International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

A Neural Network Model for Attacker Detection using GRU and Modified Kernel of SVM

Sharfuddin Waseem Mohammed, C Madan Kumar, Narasimha Reddy Soora

Abstract: ***over past few decades neural network changed the way of traditional computing many different models has proposed depending upon data intensity, predictions, and recognition and so on. Among which Gated Recurrent Unit (GRU) is created for variety of long short-term memory (LSTM) unit, which is part of recurrent neural network (RNN). These models proved to be dominant for range of machine learning job such as predictions, speech recognition, sentiment analysis and natural language processing. In this proposed model, a support vector machine (SVM) with modified kernel as final output layer for prediction is used instead of traditional approach of softmax and log loss function is used to calculate the loss. Proposed technique is applied for binary classification for intrusion detection using honeypot dataset (2013) network traffic sequence of Kyoto University. Results shows a prominent change in training efficiency of ≈89.45% and testing efficiency of ≈88.15% when compared with softmax output layer. We can conclude that linear SVM with modified kernel as output layer outperform compared with softmax in prediction time.***

Keywords: Gated Recurrent Unit (GRU), Long Short Term Memory (LSTM), Support Vector Machine (SVM), Recurrent Neural Network(RNN)

I. INTRODUCTION

According to cybercrime report of juniper research [1] by the end of this year 2019 will increase the cost of data breach to $2.1 trillion increasing almost four times compared to 2015, due to increase of intruders in network. To detect the intruder the most common mechanism is to identify the behavior and activities of user [2]. For analyzing the user activities many administrative tools are available [3], much more research is conducted on profiling and identifying user activities [4], which is tedious task to manually identify the intruder hence a deep learning principle can be applied to detect the intruder. A neural network model is computational model that resembles the human brain with neuron interconnection [5]. According to study by Mukkamal Janoski & Sung (2002) [6] Support Vector Machine and Artificial Neural Network can be used to solve the intruder detection. In machine learning SVM can be used to solve the two classes of data points using hyper plane. Alalshekmubarak & Smith [7] combined ANN and SVM for time series classification especially for combined echo state network (ESN, a variety of recurrent neural network) and SVM. In this paper we propose a modified SVM combined with Gated Recurrent Unit (GRU) in place of ESN. Basically RNN is used to predicting sequential or time series data, here we used network traffic data which is sequential to predict intruder.

In our propose method GRU is used to predict the network data and at output layer we use modified kernel SVM with log loss function. Which shows effective results in training and testing are fine and satisfactory.

Revised Manuscript Received on 30 July 2019.

***** Correspondence Author

Sharfuddin Waseem Mohammed*, CSE department, Kakatiya Institute of Technology & Sciences, Warangal, India. Email: waseem7602@gmail.com

C Madan Kumar , CSE department, Kakatiya Institute of Technology & Sciences, Warangal, India. Email: madan.kumar547kitsw@gmail.com

Narasimha Reddy Soora, CSE department, Kakatiya Institute of Technology & Sciences, Warangal, India. Email: 12dt04cse005@cse.vnit.ac.in

II. METHODOLOGY

Following steps has followed for network intruder detection.
Step1: Preprocessing the data.

Step2: Applying GRU and SVM Training on data

Step3: Testing and validating the data.

Dataset: The Kyoto University honeypot system's network traffic data [8] is used to detect the intruders, it consist of 24 statistic features are available, 14 are related to network instance from KDD Cup 1999 dataset[9] and 10 additional features which are more important to detect intrusion.

Step1: Preprocessing the data.

Dataset consist of 16.5 GB of network traffic data from which only 25% is used approx. 4 GB to evaluate our proposed system (from January 1, 2013 to June 1, 2013).

We normalized the data, standardization for continuous data, indexing is performed then it is binned for discretization.

- $Z=(X-\mu)/\sigma$ (1)

Where X is the feature value to be standardize, μ is the mean of feature values and σ is the standard deviation.

For indexing the categories are mapped to [0,n-1], after dataset normalization continuous features are binned with a quantile of the features. Binning reduces the computation effort and improve the performance of training.

Step 2: Applying GRU and SVM on data.

Similar to Alalshekmubarak & Smith(2013)[7] we also used kernel modified SVM for classifier with Gated Recurrent Unit(GRU) for training.

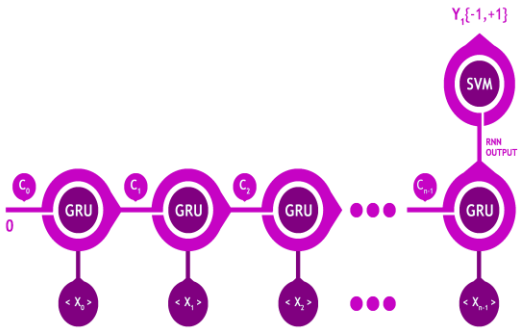

Figure 1: The proposed GRU-SVM model with n-1 GRU units and SVM as classifier.

In the above architecture 21 features are used as input to GRU units and kernel modified SVM is used as classifier.

An unfolded GRU unit computes the output as follows:

$z = \sigma(W_z \cdot [h_{t-1}, x_t])$ (2)

h_{t-1} is the previous state of the GRU unit but for very first input its value will be zero.

$m = \sigma(W_r \cdot [h_{t-1}, x_t])$ (3)

$\check{h}_t = \tanh(W \cdot [m_t * h_{t-1}, x_t])$ (4)

$h_t = (1 - z_t) * h_{t-1} + z_t * h_t$ (5)

h_t will be the final output for the GRU unit and this is fed as input to SVM.

In our proposed work a modified kernel linear SVM is used as output layer to classify the output, parameters are tuned to optimize the objective function of SVM. Then, instead of traditional cross entropy function to measure the network loss we used the log loss function which yields to better result. Following equation of L2-SVM is used which is modified kernel of SVM.

$\min * 0.5 \|w\|_2 + C(\text{Sum}(\max(0, 1 - \hat{y}_i (W^T * 0.05 * x_i + b_i)^2)$ (6)

This equation-6 is modified version of L2-SVM. As for the prediction the function sign(Wx+b) will produce an score vector for each classes to get the predicated class label y of a data.

Predicted class =argmax(sign(wx+b)) (7)

Argmax can be used to return the maximum score among the predicted classes.

Loss of the trained network is calculated using the log loss equation as follows.

$L_{log}(y,z) = \ln(1 + e^{-yz})$. (8)

Log loss SVM is used for imbalanced data, here in our model network data is imbalance, hence results are shown effective with this function.

We can summarize the above SVM process with following procedure.

1. Input the preprocessed data to the GRU model.
2. Initialize the hyper-parameters of the network like weight and bias.
3. The cell state of GRU is computed using the equations 3, 4, 5 with respect to input feature and learning rate parameter.
4. At the last time step, the prediction of the model is computed using the equation 7.
5. Loss of the network is computed using the log loss function of equation 8.
6. An optimizer is used to minimize the loss by tuning or adjusting the weights and biases.
7. This process is repeated until a maximum training accuracy is reached.
8. Above trained model can be used for binary classification for given data.

Table 1 Describe the Hyper parameters values.

Hyper Parameter name	GRU-Values
Batch size	256
Cell size	256
Epochs	10
Dropout Rate	0.85
Learning rate	$1e^{-6}$
SVM Constant	0.45

Step 3: Testing and validating the data

Data is being classified as training data and testing data with a ratio of 80:20.

After performing the data preprocessing the duplicate points are removed and network is trained of unique data of about approx. 5 lacks of record (nearly 40 MB of data).

Following parameters are used to evaluate the proposed network.

1. Accuracy
2. Epochs
3. Learning rate
4. Loss
5. Run time
6. Number of data points
7. Number of false positive
8. Number of false negative.

III. RESULT ANALYSIS

After training the network it is evaluated for prediction of intruder with testing with validate data and test data.

Proposed method is evaluated on MacBook air 1.8GHz, Dual-core Intel core i5, 8GB RAM, Intel HD Graphics 6000.

Following table demonstrate the evaluation of network.

Table: 2 Summary showing the parameter evaluation from GRU-SVM modified kernel to GRU-softmax approach.

Parameter	GRU-SVM (Modified)	GRU-softmax
No. of input data points-training	18,98,240	18,98,240
No. of data points-testing	4,20,608	4,20,608
Epochs	10	10
Accuracy-Training	≈ 88.34%	≈ 63.7%
Accuracy-Testing	≈ 87.32%	≈ 70.35%
Loss	≈ 0.91%	≈ 0.61%
No. of false positives-Training	8,90,525	25,58,894
No. of false positives- Testing	1,80,585	4,75,989
No. of false negatives- Training	8,29,555	4,87,242
No. of false negative Testing	1,40,878	5,82,878

From the above table we can observe that proposed technique work better than GRU-softmax. Here we consider the different parameters to evaluate both models, input data points are same for both models i.e. 18,98,240 and for testing we consider 4,20,608 data points.

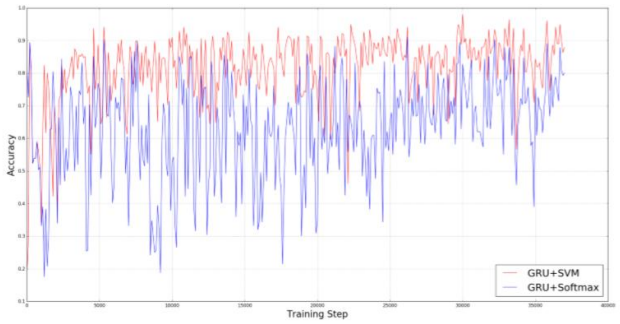

Figure 2: Training Accuracy of the proposed model GRU+SVM with conventional GRU+softmax.

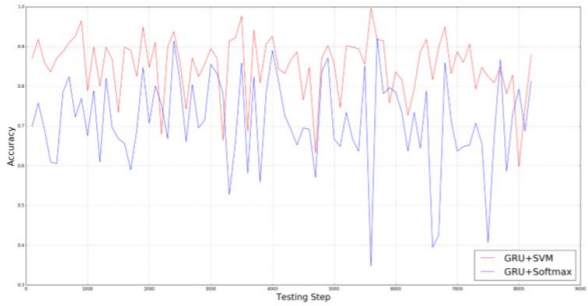

Figure 3 Testing Accuracy of the proposed model GRU+SVM with conventional GRU+softmax.

To support the result we can justify why our proposed model performed better than GRU-softmax.
Softmax is best suited for multinomial classification whereas SVM is for binary classification, SVM doesn't care about the individual score of data points hence consider the batch score. GRU-softmax doesn't perform well on binary classification due to vanishing gradient problem still it consist of LSTM unit inside, it make high probability of misclassification.

IV. CONCLUSION AND FUTURE SCOPE

We proposed a model of modified SVM kernel as an output layer to GRU for binary classification problem, were we used log-loss function instead of hing-loss. Experiment is conducted with approx. 4GB of network traffic dataset which performs better results for both training and testing.

In future scope we can apply this modified SVM kernel of other binary classification and comparative result can be evaluate to prove a strong support towards proposed system

Acknowledgement: I specially thanks to my mentor Dr. B. Janet, Assistant Professor, Department of CA, National Institute of Technology, Tiruchchirappalli to be a mentor to conduct this experiment.

REFERENCES:

1. Juniper. May 12, 2015. Cybercrime will cost Businesses over $2 Trillion by 2019. https://www.juniperresearch.com/press/press-releases/ cybercrime-cost-businesses-over-2trillion. (May 12, 2015). Accessed: May 6, 2017.
2. J. Cannady. Artificial neural networks for misuse detection. In Proceedings of the 1998 National Information Systems Security Conference (NISSC'98), pages 443{456, October 5-8, 1998. Arlington, VA.
3. https://www.cse.wustl.edu/~jain/cse56706/ftp/net_perf_monitors1.pdf
4. M. Tao, Y. C. Ming and C. Juan, "Profiling and identifying users' activities with network traffic analysis," 2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 2015, pp. 503-506.
5. M.Negnevitsky.2011. Artificial Intelligence: A Guide to Intelligent Systems (3rd ed.). Pearson Education Ltd.,Essex, England.
6. Ryan J, Lin M-J, Miikkulainen R (1998) Intrusion Detection with Neural Networks. Advances in Neural Information Processing Systems 10, Cambridge, MA: MIT Press
7. A.Alalshekmubarak and L.S.Smith.2013. A Novel Approach Combining Recurrent Neural Network and Support Vector Machines for Time Series Classification. In Innovations in Information Technology (IIT), 2013 9th International Conference on IEEE,42–47.
8. Jungsuk Song, Hiroki Takakura, and Yasuo Okabe. 2006. Description of kyoto university benchmark data. Available at link: http://www.takakura.com/Kyoto_data/BenchmarkDataDescriptionv5.pdf. [Accessed on 15 March 2016] (2006).
9. J Stolfo, Wei Fan, Wenke Lee, Andreas Prodromidis, and Philip K Chan. 2000. Cost based modeling and evaluation for datamining with application to fraud and intrusion detection. Results from the JAM Project by Salvatore (2000).

AUTHORS PROFILE

Sharfuddin Waseem Mohammed Working as Assistant Professor, Department of CSE, received his post-graduate degree (MTech(CSE)) from KU, Warangal, Telangana, India in 2015 and his under-graduate degree (BE(CSE)) from JNTU, Hyderabad, Telangana, India in 2009. Currently, he is pursuing his PhD from NIT-Trichy, Tamilnadu, India. He has a total of 10 years' experience in academic, area of interest image processing, Deep Neural Networks, Pattern Recognition.

C Madan Kumar Working as Assistant Professor, Department of CSE, received his post-graduate degree (MTech(SE)) from JNTU, Hyderabad, Telangana, India in 2012 and his under-graduate degree (B.Tech(CSE)) from JNTU, Hyderabad, Telangana, India in 2004. Currently, he is pursuing his PhD from NIT-Trichy, Tamilnadu, India. He has a total of 12 years' experience, area of interest image processing, Multimedia Security.

Dr. Narasimha Reddy Soora PhD from VNIT, Nagpur, Maharashtra, India. And his post-graduate degree (MTech(CSE)) from JNTU, Hyderabad, Andhra Pradesh, India in 2007 and his under-graduate degree (BE(CSE)) from Osmania University, Hyderabad, Andhra Pradesh, India in 1999. Currently, he is pursuing his PhD from VNIT, Nagpur, Maharashtra, India. He has a total of 12 years experience, nine years of which were in the IT industry with three years as an academic. He is life member of ISTE, CSI, India and a Fellow of IETE, India.