www.ijsrnsc.org

Available online at www.ijsrnsc.org

IJSRNSC

Volume-10, Issue-4, October 2022
Review Paper

Int. J. Sc. Res. in Network Security and Communication

E-ISSN:2321-3256

Application of Adaptive Neuro-Fuzzy Inference Systems in Mobile's Spam

Jyoti Chouhan[1*], Raju Barskar[2], Uday Chourasia[3]

[1,2,3]Department of Computer Science and Engineering, UIT-RGPV Bhopal, India

Received: 19/Aug/2022, Accepted: 22/Sept/2022, Published: 31/Oct/2022

***Abstract*:** The function of science is to explain reality logically. With the increased usage of the internet, spamming is the most typical problem that we faced every day. Service of mailing and web applications are much spammed, nowadays this expansion in the number of spam messages has turned into a significant issue. Accessibility of messaging services for a minimal price has come about in the expansion of spam messages. Spam location is attempting a direct result of the necessity for semantic analysis of the compact spam messages, which overall will for the most part have overlapping polarities. In this work, a portable spam order strategy is created in light of an Adaptive Neuro Inference System(ANFIS) containing Gini record fuzzy and Back-Propagation in machine learning. The methodology includes Gini's index models for information indexes and a back-propagation-based neural network as the AI classifier. This paper presents the assessment of the introduced system based on the error, accuracy of classification, and the number of iterations.

Keywords: Mobile Spam Classification, ANFIS, Gini's Index, Back Propagation, Training Iterations, Classification Accuracy.

I. INTRODUCTION

Now a days, mobile phones have become an essential part of communication for a large number of people. More than half of population of the world is using mobile phones due to a wide range of functionalities these devices provide. Ever since this technology has evolved, one of its signifcant features that is very popular is text messaging. This internet free service has become an important means of communication for people but one major disadvantage of this technology is the associated degree forbidding rate of spam messages that are sent over the snetwork to a large number of people[1]. The number of mobile users has dramatically increased in the recent years with an estimate of over 7 billion subscriptions globally. The common form of textual communication between mobile devices is the use of SMS, which utilizes standardized communication protocols to enable mobile phones exchange short text messages with 160 character long. On the other hand, micro blogging online social networks, such as Twitter have been utilized for a range of social activities including the posting of interesting contents about past experiences, locating longlost friends, posting photos and videos, building communities joined by families, and friends. Micro blogging social networks have been in existence for almost a decade. For instance, the launch of Twitter in 2006 witnessed a rise in the number of micro blogging platforms[2]. Spam is also expensive problem that can costs lots of expenses in a year to the service providers for the loss of bandwidth. Therefore, it is very essential to distinguish spam emails, many methods have been suggested to identify and classify email messages as spam or non spam or legitimate mail and it has been found that the algorithm success rate of machine learning is extremely high. Several algorithms that are processed for the classification of unwanted emails that are widely used and analyzed amongst them are vector machine, Naive Bayes, decision tree, neural network classifiers are well-known classifiers. In this article, we experiment with algorithms: Naive Bayes, Vector Machine Support (SVM)[3]. World Wide Web is being used ubiquitously for information retrieval, but at the same time, it is being exploited by the mischievous segment of society that induces spam on the Web. According to spam is one of the major challenges of Web search engines and an important factor that affects the quality of their search results. Redirection spam refers to the technique whereby a user is forced to pass through a series of intelligently crafted redirections that finally terminate on the compromised Web page. A redirection attack is triggered when a search user unknowingly clicks a hyperlink pointing to a malicious Websites. According to redirection spam takes the user to a page that is semantically different from the requested page. Malicious redirections create an unnecessary network traffic, thereby leading to underutilization of network bandwidth. It also puts a bad impact on the reputation of search engines and affects the trust of search users[4].

Fundamental aspects of fuzzy logic

Fuzzy analysis is based on set theorem of pure mathematics. For SPAM research, usually we use basic statistical tools, scales, indices both for cross-sectional and longitudinal study. The basic difference between crisp set and fuzzy set might generate a new thinking for using fuzzy tolls for Spam analysis. For instance, a conventional (or "crisp") set is dichotomous: a case is either "in" or "out" of a set, for example, the set of professors. Thus, a

conventional set is comparable to a binary variable with two values, 1 ("in," i.e., professors) and 0 ("out," i.e., not-professors). A fuzzy set, by contrast, permits membership in the interval between 0 and 1 while retaining the two qualitative states of full membership and full non-membership. Thus, the fuzzy set of professors could include individuals who are "fully in" the set (fuzzy membership = 1.0, full professors), some who are "almost fully in" the set (membership =.90 has applied for professor!), some who are neither "more in" nor "more out" of the set (membership = .5, also known as the "crossover point", still in the middle in the stair of associate professor and professor), some who are "barely more out than in" the set (membership = .45, still enjoying associate professorship), and so on down to those who are "fully out" of the set (membership = 0, out of the set of professors, merely assistant professors). It is up to the researcher to specify procedures for assigning fuzzy membership scores to cases, and these procedures must be both open and explicit so that they can be evaluated by other scholars. Basic phenomenological alternatives understandings allow us to set down the notion of 'stock knowledge at hand' by which it is easy for a researcher to attach membership function in a social phenomenon.

Number of Mobile Spams for Quarter-3, 2020
Courtesy: Kaspersky Labs Security Report

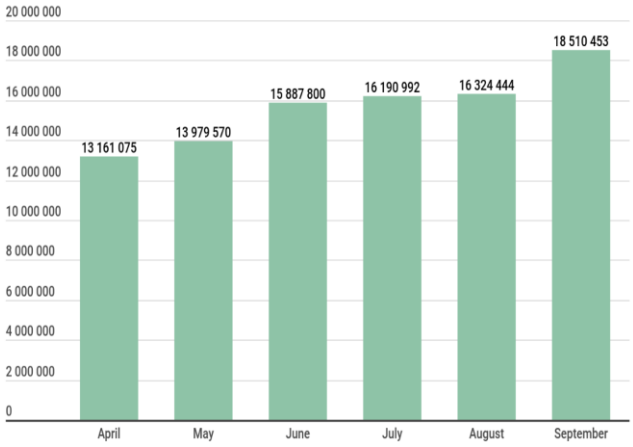

Figure1. Number of Mobile Spam for Quarter-3, 2020

II. AN OVERVIEW OF MOBILE SPAM CLASSIFICATION

There are various approaches have been devised for mobile spam classification. Some of the selected studies are as under:

A.K. Jain, D. Goel, S. Agarwal, Y. Singh, and G. Bajaj(2020) This paper presents, The model proposed in this paper proves to be an efcient method for detecting spam messages. The existing current applications available to detect spam messages works on the basis of user reviews. However, the algorithms given by us analyze the message on the basis of various features identifed by us that are able to efectively detect spam messages[5].

K. S. Adewole, N. B. Anuar, A. Kamsin, and A. K. Sangaiah(2017) This paper proposes a unified framework that can be used to identify spam messages and spam accounts successfully. The performance of the proposed unified framework is investigated using four datasets, two of which are from SMS spam detection domain and are publicly available for research purpose. The remaining two datasets were collected from Twitter microblog to investigate the capability of the proposed framework for spam message and spam account detection on Twitter[2].

N. Mirza, B. Patil, T. Mirza, and R. Auti(2017) In this paper, Spam mails are generally used to send bulk mails to a sender. Spam floods the Internet with many similar copies of messages that is distributed in depth; these messages are sent strongly to recipients who would not otherwise choose to receive it. We will analyze several methods of mining data for spam data in order to figure out the best classifier for sorting emails. As part of this paper, we explain the classification of emails to identify spam and not spam. For this purpose, we use the Naive Bayesian Classifier and created an email classification system to classify spam and not spam[3].

K. Hans, L. Ahuja, and S. K. Muttoo(2017) In this paper, we propose a neural framework for detecting redirection spam. We incorporated the feed-forward multilayer perceptron network and used scaled conjugate gradient algorithm that is able to perform very fast classification of URLs leading to redirection spam. We investigated the network empirically to choose the number of hidden layers and observed that when network is trained with two hidden layers, it gives better accuracy. We validated our proposed approach against the dataset of 2383 URLs and were able to detect the spammed redirections with high accuracy. The results indicate that neural networks are very effective technique to model the redirection spam detection[4].

S. Sedhai and A. Sun(2017) In this paper, we propose a semi-supervised spam detection framework, named S3D. S3D utilizes four lightweight detectors to detect spam tweets on real-time basis and update the models periodically in batch mode. The experiment results demonstrate the effectiveness of semi-supervised approach in our spam detection framework. In our experiment, we found that confidently labeled clusters and tweets make the system effective in capturing new spamming patterns[6].

H. Afzal and K. Mehmood(2016) This paper analyses different classification techniques that are currently being used in spam filtering in the context of social media. The contents of tweets are unique in nature, and are different from emails due to their less content so some techniques used in emails might be effective while some might not be effective[7].

C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou, and G. Min(2016) In this paper, we firstly identify the "Spam

Drift" problem in statistical features based Twitter spam detection. In order to solve this problem, we propose a L fun approach. In our L fun scheme, classifiers will be re-trained by the added "changed spam" tweets which are learnt from unlabelled samples. We evaluate the performance of L fun approach in terms of Detection Rate and F-measure. Experimental results show that both detection rate and F-measure are improved a lot when applying with our L fun approach[8].

T. Vyas, P. Prajapati, and S. Gadhwal(2015) This paper can be classified into ham and spam and with their increasing use, the ratio of spam is increasing day by day. There are several machine learning techniques, which provide spam mail filtering methods, such as Clustering, J48, Naive Bayes etc, Also a comparative study of each technique in terms of accuracy and time taken is provided[9].

N. Jatana(2014) The main objective of this paper is to reduce overall time in the process of spam detection. Quantitative and qualitative analysis of the proposed technique, performed on two public spam databases (Spam Assasin and Ling Spam) has shown improved time performance. The proposed method has performed up to six times faster than the existing Paul Graham's Bayesian approach[10].

K. Kandasamy and P. Koroth(2014) In this paper we are proposing an application which can classify a Twitter user into spam or legitimate. To achieve this, an integrated approach, which contains URL analysis, Natural Language Processing and Machine Learning techniques are used[11].

N. Prasad, R. Singh, and S. P. Lal(2013) In this paper we compare the performance of back propagation and resilient propagation algorithms in training neural networks for spam classification. Back propagation algorithm is known to have issues such as slow convergence, and stagnation of neural network weights around local optima. Researchers have proposed resilient propagation as an alternative. Resilient propagation and back propagation are very much similar except for the weight update routine[12].

A. Rajadesingan and A. Mahendran(2012) In this paper, we consider the problem of spamming in blogs. In blogs, spammers usually target commenting systems which are provided by the authors to facilitate interaction with the readers. Unfortunately, spammers abuse these commenting systems by posting irrelevant and unsolicited content in the form of spam comments. Thus, we propose a novel methodology to classify comments into spam and non-spam using previously-undescribed features including certain blog post-comment relationships[13].

A. K. Uysal, S. Gunal, S. Ergin, and E. S. Gunal(2012) In this paper, a novel framework for SMS spam filtering is proposed to be able to block unsolicited SMS messages. In the filtering framework, distinctive features representing SMS messages are identified using CHI2 and IG- based feature selection methods. The selected feature subsets with varying sizes are then fed into two different Bayesian-based classification algorithms, namely the binary and probabilistic models, to classify SMS messages as either legitimate or spam. Additionally, the proposed SMS spam filtering scheme is employed to develop a real-time mobile application running on the mobile phones with Android operating system[14].

S. Vahora, M. Hasan, and R. Lakhani(2011) In this paper, we are using novel approach which uses Vector space model with Naïve Bayes to correctly classify mails as spam mail. Naïve Bayes method is used for spam classification but still binding with personalize word vector helps in increasing the accuracy of the system because user receives special type of message only[15].

L. Araujo and J. Martinez-Romo(2010) In this paper, we proposed a new methodology to detect spam on the Web, based on an analysis of QLs and a study of the divergence between linked pages. To use QLs and the LM features effectively, we proposed a robust classifier based on a cost-sensitive algorithm. We have evaluated our methodology using the public WEBSPAM-UK2006 and WEBSPAM-UK2007 datasets and we focus on the F-measure, using the proposed features in a separate and also in combined way[16].

S. M. Lee, D. S. Kim, J. H. Kim, and J. S. Park(2010) In this paper, we have presented an optimal spam detection model based on RF. We performed parameters optimization and feature selection simultaneously using RF[17].

C. Y. Tseng and M. S. Chen(2009) In this paper, we propose an incremental support vector machine (SVM) model for spam detection on dynamic email social networks. A complete spam detection system MailNET is devised to better adjust to diverse networks. Several features of each user in the network are extracted to train an SVM model. we present an incremental update scheme to efficiently re-train an SVM model[18].

III. COMPARISION TABLE

Table 1

Title of paper	Authors	Approached used	Demerits	Publication & year
"Predicting Spam Messages Using Back Propagation Neural Network",	AK Jain, D Goel, S Agarwal,Y Singh, G.Bajaj,	Back Propagation Used for Training Neural Network for spam classification.	No pre-processing and probabilistic approach.	Springer 2020-21, vol. 110, pp. 403-422.

SMSAD: a frameworkfor spam message and spam account detection	KS Adewole,NB Anuar,A Kamsin,	Support Vector Machine Used as a classifier. Main features are frequency of spam words and location information/phone number information.	No separate pre-processing. SVM is prone to saturation in performance with increasing size of dataset.	Springer 2019, vol. 78, pp. 78, 3925–3960.
Detecting redirection spam using multilayer perceptron neural network	Kanchan Hans, Laxmi Ahuja, S. K. Mutto,	Scaled Conjugate Gradient (SCG) algorithm used with the aim to reduce computational complexity for large datasets.	Doesn't use lexical analysis.	Springer 2017. volume-21, pp.3803–3814

IV. CONCLUSION

In this paper, it can be concluded design an Adaptive Neuro Fuzzy Inference System(ANFIS) based on the Gini index for mobile spam classification. In this generation increase in the usage of web applications is very challenging to detect spam detection. The most trustworthy text-based communication channel is SMS, with the help of this technique we have to send any message to another destination. But highly preferred by the attackers to spread spam messages using text SMS and emails. In the proposed technique performance the existing techniques in terms of classification accuracy.

REFERENCES

[1] A. K. Jain, D. Goel, S. Agarwal, Y. Singh, and G. Bajaj, "Predicting Spam Messages Using Back Propagation Neural Network," *Wirel. Pers. Commun.*, Vol.**110**, Issue.**1**, pp.**403–422, 2020**, doi: 10.1007/s11277-019-06734-y.

[2] K. S. Adewole, N. B. Anuar, A. Kamsin, and A. K. Sangaiah, "SMSAD: a framework for spam message and spam account detection," *Multimed. Tools Appl.*, Vol.**78**, Issue.**4**, pp.**3925–3960, 2019**, doi: 10.1007/s11042-017-5018-x.

[3] N. Mirza, B. Patil, T. Mirza, and R. Auti, "Evaluating efficiency of classifier for email spam detector using hybrid feature selection approaches," *Proc. 2017 Int. Conf. Intell. Comput. Control Syst. ICICCS 2017*, Vol.**2018**, pp.**735–740, 2017**, doi: 10.1109/ICCONS.2017.8250561.

[4] K. Hans, L. Ahuja, and S. K. Muttoo, "Detecting redirection spam using multilayer perceptron neural network," *Soft Comput.*, Vol.**21**, Issue.**13**, pp.**3803–3814, 2017**, doi: 10.1007/s00500-017-2531-9.

[5] A. K. Jain, D. Goel, S. Agarwal, Y. Singh, and G. Bajaj, "Predicting Spam Messages Using Back Propagation Neural Network," *Wirel. Pers. Commun.*, Vol.**110**, Issue.**1**, pp.**403–422, 2020**, doi: 10.1007/s11277-019-06734-y.

[6] S. Sedhai and A. Sun, "Semi-Supervised Spam Detection in Twitter Stream," *IEEE Trans. Comput. Soc. Syst.*, Vol.**5**, Issue.**1**, pp.**169–175, 2018**, doi: 10.1109/TCSS.2017.2773581.

[7] H. Afzal and K. Mehmood, "Spam filtering of bi-lingual tweets using machine learning," *Int. Conf. Adv. Commun. Technol. ICACT*, Vol.**2016**, pp.**710–714, 2016**, doi: 10.1109/ICACT.2016.7423530.

[8] C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou, and G. Min, "Statistical Features-Based Real-Time Detection of Drifted Twitter Spam," *IEEE Trans. Inf. Forensics Secur.*, Vol.**12**, Issue.**4**, pp.**914–925**, **2017**, doi: 10.1109/TIFS.2016.2621888.

[9] T. Vyas, P. Prajapati, and S. Gadhwal, "A survey and evaluation of supervised machine learning techniques for spam e-mail filtering," *Proc. 2015 IEEE Int. Conf. Electr. Comput. Commun. Technol. ICECCT*, **2015**, doi: 10.1109/ICECCT.2015.7226077.

[10] N. Jatana, "Radix Encoded Fragmented Database Approach," pp.**939–942, 2014**.

[11] K. Kandasamy and P. Koroth, "An integrated approach to spam classification on Twitter using URL analysis, natural language processing and machine learning techniques," *2014 IEEE Students' Conf. Electr. Electron. Comput. Sci. SCEECS 2014*, **2014**, doi: 10.1109/SCEECS.2014.6804508.

[12] N. Prasad, R. Singh, and S. P. Lal, "Comparison of back propagation and resilient propagation algorithm for spam classification," *Proc. Int. Conf. Comput. Intell. Model. Simul.*, pp.**29–34, 2013**, doi: 10.1109/CIMSim.2013.14.

[13] A. Rajadesingan and A. Mahendran, "Comment spam classification in blogs through comment analysis and comment-blog post relationships," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, Vol.**7182**. no. PART 2, pp.**490–501, 2012**, doi: 10.1007/978-3-642-28601-8_41.

[14] A. K. Uysal, S. Gunal, S. Ergin, and E. S. Gunal, "A novel framework for SMS spam filtering," *INISTA 2012 - Int. Symp. Innov. Intell. Syst. Appl.*, **2012**, doi: 10.1109/INISTA.2012.6246947.

[15] S. Vahora, M. Hasan, and R. Lakhani, "Novel approach: Naïve Bayes with vector space model for spam classification," *2011 Nirma Univ. Int. Conf. Eng. Curr. Trends Technol. NUiCONE 2011 - Conf. Proc.*, pp.**11–15, 2011**, doi: 10.1109/NUiConE.2011.6153245.

[16] L. Araujo and J. Martinez-Romo, "Web spam detection: New classification features based on qualified link analysis and language models," *IEEE Trans. Inf. Forensics Secur.*, Vol.**5**, Issue.**3**, pp.**581–590, 2010**, doi: 10.1109/TIFS.2010.2050767.

[17] S. M. Lee, D. S. Kim, J. H. Kim, and J. S. Park, "Spam detection using feature selection and parameters optimization," *CISIS 2010 - 4th Int. Conf. Complex, Intell. Softw. Intensive Syst.*, no. i, pp.**883–888, 2010**, doi: 10.1109/CISIS.2010.116.

[18] C. Y. Tseng and M. S. Chen, "Incremental SVM model for spam detection on dynamic email social networks," *Proc. - 12th IEEE Int. Conf. Comput. Sci. Eng. CSE 2009*, Vol.**4**, pp.**128–135, 2009**, doi: 10.1109/CSE.2009.260.

