Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 2 : 2024
ISSN :**1906-9685**

Journal of Nonlinear Analysis and Optimization : Theory & Applications
Editors-in-Chief :

A DEEP LEARNING MODEL FOR DETECTING PHISHING EMAILS WITH AN IMPROVED ATTENTION MECHANISM AND MULTILEVEL VECTORS

P Manjulatha PG Scholars, Department of CSE, Sree Dattha Institute of Engineering and Science, Sheriguda, Hyderabad**.**
Dr S V Achutha Rao Dean Academics & Professor, Department of CSE, Sree Dattha Institute of Engineering and Science, Sheriguda, Hyderabad**.**

Abstract: *Nowadays, one of the most prevalent forms of cybercrime is phishing emails, which may steal important information such as login credentials, credit card numbers, and more. There has to be an introduction of excellent phishing detection technology since the number of phishing emails is increasing. There are a plethora of new approaches, strategies, and technology that have been developed to identify phishing emails. The use of an attention mechanism with multilevel vectors to improve RCNN is one example of such a technique. Prior to detecting attention mechanism phishing content, the email's structure is examined using RCNN model with multilayer vectors. To evaluate the efficacy, we subject the system to noise.*

INTRODUCTION

Security has emerged as a top priority for internet users in light of the rapid development of internet technology. Whether it's for personal or professional reasons, emails are a common way to communicate and share information. Phishers may attempt to steal sensitive information using these emails. When a victim clicks on a link in an unsolicited email from a phisher, sensitive information might be collected and utilised in an unethical way. We are using techniques such as RCNN with multilayer vectors and the attention mechanism to identify these phishing emails.

The following methods are used to detect the phishing email:

1) First the email structure is analyzed and mine the text features from email header, email body, work-level and char-level.

2) Using RCNN similar patterns from the email is recognized

3) The email goes through many layers for 0the check using RCNN

4) Using attention mechanism different weights are assigned to different parts of the email like email header and email body to focus more on important information.

ALGORITHM

A. *RCNN Algorithm*

The RCNN is a versatile tool that can classify both images and text. The act of sorting text into predetermined categories is called text classification. Text classifiers automate the process of analysing text and assigning it to pre-defined tags or categories using Natural Language Processing (NLP). Finding and categorising things in a picture is known as object detection. One method for deep learning is R-CNN, which combines the characteristics of convolutional neural networks with rectangle region suggestions. The R-CNN algorithm is a two-step detection method. The next step is to choose which areas of the picture have a good chance of containing an item. In the second step, the item is categorised in each area. When it comes to object identification, R-CNN and its faster counterpart, Faster R-CNN, are the way to go. The R-CNN model may be trained to identify license plate photos or text and backgrounds. The next step, after obtaining the object proposal, is to apply text understanding using a CNN or another ML system.

COMPONENTS

A. *Multilevel Vector*

Manual extraction of features is difficult and time taking ,we can not achieve the effective results .Hence we opt for Multilevel vectors ,which are very useful to extract the features from image or text . As in the case of Phishing email detection, the email has two parts namely email header and email body ,Multilevel vector checks the email header at character-level and word-level . It checks the email body at character-level and word-level .Mostly the phishing content can be found in the email body because the structure of email header is mostly same for all the emails but the email body differs from email to email .The email body is more attractive to get the attention from the victim which differs from legitimate mails .

B. *Attention Mechanism*

Attention is an increasingly popular mechanism used in a wide range of neural architectures. The attention mechanism is a part of a neural architecture that enables to dynamically highlight relevant features of the input data, which, in NLP, is typically a sequence of textual elements. It can be applied directly to the raw input or to its higher-level representation. The core idea behind attention is to compute a weight distribution on the input sequence, assigning higher values to more relevant elements. Attention can be used to compare the input data with a query element based on measures of similarity or significance. It can also autonomously learn what is to be considered relevant, by creating a representation encoding what the important data should be similar to. attention is a technique that mimics cognitive attention. The effect enhances some parts of the input data while diminishing other parts — the motivation being that the network should devote more focus to the small, but important, parts of the data.

C. *Neural Networks*

An artificial neural network, or neural network, is a mathe-matical model inspired by biological neural networks. Inmost cases it is an adaptive system that changes its structure during learning . There are many different types of NNs. For the purpose of phishing detection, which is basically a classification problem, we choose multilayer feedforward NN. In a feedforward NN, the connections between neurons do not form a directed cycle. Contrasted with recurrent NNs, which are often used for pattern recognition, feedforward NNs are better at modeling relation-ships between inputs and outputs. In our experiments, we use the most common structure of multilayer feedforwardNN, which consists of one input layer, one hidden layer and one output layer. The number of computational units in the input and output layers corresponds to the numberof inputs and outputs. Different numbers of units in the hidden layer are attempted .

D. *Deep Learning*

Deep learning is a subset of machine learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behavior of the human brain. ―Deep‖ refers to the many layers the neural network accumulates over time, with performance improving as the network gets deeper. Each level of the network processes its input data in a specific way, which then informs the next layer. So the output from one layer becomes the input for the next . The adjective "deep" in deep learning refers to the use of multiple layers in the network. Early work showed that a linear perceptron cannot be a universal classifier, but that a network with a nonpolynomial activation function with one hidden layer of unbounded width . Deep neural networks consist of multiple layers of interconnected nodes, each building upon the previous layer to refine and optimize the prediction or categorization .

E. *NLP*

Natural Language Processing (NLP) is a subfield of artificial intelligence (AI). It helps machines process and understand the human language so that they can automatically perform repetitive tasks. In natural language processing, human language is separated into fragments so that the grammatical structure of sentences and the meaning of words can be analyzed and understood in context. This helps computers read and understand spoken or written text in the same way as humans. Email filters are one of the most basic and initial applications of NLP online. It started out with spam filters, uncovering certain words or phrases that signal a spam message. But filtering has upgraded, just like early adaptations of NLP. NLP combines computational linguistics—rule-based modeling of human language—with statistical, machine learning, and deep learning models. Together, these technologies enable computers to process human language in the form of text or voice data and to _understand‘ its full meaning, complete with the speaker or writer‘s intent and sentiment . NLP also plays a growing role in enterprise solutions that help streamline business operations, increase employee productivity, and simplify mission-critical business processes.

PROPOSED SYSTEM

In this project, Emails are divided into two categories: legitimate emails and phishing emails. A binary variable y is used to represent an email. If y=1, the email is a phishing email and y=0 means that email is legitimate. This phishing email detection model is used to model emails at the email header, the email body, the character level, and the word level simultaneously. The email is modelled from multiple levels using an improved RCNN model. The attention mechanism is applied between the email header and and the email body ,and different weights are assigned to the two parts so that the model can focus on more different and more useful information.

RESULTS

The phisher who wants to steal the information compose a mail and send to the victim .By using this application you can check whether the website URL is legitimate or phishing .The URL is tested for the phishing content .If the phishing content is present it will be detected .As shown in the picture .

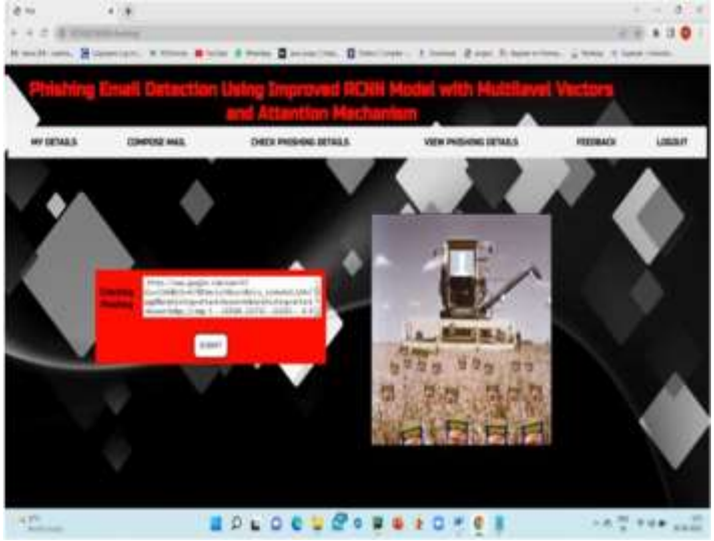

Fig 5.1 Entry of the URL

In the above picture the URL is checked for phishing. The admin logon to the page and give the analysis of the phishing emails and legitimate emails . Graphical analysis of the phishing and legitimate emails are given .As shown in the given picture

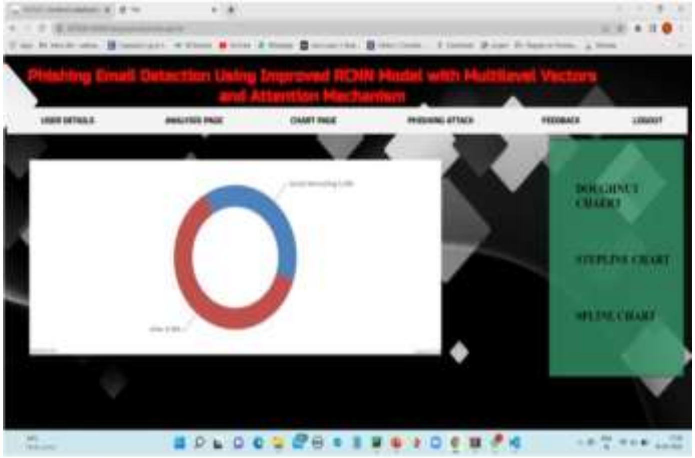

Fig 5.2 Doughnut Graph

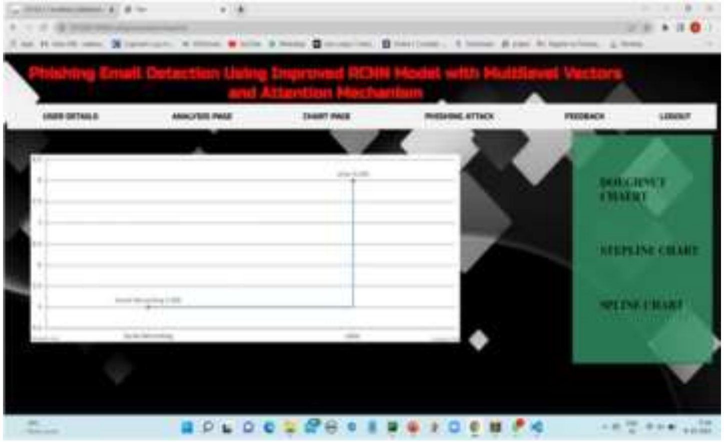

Fig 5.3 Step line Graph

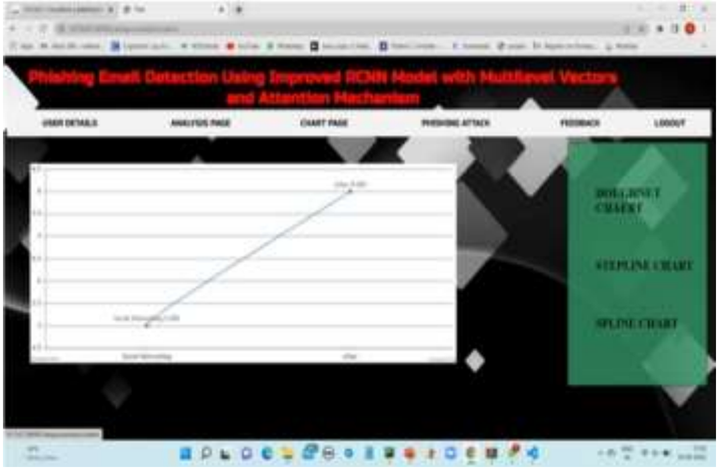

Fig 5.4 Spline Graph

CONCLUSION

Advanced phishing emails may now be detected with the use of a Deep Learning approach. Similar patterns in the email may be identified with the use of RCNN. The email is checked via many levels of check using a multilevel vector. In order to draw greater attention to the most crucial elements of an email, such as the header and the text, an attention mechanism is used.The application's efficacy is evaluated by introducing noise. Using an imbalanced dataset that is more representative of the actual world to test and assess the model.

FUTURE ENHANCEMENTS

1) An alarming trend can be adopted to detect phishing emails

2) More Phishing protections such as email security to prevent the majority of phishing attacks.

REFERENCES

[1] A.-P. W. Group et al., "Apwg attack trends report," USA: Anti-Phishing Working Group (APWG), 2014.

[2] A.-P. W. Group et al., "Phishing activity trends report 1st quarter 2018," USA: Anti-Phishing Working Group (APWG), 2018.

[3] A.-P. W. Group et al., "Phishing activity trends report 4th quarter 2016," USA: Anti-Phishing Working Group (APWG), 2017.

[4] L. M. Form, K. L. Chiew, W. K. Tiong, et al., "Phishing email detection technique by using hybrid features," in IT in Asia (CITA), 2015 9th International Conference on, pp. 1–5, IEEE, 2015.

[5] M. Nguyen, T. Nguyen, and T. H. Nguyen, "A Deep Learning Model with Hierarchical LSTMs and Supervised Attention for Anti-Phishing," arXiv preprint arXiv:1805.01554, 2018.

[6] R. Verma, N. Shashidhar, and N. Hossain, "Detecting phishing emails the natural language way," in European Symposium on Research in Computer Security,
pp. 824–841, Springer, 2012.

T. Venkata Sesha Srikanth Mr. Sk. Mahaboob Basha , R. Neelima , G. Sai Ranjitha , Sohail Eajaz Mohammad. A FRAMEWORK TO DETERMINE CYBERCRIME INFORMATION THROUGH DATA ANALYTIC APPROACH.