proposed method is divided into two modules: Recurrent Neural Network and Recursive Feature Elimination.

1. **Recurrent Neural Network**
   The notion of Deep Neural Networks architectures, more specifically RNN, is utilised in the suggested system for the detection of intrusions. As a result of the RNN cybersecurity enabled network topology, the issue of identifying and classifying various threats can finally be put to rest. In this work, we use two distinct RNNs algorithms: long short-term memory (LSTM) and gated recurrent units (GRUs). These are two applications of recurrent neural networks (GRU). Single information points, as well as entire sequences, are no problem for LSTM. It's also helpful because it doesn't care how big the separation is. Additionally, LSTM variants such as Gated Recurrent Units (GRU) are nearly identical to LSTM layers. GRU layers, in contrast to LSTM layers, contain fewer parameters and no outputting gates. The CNN only looks at the most recently input, resulting in it

being also known as a feed forward neural network. Evidently, the design's many levels each have their own functionality, which finds the hidden units and extracts features. For this work, we employ a technique called Recursive Feature Removal, which allows us to choose and extract features automatically.

The RNN architecture makes use of two LSTM layers, each of which can recall its inputs for a considerable amount of time. Additionally, a dropout layer is used after each LSTM layer, followed by a thick layer and then an activation layer. For the GRU method, the only difference is that the LSTM layers have been swapped out for GRU ones in the same basic architecture.

2. **Recursive Feature Elimination**

One kind of feature selection is known as Recursive Feature Elimination. This method that gets rid of features in a cyclical fashion; it creates a prototype that, by making use of the remaining parameters, determines how accurate it is. The Recursive Feature Elimination (RFE) technique analyses the configuration of features in order to make an accurate prediction of the intended output. RFE, or sequential backward reduction, is a technique for picking features that suit a model and removing the characteristic (or features) that represent the weakest one at time until the required number of characteristics has been attained. Recursive feature removal is often commonly referred to as RFE. The coefficients of the framework or the correlation - based feature attributes are employed to rank this same feature, and the RFE methodology aims to remove any interconnections and collinearity that could be present in the prototype by iterative manner deleting a specific feature with each iterative process of the loop. This is done in order to reduce the likelihood that the model will produce incorrect results. The number of valid characteristics is not always known in advance, but RFE insists on keeping a certain minimum number of them. Cross-validation with RFE is used to evaluate many feature subsets and pick the highest-scoring set of features in order to determine the ideal feature set size.

IV. EXPERIMENTAL ANALYSIS

It is necessary to have a large dataset of high quality for the Deep Learning IDS implementation in order to be able to train and test an algorithm in a simulation of the conditions that exist in the actual world. In this study, the NSL-KDD data set that is provided by the Canadian Institute for Cybersecurity is utilised not only for the purpose of training but also of assessment. Because there is no repeat in this dataset between the data used for training and the data used for testing. Due to the reasonable size of the dataset, the complete set can be used for the experiment rather than just a subset of it. This improves the reliability of the experiment. Approximately 85% of the dataset is reserved for training purposes, while the remaining 15% is used for testing and validation. There are 41 features per record in the NSL-KDD that represent different characteristics, as well as a label that indicates whether the traffic was malicious or not.

The evaluation of the functioning of the model that uses deep learning the accuracy, recall, precision, and F1 score were the four primary evaluation metrics that were used in the computing process throughout the testing phase.

The process of selecting the features that have the potential to produce the most accurate results or the output that is desired is referred to as feature selection. This removal of irrelevant features that have the potential to decrease the accuracy of our system can take place either manually or automatically. Feature selection contributes to the distinctive description of patterns from a variety of classes. In our proposed method feature selection is done by recursive feature elimination method and then the selected feature are passed to RNN for further classification. Proposed model is compared with CNN, LSTM, CNN-LSTM and SVM-CNN to evaluate its performance as shown in Table 1.

Method	**Precision**	**Recall**	**Accuracy**
CNN	95.34	95.34	96.89
LSTM	96.01	96.89	96.99
CNN-LSTM	95.99	96.12	96.774
SVM-CNN	95.99	95.12	95.95
Proposed model	97.01	97.34	97.90

Table: Model Comparison

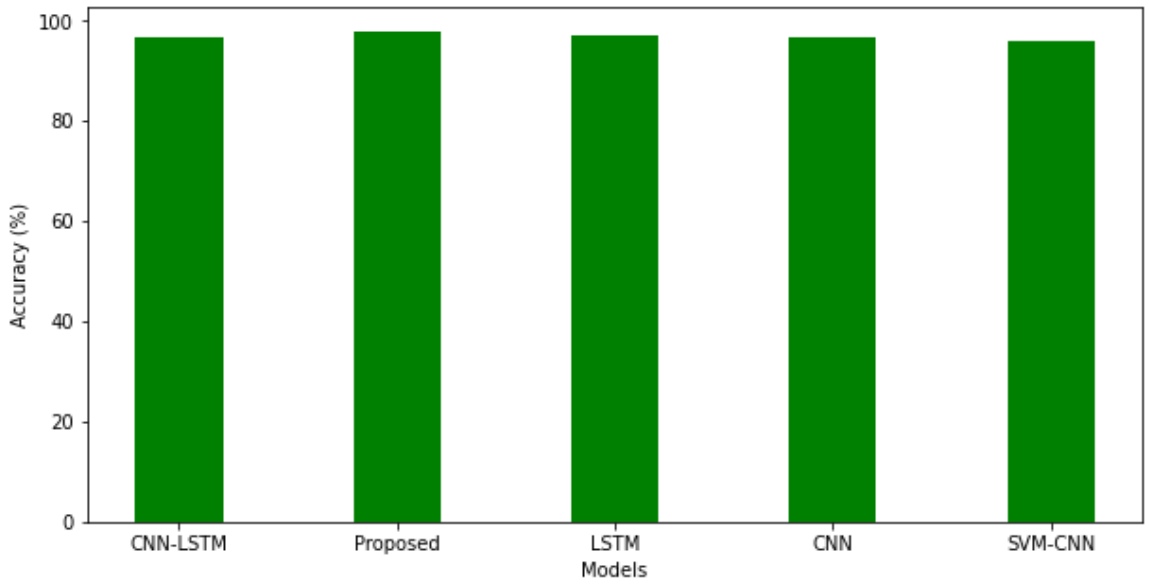

Figure 1: Accuracy Comparison with Proposed Model

From Table 1 and Figure 1, it's obvious that our conceptual model outperforms competing models. In addition, it was found that a lot of computational resources was needed for the development of the models also during empirical stage of the research. In order to run the computation for 1000 epochs on both machines, we had to make some changes to the learning phase, specifically by raising the packet size of the simulations

from 32 to 64. This led to a more homogenous distribution of outcomes and helped speed up the development process.

V. CONCLUSION

We implemented an intrusion detection system in the presented study. We trained and tested this approach on RNN. We also discovered that certain features in the dataset were useless and redundant. As a result, recursive feature reduction played a significant role in lowering the dataset's dimensionality. The current study used the NSL-KDD dataset for training and testing to apply the technique of anomaly-based intrusion detection utilising Recursive Feature Elimination and RNN with LSTM. One of the most significant challenges that was encountered was the absence of sufficient computational resources, which resulted in a prolonged period of training for model.

REFERENCES

1. H. Shi, H. Li, D. Zhang, C. Cheng, and W. Wu, ‘‘Efficient and robust feature extraction and selection for traffic classification,’’ Comput. Netw. Int. J. Comput. Telecommun. Netw., vol. 119, pp. 1–16, Jun. 2017
2. J. Shen, J. Xia, X. Zhang, and W. Jia, ‘‘Sliding block-based hybrid feature subset selection in network traffic,’’ IEEE Access, vol. 5, pp. 18179–18186, 2017.
3. S. M. H. Bamakan, H. Wang, T. Yingjie, and Y. Shi, ‘‘An effective intrusion detection framework based on MCLP/SVM optimized by timevarying chaos particle swarm optimization,’’ Neurocomputing, vol. 199, pp. 90–102, Jul. 2016.
4. A. S. Eesa, Z. Orman, and A. M. A. Brifcani, ‘‘A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems,’’ Expert Syst. Appl., vol. 42, no. 5, pp. 2670–2679, 2015
5. A. Sharma, I. Manzoor, and N. Kumar, ‘‘A feature reduced intrusion detection system using ANN classifier,’’ Expert Syst. Appl., vol. 88, pp. 249–257, Dec. 2017.
6. L. Koc, T. A. Mazzuchi, and S. Sarkani, ‘‘A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier,’’ Expert Syst. Appl., vol. 39, no. 18, pp. 13492–13500, 2012.
7. P. R. K. Varma, S. S. Kumari, and V. V. Kumar, ‘‘Feature selection using relative fuzzy entropy and ant colony optimization applied to real-time intrusion detection system,’’ Procedia Comput. Sci., vol. 85, pp. 503–510, Dec. 2016
8. F. Kuang, W. Xu, and S. Zhang, ‘‘A novel hybrid KPCA and SVM with GA model for intrusion detection,’’ Appl. Soft Comput., vol. 18, pp. 178–184, May 2014.
9. A. A. Aburomman and M. B. I. Reaz, ‘‘A novel SVM-kNN-PSO ensemble method for intrusion detection system,’’ Appl. Soft Comput., vol. 38, pp. 360–372, Jan. 2016.
10. R. C. Staudemeyer, ‘‘Applying long short-term memory recurrent neural networks to intrusion detection,’’ South Afr. Comput. J., vol. 56, no. 1, pp. 136–154, 2015.