IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 10, October 2015.
www.ijiset.com
ISSN 2348 – 7968

Efficient Classifier for Detecting Spam in Social Networks

E.Nalarubiga
M.E-Software Engineering
Rajalakshmi Engineering College
Chennai, India
nalarubiga.e.2014.mese@rajalakshmi.edu.in

M.Sindhuja, M.E.,
Assistant Professor
Rajalakshmi Engineering College
Chennai, India
Sindhuja.m@rajalakshmi.edu.in

***Abstract*— Social networking services are used for communication between people to share information through internet. The unbounded growth of content and users pushes the internet technologies to certain limitations. Data mining plays a major role in the field of social network to extract relevant content from the voluminous data which is being a phenomenal task, because of its dynamic nature the participation is more complex. The major problem, users face spammer's interaction which leads to misunderstanding and inconvenience for social activities. This work concentrates on detecting the spammer actions using feature relevance analysis and applying efficient classifier. The main objective of the proposed work is to find relationship between features and classifying patterns for detecting spam message from the unwanted sites. The system applies efficient classification algorithms after feature relevance analysis for detecting spam in better way. The outcome of this project will serve for the users participating in social networks for effective purpose like business, marketing and establishing contacts**.

KEY WORDS: Classification, Data Mining, Machine Learning, Predictive analysis, Social Networking Spam, Spam detection.

I. INTRODUCTION

In recent years, internet has become an integral part of our life. With increased use of internet, numbers of email and social media (Facebook, Twitter, Linkln, and Google+) users are increasing day by day. In which user can connect to other users and post messages to each other and on other pages within in the network. Some social networking sites rely on their users not only to generate content, but also fight spam and other inappropriate content. The content of the spam messages include pharmaceuticals, jewelry, electronics, loans, stocks, weight loss, and gambling: as well as malware and phishing (identify theft) lures.

Within social network, spam can be found in publicly visible pages (celebrity, groups, profiles), on profile pages and in private inboxes. Most networking sites allow their users to issue a report on all of these levels when they feel a message is spam or otherwise inappropriate. These user spam report are the main ingredient of this paper. We explore the usefulness user spam reports in classifying spam in social networks. More precisely, we present a framework that indicates the likelihood of a message being spam, based on users spam report.

A spam filter is a program that is used to detect unsolicited and unwanted messages and prevent those messages from getting to a user's inbox. Like other types of filtering programs, a spam filter looks for certain criteria on which it bases judgments. For example, the simplest and earliest versions (such as the one available with Microsoft's Hotmail) can be set to watch for particular words in the subject line of messages and to exclude these from the user's inbox. This method is not especially effective; it may omit legitimate messages (called false positives) and passing actual spam messages. More sophisticated programs such as Bayesian filters or other heuristic filters, attempt to identify spam through suspicious word patterns or word frequency. Filter classification strategies can separated into two categories: those based on machine learning (ML) principles and those not based on ML. ML approaches are capable of extracting knowledge from a set of messages supplied, and using the obtained information in the classification of newly received messages.

Non-machine learning techniques, such as heuristics, blacklisting and signatures, have been complemented in recent years with new, ML-based technologies. In the last few years, substantial academic research has taken place to evaluate new ML-based approaches to filtering spam. ML filtering techniques can be further categorized into complete and complementary solutions. Complementary solutions are designed to work as a component of a larger filtering system, offering support to the primary filter (whether it be ML or non-ML based). Complete solutions aim to construct a comprehensive knowledge base that allows them to classify all incoming messages independently.

ISSN 2348 – 7968

Social network analysis focuses on measuring various aspects of entities and the relationships between them. This includes identifying "central" nodes within a network and determining the distance between nodes. The entities of interest are often people and the relationships of interest are often social interactions, but the concepts are easily transferred to computers and email connection between them.

II. Classification and Prediction

Classification is a machine learning technique which assigns class labels to different groups. Classification is the most frequently used data mining function with a predominance of the implementation of Bayesian classifiers, K-NN, and Support Vector Machines. The analysis of social media data improves the spam detection by improving the performance of social media user management tasks.

The main application of data mining in the field of social media is predictive analysis. Spam can be predicted if the users past history is analyzed. Predictive analysis is a data mining technique applied on electronic spam records in order to effectively predict the spam at a very early stage. After identification, the social users can avoid spam messages and spammer interaction in the usage of social network.

III. Significance of the System

The paper mainly focuses on how machine learning techniques in Data mining can be applied to predict the risk factors of spam in the data that is being used.

The study of literature survey is presented in section IV, Methodology is explained in section V, section VI covers the experimental results of the study, section VII discuses the future study and Conclusion.

IV. LITERATURE SURVEY

Xianghan Zheng et al. [1] describe the user's social activities and spammer malicious activities in social network. Supervised machine learning solution is used to detect spammers efficiently with the help of collected dataset from Sina Weibo including 30,116 users and more than 16 million messages. Support Vector Machines based spammer detection algorithm is used to extract a set of feature from message content and users' social behavior. It has lower training time with high accuracy.

Atefeh Heydari et al. [2] made a survey on online reviews are most important resource of customers opinions. To prevent and avoid fraudster's activities use opinion mining techniques. Unsupervised and supervised learning approaches are used. Spam detection in data can be reviewed by content, metadata and information about products. Spam review is very difficult and accuracy cannot be measured exactly.

Fahim A et al. [3] describes spam is an unwanted electronic message post to the internet users in any form. Facebook become the leading one and different users in Facebook by posting or creating ways to post spam. Facebook users' faces potential spam types. User become victim to spam attack and to handle different types of spam knowledge based and artificial neural networks spam detection technique is used.

Mohammed N et al. [4] developed a top Arabic websites which are selected for evaluating possible web spam behavior. Spam techniques are used for boost their ranks within search engine result page. Naive based classifier is used to classify web pages and Term Frequency Inverse Document frequency, HITS algorithm and page ranking algorithms are used to increase their website ranks.

Xin Liu et al. [5] proposed a spam filtering approach with push technology to share user's individual spam knowledge in social network. Spam filtering approaches like source based method and content based method are used. Improve performance and accuracy rate using Bayesian filter.

Vipin N S et al. [6] describes a distributed filtering scheme perform spam filtering on secure messages without decrypting them. Filters for such messages should operate in real time on large volume of data. Merkle-Hellman encryption scheme is used and provides real time filter without loss of privacy. Solve load overhead caused by message explosion and reduces running time by filtering encrypted text.

Zhipeng Zeng et al. [7] survey on supervised machine learning based spammer filtering with Sina Weibo dataset. Support vector machine classifier is used and shows the true positive rate of spammers and non spammers. Content based and user based features are used for cumulative distribution function. A dataset collected from Sina Weibo that includes 30,116 users and more than 16 million messages.

Dae-Ha Park et al. [8] proposed a multiple Bayesian Network classifiers to provide social spam detection framework with effective features based on user's behavior and previous patterns of spammers. Social networking sites provide the way for communication between peoples and various features such as behavior, celebrity, trust, and common interest should be updated to the users to make the decision efficiently and away from the spam.

David Mandell Freeman [9] describes a set of features that can be used by a Naive Bayes classifier to find accounts whose names do not represent real people. Name scoring algorithm and false positive rate algorithm are used to

ISSN 2348 – 7968

provide effective way to catch bad actors and also poor email generation algorithm. Data from LinkedIn to train and validate our model and best-scoring model achieves AUC 0.85 on a sequestered test set. LinkedIn data for one month in parallel with our previous name scoring algorithm based on regular expressions. The false positive rate of our new algorithm (3.3%) was less than half that of the previous algorithm (7.0%).

Sajid Yousuf Bhat et al. [10] present a Community based framework to identify spammers in OSNs for identifying spamming accounts. Community based node features of OSN users improve performance of classifying spammers and legitimate users. Node level community detection and feature extraction are based on machine learning technique is used to known spam.

Kuan Zhang et al. [11] proposed an effective social based updatable filtering protocol (SAFE) with privacy preservation in MSNs. Mobile social network used to provide social interaction and information sharing among users. Spam filtering protocol helps to reduce communication and storage overhead. MSN with effective SAFE privacy preservation by using Merkle tree property. SAFE can protect user's private information from filters and avoid forgery attack.

Geerthik.S [12] surveys the classification of various spam in the internet based on their properties. The impact of various spam's in social networks, email, image, content and links is discussed and the technique applied to prevent the spam in various areas is listed. A detailed analyzes of cloaking and redirection in web search is also given is this paper. It analyzed all the types of spam in the internet the things to consider for designing effective spam filter is also surveyed.

Rupam Some [13] survey Social network analysis based on relationship of people, organization. Link based techniques for analyzing social networks improve text based retrieval and ranking strategies. Recent trends on research are in area of link analysis, dark web analysis and spam behavior detection.

Helen Costa et al. [14] proposed a tip spam in location based social networks which Identifying tip spam on a popular Brazilian LBSN system crawled information about users and locations. Dataset are collected and labeled as spam and non spam tips. Random forest algorithm is implemented in weka tool. Using a classification technique, correctly identify a significant fraction of spam and also non spam tips accuracy.

Yang Yu et al. [15] proposed a development of mobile short message services. Online spam filters analysis based on content representation and relationship between sender and receivers. Naïve Bayesian classifier used to the filter including the content features and social network features. Runtime optimization makes the algorithm effective.

Dave DeBarr et al. [16] proposed two methods random project and Logit boost which is a combination of random boost. Random boost method improves spam filter compared to logit boost algorithm. Random Boost algorithm reduces training time. Logit boost algorithm uses a greedy approach to learning, focusing on best features for distinguish spam from non spam.

Xin Jin et al. [17] proposed a social Spam Guard system depend on users for content contribution and sharing. Feature extractions are extracted based on image content features, text content features and social network features GAD clustering algorithm used for large scale clustering and integrate to avoid duplicates.

Gianluca Stringhini et al. [18] describe social networking users storing and sharing a wealth of personal information in social network platform such as Facebook, MySpace, or Twitter. Analyze spam in social networking sites based on spammy activities. Machine learning techniques are used to detect malicious activity and identify the accounts used by spammers.

V. METHODOLOGY

Data mining algorithms used

Classification is one of the important tasks in Data mining. There are many types of classification algorithms for classifying the data. These classification algorithms also play a significant role in analyzing and predicting the social media data. Some of the commonly used classification algorithms for predicting spam are SVM, Naïve Bayes, ID3, KNN, Random Tree, and Random Forest. These algorithms are used in accordance with the problem specificity. On the other hand, the algorithms have their own advantages and disadvantages.

Discussion

The method is based on the idea of using several data sources as input to an engine that classifies a message as either spam or ham. These data sources could comprise pieces of information from several social media. Given data from these data sources, the engine creates a graph of users and extracts basis for the classification of incoming messages, regardless of which medium is used to transfer the message. Since data collected may not be correct always, data is preprocessed to avoid any inconsistencies in the data. Filtering is done for the feature selection process where the most relevant attributes are given highest priority while classifying the data.

ISSN 2348 – 7968

Dataset Description

The Dataset used in this work is the Weibo Dataset from the popular social network in China. This dataset is used throughout the study for classifying the spammers and non-spammer. A large dataset containing 30,116 users and more than 16 million messages were crawled. In order to label users as spammers or non-spammers, Majority voting was introduced to decide which class user should be if one user was labeled to different classes. In total, 8858 spammers and 17646 non-spammers were labeled. Since user labeling process is greatly depend on human judgment, which would directly lead to inevitable human error. Thus, we only randomly select about 80% spammers and non-spammers from labeled dataset as our training data collection, and the rest

A. Data Preprocessing

Data collected is not always complete and consistent. In order to remove all the inconsistencies those are associated with the data, we go for data preprocessing. There are many data preprocessing techniques in existence.

B. System Design

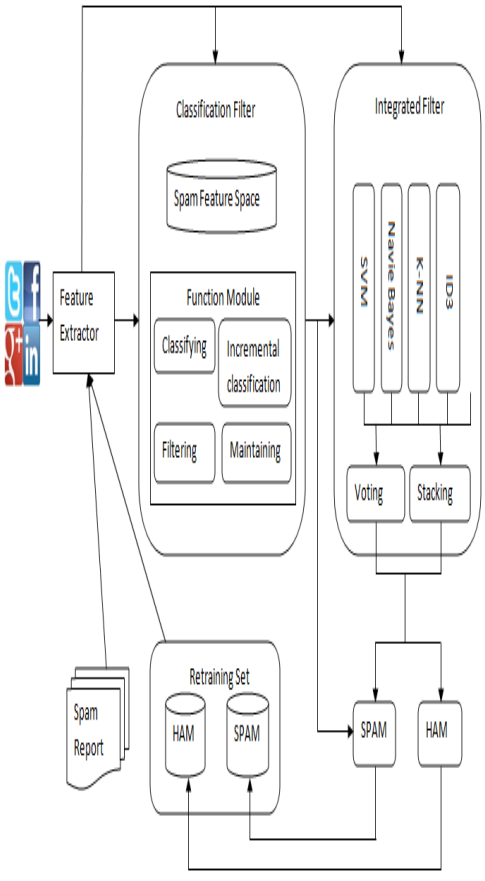


Fig1. System Design

This section explains the steps involved in building the classifier model. The feature extractor is used to extract low-level message features. During the filtering process, the classification filter makes the filtering decisions before the integrated filter. All messages that are classified as spam will be put into the spam dataset directly. Since the classification filter is more resistant to obscuring tricks, let it make the filtering decision in advance can improve the filtering precision. The messages that have passed through the classification filter will be further inspected by the integrated filter. Thus, spam from unknown spam sources can also be detected. The combination of the two filters will improve the filtering capacity. Furthermore, this hierarchical framework can outperform the concurrent framework in terms of filtering speed. Finally, social media will be classified as spam once the proportion of contained spam messages is above certain ratio. Then the data is validated and finally report is being generated.

VI. EXPERIMENTAL RESULTS

A. Accuracy Measure of individual classifiers

In order to perform data analysis and prediction, a lot of data mining classification algorithms are applied in the social media dataset and is implemented using WEKA. In this study, we have compared the performance of most of the classification algorithms such as Naïve Bayes, ID3, Random Forest, K-NN, Support Vector Machine, etc and fig2. Shows their accuracy measures with different ratios of spammer to non spammer in training dataset.

Table1. Accuracy Measure of Classifiers

S.no	Algorithm	Accuracy percentage		
		Without Filtering	Fisher filtering	Correspondence analysis
1	C 4.5	90.62	87.7	88.15
2	MC4	79.42	79.42	79.42
3	ID3	77.21	77.21	77.21
4	K-NN	80.33	81.90	81.90
5	LDA	78.38	77.21	77.21
6	MLP	78.77	78.77	78.9
7	NBC	75.39	75.78	75.78
8	PLS-LDA	77.21	75.52	75.52
9	RND	100	100	100
10	SVM	77.47	77.08	77.08

Correspondence analysis involves finding coordinate values which represents the row and the column

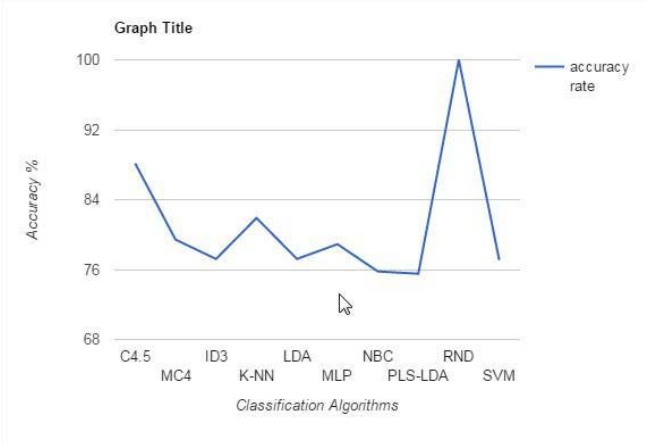
categories in an optimal way. The global Independence between two variables is generally measured by Chi-squared ($\chi 2$) and is calculated as,

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(N_{ij} - E_{ij})^2}{E_{ij}}$$

Where *Eij* is expected to count under independence

$$E_{ij} = \frac{N_{i.}N_{.j}}{N_{..}}$$

Fig2. is a graph showing the accuracy rate of various classification algorithms.



B. Classification result and comparison

Compare different approach with each other classifiers: Decision tree, SVM, Naïve Bayes and Bayes network with implementation provided by Weka. For each classifier, the same evaluation metrics (precision, recall and F-measure) are calculated for both spammers and non-spammers, with the result illustrated in Table1.

Table2. Comparison between classifiers

Classifier	Precision		Recall	
	Spammer	Non-Spammer	Spammer	Non-Spammer
SVM	0.999	0.995	0.991	0.999
Decision Tree	0.942	0.95	0.953	0.958
Naïve Bayes	0.939	0.96	0.922	0.966
Bayes Network	0.946	0.915	0.907	0.956

VII. CONCLUSION AND FUTURE WORK

Social networking sites have millions of users from all over the world. The ease of reaching these users, as well as the possibility to take advantage of the information stored in their profiles, attracts spammers and other malicious users.

In this paper, we showed that spam on social networks is a problem. The proposed methodology aims at providing an efficient classification framework for predicting and monitoring the spammer. The main aim of this research is to identify and construct models that would help social networking users in an efficient way purpose like business, marketing and establishing contacts.

The future work will explore a working model of integrating a smart system to this classification framework for continuous monitoring of the spam and also to include an automatic mechanism for detection spam continuously.

REFERENCES

[1]Xianghan Zheng, ZhipengZeng, ZheyiChen, YuanlongYu, ChunmingRong "Detecting spammers on social networks "journal Neurocomputing 159(2015)27–34.

[2]Atefeh Heydari, Mohammad Ali Tavakoli, Naomie Salim, Zahra Heydari "Detection of review spam: A survey" journal Expert Systems with Applications 42 (2015) 3634–3642.

[3] Fahim A., Mutahira N. Naseem "Facebook Spam and Spam Filter Using Artificial Neural Networks" International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:9, No:1, 2015.

[4]Mohammed N. Al-Kabi, Izzat M. Alsmadi, Heider A. Wahsheh "Evaluation of Spam Impact on Arabic Websites Popularity" Journal of King Saud University – Computer and Information Sciences (2015) 27, 222–229.

[5]Xin Liu, Zhaojun Xin, Leyi Shi, Yao Wang "A Decentralized and Personalized Spam Filter Based on Social Computing" IEEE 2014.

[6]Vipin N S, Abdul Nizar M "A Proposal for Efficient On-line Spam Filtering" First International Conference on Computational Systems and Communications 2014.

[7]Zhipeng Zeng, Xianghan Zheng, Guolong Chen, Yuanlong Yu "Spammer Detection on Weibo Social Network" 2014 IEEE 6th International Conference on Cloud Computing Technology and Science.

ISSN 2348 – 7968

[8]Dae-Ha Park,Eun-Ae Cho, Byung-Won On "Social Spam Discovery using Bayesian Network Classifiers based on Feature Extractions" 2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications.

[9] David Mandell Freeman "Using Naive Bayes to Detect Spammy Names in Social Networks" AISec'13, November 4, 2013, Berlin, Germany.

[10]Sajid Yousuf Bhat, Muhammad Abulaish "Community based features for identifying spammers in online networks" IEEE/ACM International Conference on Advances in social networks analysis and mining,pp.100-107, Aug. 25-28,2013.

[11] Kuan Zhang, Xiaohui Liang, Rongxing Lu, and Xuemin (Sherman) Shen "SAFE: A Social Based Updatable Filtering Protocol with Privacy-preserving in Mobile Social Networks" IEEE ICC 2013- Wireless Networking Symposium.

[12] Geerthik.S "Survey on Internet Spam: Classification and Analysis" IJCTA | May-June 2013.

[13] Rupam Some "A Survey on Social Network Analysis and its Future Trends" International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 6, June 2013.

[14] Helen Costa, Fabricio Benevenuto, Luiz H. C. Merschmann "Detecting Tip Spam in Location-based Social Networks" ACM 978-1-4503-1656-9/13/03.

[15] Yang Yu, Yuzhong Chen "A Novel Content Based and Social Network Aided Online Spam Short Message Filter" Proceedings of the 10th World Congress on Intelligent Control and Automation July 6-8, 2012.

[16] Dave DeBarr, Harry Wechsler "Spam detection using Random Boost" journal Pattern Recognition Letters 33 (2012) 1237–1244.

[17] Xin Jin, Cindy Xide Lin, Jiebo Luo, Jiawei Han "Social Spam Guard:A Data Mining Based Spam Detection System for Social Media Networks" 37th International conference 2011, Vol.4, No.12.

[18] Gianluca Stringhini, Christopher Kruegel, Giovanni Vigna "Detecting Spammers on Social Networks" ACM 978-1-4503-0133-6/10/12.