ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 4, Issue 3 , March 2017

Survey on Efficient Classifier for Detecting Spam in Social Networks

Vishalakshi N S, S.Sridevi

P.G. Student, Department of Computer Science, New Horizon College Of Engineering, Bangalore, Karnataka, India
Assistant Professor, Department of Computer Science, New Horizon College Of Engineering, Bangalore,India

ABSTRACT:Social networking services are used for communication between people to share information through internet.Reaching hundreds millions of users, major social networks have become important target media for spammers. Social networks provide communication between people to share information through internet. The unbounded growth of content and users pushes the Internet technologies usage to certain limitations. The main objective of the proposed work is to find relationship between features and classifying patterns for detecting spam message from the unwanted sites. .In this paper we have reviewed the existing techniques for detecting spam users in social network. Features for the detection of spammers could be user based or content based or both and spam classifier methods.

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

I.INTRODUCTION

Within the past few years, online social network, such as Face-book, Twitter, Weibo, etc., has become one of the major way for internet users to keep communications with their friends. According to Statist report [1], the number of social network users has reached 1.61 billion until late 2013, and is estimated to be around 2.33 billion users globe, until the end of 2017. However, along with great technical and commercial success, social network platform also provides a large amount of opportunities for broadcasting spammers, which spreads malicious messages and behaviour. According to Nexgate's report [2], during the first half of 2013, the growth of social spam has been 355%, much faster than the growth rate of accounts and messages on most branded social networks.

The impact of social spam is already significant. A social spam message is potentially seen by all the followers and recipients' friends. Even worse, it might cause misdirection and misunderstand-ing in public and trending topic discussions. For example, trending topics are always abused by spammers to publish comments with URLs, misdirecting all kinds of users to completely unrelated web-sites.

Because most social networks provide shorten service on URLs inside messages it is difficult to identify the content without visiting the site.

A. Types of Spammers

1**) Spammers**: are the malicious users who contaminate the information presented by legitimate users and in turn pose a risk to the security and privacy of social networks. Spammers belong to one of the following categories.
2) **Phishers**: are the users who behave like a normal user to acquire personal data of other genuine users.
3) **Fake Users**: are the users who impersonate the profiles of genuine users to send spam content to the friends' of that user or other users in the network.
4) **Promoters**: are the ones who send malicious links of advertisements or other promotional links to others so as to obtain their personal information.

B. **Motives of Spammers**
1) Disseminate pornography
2) Spread viruses
3) Phishing attacks
4) Compromise system reputation.

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.

II. 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 III, Methodology is explained in section IV, section V covers the experimental results of the study, and section VI discusses the future study and Conclusion.

III. LITERATURE SURVEY

In the past ten years, email spam detection and filtering mechanisms have been widely implemented. The main work could be summarized into two categories: the content-based model and the identity-based model. In the first model, a series of machine learning approaches are implemented for content parsing according to the keywords and patterns that are spam potential. In the identity-based model, the most commonly used approach is that each user maintains a whitelist and a blacklist of email addresses that should and should not be blocked by anti-spam mechanism. More recent work is to leverage social network into email spam identification according to the Bayesian probability.

The concept is to use social relationship between sender and receiver to decide closeness and trust value, and then increase or decrease Bayesian probability according to these values. With the rapid development of social networks, social spam has attracted a lot of attention from both industry and academia. In industry, Facebook proposes an Edge Rank algorithm that assigns each post with a score generated from a few feature (e.g., number of likes, number of comments, number of reposts, etc.). Therefore, the higher Edge Rankscores, the less possibility to be a spammer. The disadvantage of this approach is that spammers could join their networks and continuously like and comment each other in order to achieve a high Edge Rank score.

Mohammed N et al. developed a top Arabic websites which are selected for evaluating possible web spam behaviour. 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. 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. 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.
ZhipengZeng et al. Survey on supervised machine learning based spammer filtering with SinaWeibo 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 SinaWeibo that includes 30,116 users and more than 16 million messages

Yang Yu et al.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. proposed two methods random project and log it boost which is a combination of random boost. Random boost method improves spam filter compared to log it boost algorithm. Random Boost algorithm reduces training time. Legit boost algorithm uses a greedy approach to learning, focusing on best features for distinguishing spam from non-spam.

Xin Jin et al. 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

IV. METHODOLOGY

Data mining is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and systems. It is an interdisciplinary subfield of computer science. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD.

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 analysing 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 pre-processed 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.

Dataset Description

A data set (or dataset) is a collection of data. Most commonly a data set corresponds to the contents of a single database table, or a single statistical data matrix, where every column of the table represents a particular variable, and each row corresponds to a given member of the data set in question. The data set lists values for each of the variables, such as height and weight of an object, for each member of the data set. Each value is known as a datum. The data set may comprise data for one or more members, corresponding to the number of rows. In total, 8858 spammers and 17646 non-spammers were labelled. Since user labelling 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 labelled dataset as our training data collection, and the rest

A) Data Pre-processing

Data pre-processing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviours or trends, and is likely to contain many errors. Data pre-processing is a proven method of resolving such issues. Data pre-processing prepares raw data for further processing.
Data goes through a series of steps during preprocessing:

- Data Cleaning: Data is cleansed through processes such as filling in missing values, smoothing the noisy data, or resolving the inconsistencies in the data.
- Data Integration: Data with different representations are put together and conflicts within the data are resolved.
- Data Transformation: Data is normalized, aggregated and generalized.
- Data Reduction: This step aims to present a reduced representation of the data in a data warehouse.
- Data Discretization: Involves the reduction of a number of values of a continuous attribute by dividing the range of attribute intervals.

B) 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. Then the data is validated and finally report is being generated.

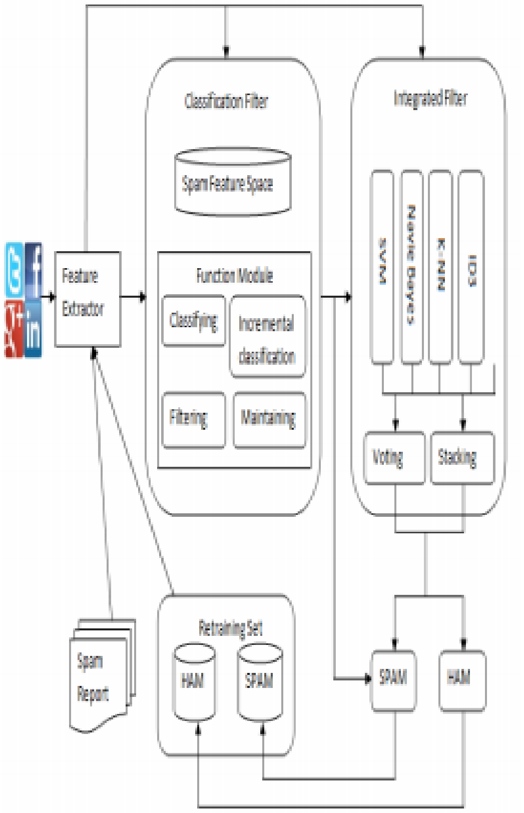


Fig1. System Design

V. EXPERIMENTAL RESULTS

It is implemented on a training dataset consisting of five legitimate messages and five spam message. The true rate and false rate for spam and good messages for the proposed system is calculated from equation .True rate is number of messages truly classified as spam message and good message. False rate is number of messages falsely classified as spam message and good message. Spam Messages:

- True Rate = (No of spam messages truly classified / total no of messages) *100% (1) (4/5) *100% = 80%
- False Rate= (No of spam messages Falsely classified – True rate) *100% (2)(80-60) * 100% = 20% Good Messages:
- True Rate = (No of good messages truly classified/ total no of messages) *100%
- False Rate= (No of good messages Falsely Classified / total no of messages) * 100% (4) (100- 60) = 40%.

Based on the true rate and false rate values of spam and good message, the following graph is generated.

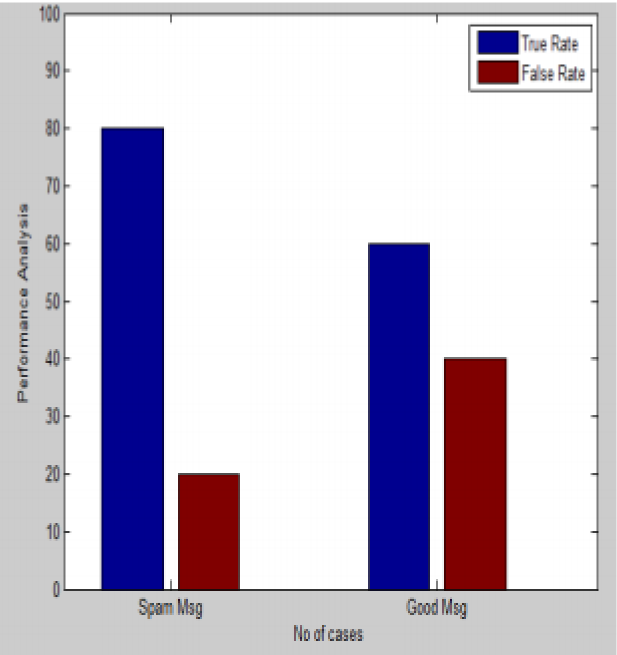
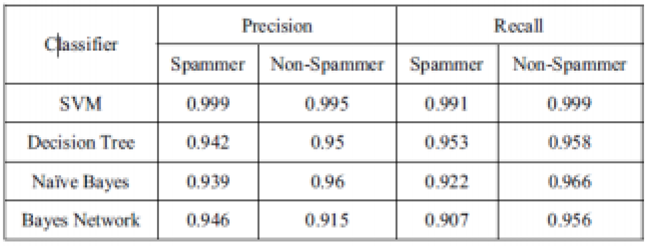


Fig 2: Graph showing True Rate and False Rate for spam and good messages.

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.

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

VI. CONCLUSION AND FUTURE WORK

In order to detect and prevent spammers in social networks several methods have been proposed and developed by many researchers. During our survey it is seen that spam detection in social networks using Decision Tree, SVM, Random Forest and Naïve Bayesian approaches is highly effective and a combination of spam prevention filters will give higher accuracy. 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.Future work

involves to implement a new SVM Kernel which has enlarged dataset for classifying messages which have non-English words and spam messages which are encrypted.

REFERENCES

[1] Agarwal S, Jain. K "Hybrid Approach For Spam Detection using Support Vector Machine and Artificial Immune System", First International Conference on Network and Soft Computing", Aug 2014, pg no: 05-09.

[2] Selamat, Mohammed .M, " An Evaluation on Efficiency of Hybrid Features for Spam Email Classification",2015 International Conference on Computer Communication and Control Technology ,April 2015, pg no : 227-231

[3] "A Hybrid Approach for Spam Filtering using Local Concentration and K- means Clustering", 2014, 5th International Conference, pg no: 194-199.

[4] Salehi, Solmat. A "Hybrid Simple Artificial Immune System and Particle Swam Detection", "5th Malaysian Conference In Software Engineering", Aug 2011, pg no: 124-129.

[5]Xin Liu, ZhaojunXin, 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 Online Spam Filtering" First International Conference on Computational Systems and Communications 2014.

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

[8] A.H. Wang, Don't follow me: spam detection in Twitter, Security and Cryptography (SECRYPT), in: Proceedings of the 2010 International Conference on. IEEE, 2010

[9] H. Gao, Y. Chen, K. Lee, D. Palsetia, A. Choudhary, Towards online spam filtering in social networks, in: Proceedings of the Symposium on Network and Distributed System Security (NDSS), 2012.

[10] F. Benevenuto, G. Magno, T. Rodrigues, V. Almeida, Detecting spammers on Twitter, in: Proceedings of the Seventh Annual Collaboration, Electronic messaging, Anti-abuse and Spam Conference (CEAS), 2010.

[11] Y. Zhu, X. Wang, E. Zhong, N.N. Liu, H. Li, Q. Yang, Discovering spammers in social networks, in: Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI), 2012.