Chapter 71
Using an Artificial Neural Network to Improve Email Security

Mohamed Abdulhussain Ali Madan Maki
Ahlia University, Bahrain

Suresh Subramanian
https://orcid.org/0000-0002-4055-8725
Ahlia University, Bahrain

ABSTRACT

Email is one of the most widely used features of internet, and it is the most convenient method of transferring messages electronically. However, email productivity has been decreased due to phishing attacks, spam emails, and viruses. Recently, filtering the email flow is a challenging task for researchers due to techniques that spammers used to avoid spam detection. This research proposes an email spam filtering system that filters the spam emails using artificial back propagation neural network (BPNN) technique. Enron1 dataset was used, and after the preprocessing, TF-IDF algorithm was used to extract features and convert them into frequency. To select best features, mutual information technique has been applied. Performance of classifiers were measured using BoW, n-gram, and chi-squared methods. BPNN model was compared with Naïve Bayes and support vector machine based on accuracy, precision, recall, and f1-score. The results show that the proposed email spam system achieved 98.6% accuracy with cross-validation.

INTRODUCTION

Electronic mail or known as e-mail it is a channel of electronically communicate with others by massages through the internet. Now a day's emails are not only used for communication but also for creating tasks and solving customer queries. Email is, simple, cheap, and fast type of communication, then it could be vulnerable to many threats (David, Lucia, and Bindura, 2013).

One of the most potential security threats in the emailing system is the "SPAM" where attackers are illegally disseminating malicious software's such us Malware's, Viruses', Trojan's and Internet worms (Ndumiyana, Magomelo and Sakala, 2013).

Spam means unwanted email or unsolicited commercial emails sent directly to a large number of addresses (Shama.N, 2017). The spam emails sent to the receivers without their permission. It is possible to send hundreds of emails to thousands of users around the world at no cost.

Spam scientifically different in content and can belong to the following categories: advertisement, money making, sexually explicit, business, scams. (Al-jarrah, Khater and Al-duwairi, 2012)

The spam has been increased in the last years and becomes a serious problem for communication. Mr. Vairagade, (2017) estimated that 48 billion out of the 80 billion emails are sent daily as spam also among 40,000 users are replying to spam emails. According to Symantec (2017), email spam rate decline in 2011 was 75%, in 2015 and 2016 dropped to 53% and the first quarter of 2017 the rate dropped to 54%. The high ratio of spams in recent years indicates that scammers are looking for fast revenue opportunity.

This research tried to improve the techniques of filtering the emails to prevent spam from spreading into customer's mailboxes. Also, we will preprocess the Enron spam dataset and extract the features required to feed NN text classifier. Accordingly this research improved the accuracy and performance of email spam filtering.

BACKGROUND

Theoretical Background

Machine learning is the development of algorithms that permit machines to learn. ML has been used in medical diagnosis, bioinformatics, Money fraud, stock market analysis, classifying DNS, speech recognition, computer games, and spam filtering (Bhuiyan *et al.*, 2018),(R Manikandan, 2018).

Neural Network (NN) is a beautiful biologically inspired programming paradigm, which enables a computer to learn from observational data. Currently, the NN algorithm used widely in many problems, such as text categorizations, image, and speech recognition.

However, extracting the emails and classify them needs knowledge of Natural Language Processing (NLP) to normalize the datasets, extract and select the features to feed the classifiers (Ndumiyana, Magomelo, and Sakala, 2013)(Jayanthi and Subhashini, 2016).

NN has more efficiency in detecting spam because its supervised learning method and also errors can be corrected NB, DT, SVM, KNN are also good classifiers (Sharma, 2014).

The study will use BPNN to improve accuracy and performance in detecting email spam.

Related Work

The tasks of managing a large volume of data are challenged because of the growing number of emails around the world. For example, detecting spam with many security roles may slow down system performance. Many studies include replies in the email, folder classification, automatic subject, contacts. Currently, email servers such as Google combined the email communication once the user replied (Alsmadi and Alhami, 2015).

Related Content

Cooperative Control of Unknown Networked Lagrange Systems using Higher Order Neural Networks
Gang Chenand Frank L. Lewis (2013). *Artificial Higher Order Neural Networks for Modeling and Simulation (pp. 214-236).*
www.irma-international.org/chapter/cooperative-control-unknown-networked-lagrange/71801

Design for Testability of High-Speed Advance Multipliers: Design for Testability
Suman Lata Tripathi (2020). *AI Techniques for Reliability Prediction for Electronic Components (pp. 175-190).*
www.irma-international.org/chapter/design-for-testability-of-high-speed-advance-multipliers/240497

Artificial Higher Order Neural Network Training on Limited Precision Processors
Janti Shawashand David R. Selviah (2010). *Artificial Higher Order Neural Networks for Computer Science and Engineering: Trends for Emerging Applications (pp. 312-332).*
www.irma-international.org/chapter/artificial-higher-order-neural-network/41673

A Decision Making Method Based on Society of Mind Theory in Multi-Player Imperfect Information Games
Mitsuo Wakatsuki, Mari Fujimuraand Tetsuro Nishino (2020). *Deep Learning and Neural Networks: Concepts, Methodologies, Tools, and Applications (pp. 317-329).*
www.irma-international.org/chapter/a-decision-making-method-based-on-society-of-mind-theory-in-multi-player-imperfect-information-games/237879

Application of Deep Learning in Speech Recognition
Rekh Ram Janghel, Satya Prakash Sahu, Yogesh Kumar Rathore, Shraddha Singhand Urja Pawar (2019). *Handbook of Research on Deep Learning Innovations and Trends (pp. 59-71).*
www.irma-international.org/chapter/application-of-deep-learning-in-speech-recognition/227844