A Web Spam based Identification and Detection for Fake Reviews in Social Network

S. Sabitha[1], S. Ramana[2], M. Rajesh[3], T. Senthil Prakash[4]

[1,2]*PG Student, Department of CSE, Shree Venkateshwara Hi-Tech Engineering College, Gobi, India*

[3]*Assistant professor, Department of CSE, Shree Venkateshwara Hi-Tech Engineering College, Gobi, India*

[4]*Head of the Department, Department of CSE, Shree Venkateshwara Hi-Tech Engineering College, Gobi, India*

Abstract—**In recent days, a huge part of people depend on available content in social media in their decisions (e.g. reviews and feedback on a topic or product). The possibility that anybody can leave a review provides a golden opportunity for spammers to write spam reviews about products and services for different interests. Identifying these spammers and the spam content is a hot topic of research and although a considerable number of studies have been done recently toward this end, but so far the methodologies put forth still barely detect spam reviews, and none of them show the importance of each extracted feature type. In this study, we propose a new framework, named NetSpam, which uses spam features for modeling review datasets as heterogeneous information networks to map spam detection procedure into a classification problem in such networks. Using the importance of spam features help us to obtain better results in terms of different metrics experimented on real-world review datasets from Yelp and Amazon websites. The results show that NetSpam outperforms the existing methods and among four categories of features; including review-behavioral, user-behavioral, review linguistic, user-linguistic, the first type of features performs better than the other categories.**

I. Introduction

1) Various social networks:

A social network is a platform for people sharing their activities, interests, background, and real life connections via specific visual computer techniques. A social network consists of a basic background of each user (often a personal profile), his or her social connections, and a variety of additional information such as career and academic backgrounds [1]. Online social networks (OSNs) have become more and more popular in nowadays society, and it would be hard to get rid of them from normal daily life. The very first online social network is an email where people shared and transferred information via different email addresses. Benefiting from the flourish of smartphones, people have multiples choices of various social network applications or Apps. Facebook, Twitter, Snapchat, Tumblr, Instagram, etc. have been a huge part of people's normal life. People share their personal thoughts, activities, arrangements, and information on daily life via different OSNs. At the same time most of the celebrities, athletes, and politicians always share their activities and information via various OSNs. Information in OSNs had been a major role for international organizations and institutions to publish their statements. Also, with the rapid developments of different functioned OSNs, more and more people probably share their activities in the forms of similar posts during different OSNs because of the various scopes of friends and followers. For example, when Justin Bieber plans to publish his new album, he or his company will post the same or similar content on all his social network accounts to notify all followers about the information of this new album. Sometimes, various OSNs design a new function to share their posts on other social networks. People are even able to use a feature on Facebook to automatically publish updates to their Twitter accounts simultaneously [2]. The similar function can also be designed in other social networks, for example, Tumblr users can share the pictures or information to twitter and Facebook accounts. Most of the web pages have various buttons to allow viewers to share the page to various OSNs. All of these make different OSN accounts for one person exhibit high similarities.

2) Social networks security:

Designers had created various kinds of policies and technologies to prevent the potential spam activities sincere the birthday of the World Wide Web. As the first global social network, the email had huge amounts of spammers' invasions and attacks. A huge amount of methodologies has been used for email spam detection and social network activities. Unfortunately, high prosperity in OSNs gives rich soils for different kinds of spams. Spammers who aim to advertise their products or post victim links are more frequently spreading their malicious activities via different OSNs. The information is not private if the information is posted to a social networking site. The more information users show in social networks; the more vulnerable users may become. Even when using high-security settings, the webpages and various activities will leak users' information to spammers. Personal information users share on OSNs could be used to conduct attacks against their associates. The more information shared via social networks, the more likely that someone could impersonate users and his or her friends into sharing personal information, providing access to restricted sites, or downloading malware. Business competitors, predators, and hackers troll social networking sites looking for information or people to target for exploitation [22]. For social network spam, researchers and scientists had developed several popular theories about spam classification and detection. For content spam detection, researchers use TF-IDF to extract word features combined with term frequency [28], retrieve the similarity between different documents. For link classification, people usually choose PageRank to calculate the link relevance by using hyperlinks. Spammers always have their own methods for the link-based ranking or PageRank policy, they use spam links to improve the score of the target page. By TF-IDF, spam words always connect with the contents so that this spam page show more similarity for some queries [28].

Information gleaned from social networking sites may be used to design a specific attack that does not come by way of the social networking site. Reports show that nearly 10% of tweets in Twitter are all spam [3], and Facebook usually blocks 200 million malicious actions every day [4]. In 2008, a market survey showed that at least 83% of the users of online social networks have received one or more unknown friend request or message. That survey based on the user perception of online social network spam [23]. Even if all companies developed approaches to limit the activities of spammers, spam volume is rapidly growing more than users' actions. Sometimes hackers can modify the code of a social network site, add malicious code into it, always about advertisements and third-party apps ads. On Facebook and Twitter, when log accounts in a work computer, several shortened URLs will point to malicious sites, and that a possible way for personal information leakage. Because it is very easy to retweet a post so that it finally could be seen by hundreds of or thousands of people, Twitter is especially vulnerable to this method [5].

Meanwhile, lacking of social network policy has been a major problem for users. Without authority policy, users cannot have the right ways to protect their interests and privates. Moreover, social network company may have chaos in the protections of users' privates. Sometimes users download more than they needed. In March 2011, Google officially deleted more than 60 applications. Those applications carried illegal or malicious software from Google Android Market [5]. Part of the malicious applications were used to steal the user's private information, then sold to a third, or modify the information or user profile via other devices, or even deleted users' accounts. Most of the social networks had made several methods prevent different kinds of spam. Users in Facebook and Twitter can click the report spam to notify the employees to delete those posts they think as spam. Moreover, those social networks designed their spam filter system to detect kinds of spam. Though these filters had made a tremendous improvement in spam detection, we still discover there are huge amounts of spam appear in various OSNs at various time.

3) Spam detection history of spam detection in email:

A large number of spam detection techniques came from the email classification [35]. Those techniques have four parts: individuals' actions, email administrator's automation, email senders' actions and those of researchers and law enforcement officials. For email spam detection, scientists developed several kinds of techniques to prevent spams:

Checking words: spam can be detected via the contents of actual email, either by detecting keywords such as "sexy picture" (content or non-content based).

Lists of sites: From the emails' end user address or consumer ISP, compared with the information in the DNSBLs, which contains the known name of spammers, open relays, and proxy servers, they can be identified as legal or spam.

Several new approaches have been proposed to improve the email system:

Cost-based systems: this solution requires that senders pay some cost to send email, making it prohibitively expensive for spammers who are eager to send large volumes of emails.

DMARC, which stands for "Domain-based Message Authentication, Reporting & Conformance", it standardizes the performance of email authentication using Sender Policy Framework (SPF) and DKIM mechanisms. Channel email is a new proposal which uses the process of sending an email to restrict anti-spam activities by forcing verification when the first email is received from new contacts.

With the amounts of various social network platforms, spammers have more options and targets to attack. Moreover, that causes the prosperities of spam accounts and malicious posts in OSNs. Lots of scientists and researchers had been focused on this area since a bunch of spam were discovered in OSNs. They also have various kinds of detection methodologies after decades of developments. Many types of research have concentrated on this area to find efficient methods to identify spam and are especially concentrated on the classification of different spam features.

4) Types of spam

The Spam E-mail today comes in different variety which is more harmful and may possess destructive characteristics. Spam is a form of abuse of the Simple Mail Transfer Protocol (SMTP). Most Spam making tools exploits the security holes of SMTP. They do this by forging E-mail headers, disguising sender addresses and hiding the sending system such that, it becomes difficult to identify the true sender. The users get Spam not only through E-mails but even while visiting a blog, chatting with a friend, browsing on community forum and also on mobile phones. These Spam E-mails have lot of forms like, some of these contains advertisements, others provide winning notifications, and sometimes gets messages with executable files, which finally emerge as malicious codes, such as viruses and Trojan horses which can affect the recipients computers.

Problem caused by Spam:

The design principles of the E-mail infrastructure, which were originally intended to provide simplicity and flexibility, have become ambivalent characteristics. The recipient pays a considerable price for receiving these unwanted Spam E-mails. The Spam E-mails causes the following economic harms:

- **Loss of Productivity**

When an employee receives Spam E-mails he/she spend time on opening, reading and classifying these E-mails as Spam and then spend time to delete them.

- **Download Cost**

It costs real money for the receivers to download their E-mails. Since, many receivers are still pay money for the time duration during which the content of mailbox are transferred from ISP to their computer.

- **Harm through malicious code**

Many Spam E-mails contain malicious code such as Viruses, Trojan horses, worms, spyware, adware and key logger. The economic harm that results from the execution of malicious software has not yet quantified.

- **Harm through frauds**

Frauds including phishing are an indirect harm that, Spam E-mail can cause, which has increased in number.

- **Decreases Bandwidth**

The bandwidth is precious resource in a corporate network. Spam E-mail essentially eats up lot of bandwidth. Spam also consumes other valuable computing resources. Smaller organizations working with minimal bandwidth are especially feeling the increasing strain that, Spam E-mail is putting on their network. This in turn reduces the productivity of business organizations as well as of employees. Thus, Spam dissipates employee time, burdens on E-mail servers with a heavy processing load, eats up the disk space on client machines and degrades the overall network performances.

It is a common mistake to believe that, Spam E-mails are sent for marketing or money-making schemes. Some viruses and worms also send out Spam E-mail in order to infect the recipients' computers and carry out Denial of Service Attack (DoS). Nowadays, it is executed in distributed fashion called as Distributed Denial of Service Attack (DDoS) recently, it is found that, the incidents of phishing attacks are increased. The E-mail which carry phishing attack is called as ‘Phishing E-mail’. The Phishers execute phishing attack by hacking important information of users such as E-mail account information, Bank details, Debit/Credit card numbers with PIN. These Phishing E-mails rides on the back of Spam E-mails. So, it is important to block Spam E-mails.

In this context many Spam filtering techniques are deployed at MTA. But majority of it do the mis-classification. Some legitimate E-mails are misclassified as Spam and vice-versa. So it is need of time to block the Spam E-mails only.

5) Spam detections in current OSNs:

X. Hu focused on a content and network information framework for social spammer detection [6]. X. Jin proposed a GAD [7] clustering algorithm to process the challenges about the scalability and real-time detection. B. Markines, C. Cattuto, and F. Menczer focused on six features at the post level, resource level, and user level to specify the spam H. Gao analyzed spam accounts of social networks to identify the percentage of malicious wall posts, compromised accounts and accounts created for the purpose of spamming [9]. C. Grier, etc. tested the usefulness of URL blacklists to intercept the spreading of Twitter spam via the link feature [11]. M. Bosma proposed a framework combined with user features and spam reports to detect spam [14]. J. Song classified the spams based on the relationships connection features between accounts on Twitter [16]. K. Thomas et al. analyzed different features and behaviors via the largest spam campaigns on Twitter accounts [17]. S. Long designed a new methodology combines word-, topic-, and user-based features to stem social spam in YouTube [20]. Y. Zhu used a user-activity count matrix to encode the users’ social activity in Renren [19]. Basing on spam profile features, K. Lee proposed a honeypot-based approach for spam detection in MySpace and Twitter [18] K. Thomas etc. found that it can spare 70% of victims by preventing the spread of compromise in 24 hours [10]. J. Caverlee, L. Liu, and S. Webb proposed a reputation-based trust aggregation framework to test spam in MySpace [15].

However, prior research mainly concentrated on spam detection in one particular social network like Twitter or Facebook, and they paid less attention a popular phenomenon, that is, the more and more similarities between various social networks. Several researchers have focused in this area for a period. De Wang, D. Irani, and C. Pu designed a framework called “SPADE” to deal with spam in different social networks and webs via one framework [12]. It can specify various types of spam-like links or contents in various OSNs via particular models.

As activities between different OSNs establish more connections, OSNs will develop more interactions with each other. Therefore, when one spam link, content or spammer attacks one social network, it is possible to appear on other social networks with similar actions. Therefore, if different social networks’ spam detection models have the ability to communicate with each other, it will greatly decrease the spam actions. Our research mainly focuses on combining spam in one social network to reveal and intercept new similar spam that may appear on other social networks. We analyzed behaviors and features of spam in various OSNs, and then use the similar features to facilitate the spam detection in other OSNs.

II. Problem Definition

Problem definition to develop effective Anti-Spam Framework for Spam management, this will keep Internet E-mail infrastructure alive as a reliable, cost-effective and flexible service. Objectives Analysis of legislative and behavioral measures in Spam management. To propose an effective Spam classification scheme/s in technological Anti-Spam measures. To analyze effectiveness of Anti-Spam measures relative to the identified Spamming options.

To develop a complete infrastructure Framework for Spam management Contribution the problem of Spam E-mail is studied with different measures which include legislative measures, behavioral measure and technological measures. Since, legislative measures and behavioral measures are related to different stakeholders, the research is focused on technological measures after studying legislative and behavioral measures. After carrying out the study of legislative measure, it is suggested that, in India it is need to have separate Anti-Spam legislation. There should be an online reporting mechanism for reporting the incidents of Spam.

The online reporting mechanism would serve as a data collection tool which would be useful for training the Content based Filters. In the study of behavioral measures the content analysis of Spam E-mail is carried out and some important observations are made which are utilized to propose technological solution.

The Content based Filter with Machine Learning (ML) based classifier is implemented. The empirical analysis of Spam E-mail is carried out by using classifiers such as Decision Tree, Rough Sets using various rule generation methods (Genetic Algorithm, Linear Algorithm, Covering Algorithm and Exhaustive Algorithm), k-Nearest Neighbor, and Support Vector Machine classifiers with various kernel functions (Linear, Multi-Layer Perceptron’s, Quadratic Functions and Radial Basis Function).

The Content based Filter using semantic similarity with edge based approach classifier is implemented which finds the semantic similarity between the words of E-mail to be tested with known Spam and Ham words.

The decision is made depending on the shortest path length between the words. If the path length between the words of E-mail to be tested and known Spam words is shortest, then E-mail is classified as Spam E-mail. Similarly, if the path length between the words of E-mail to be tested and known Ham words is shortest, then E-mail is classified as Ham E-mail. The path similarity method is used on 'WordNet' dictionary to find semantic similarity between the words. . The combination of Origin based Filter along with Content based Filter is implemented which is adaptive in nature. Finally, a complete Anti-Spam Framework is proposed for effective Spam management.

III. Literature Survey

This section describes prior and some important research work done to solve the problem of Spam E-mails. This chapter briefly reviews the important contribution done by the researchers in Origin based Filters, including Black-listing and Whitelisting of domain names / IP addresses and Challenges Response Systems. Also it reviews the Content based Filters which includes Machine Learning based Classifier and Semantic Similarity with Edge based Classifier. This review found to be very useful in comparing the results.

Initially, certain researchers concentrated on the development of Honey pots to detect spams. To detect spams, **(Webb et al., 2008)** dealt with automatic collection of deceptive spam profiles in social network communities based on anonymous behavior of user by using social honey pots. This created unique user profiles with personal information like age, gender, date of birth and geographic features like locality and deployed in MySpace community. Spammer follows one of the strategy such as being active on web for longer time period and sending friend request. The honey profile monitors spammer's behavior by assigning bots. Once the spammers sends friend request the bots stores spammer profile and crawls through the web pages to identify the target page where advertisements originated. The spammer places woman's image with a link in the "About Me" section in its profile and the honey profile bots crawls through the link, parses the profile, extracts its URL and stores the spammers profile in the spam list. URL does not redirect at times during crawling process and "Redirection Detection Algorithm" is executed to parse the web page and extract redirection URL to access it with the motive of finding the source account. Also, he proposed a "Shingling Algorithm" which verifies the collected spam profile for content duplication like URL, image, comments and to accurately cluster spam and non-spam profile based on the features. In this way, he eliminated Spams.

Another researcher named (Gianluca et al., 2010) used social honey pots to construct honey profiles manually with features like age, gender, DOB, name, surname etc. Here, the honey profiles have been assigned to three different social network communities (Myspace, Facebook, and Twitter). It considered friend request as well as the message (wall post, status updates) received from spammers and validate with honey pots. The identified spam accounts with the help of spam bots share common traits which has been formalized as features in their honey pots (first feature, URL ratio, message similarity, friend choice, message sent, friend number etc). The classifier namely Weka framework with a random forest algorithm has been used to classify spammers for best accuracy. Similarly during spam campaign the spam bots were clustered based on spam profiles using naïve Bayesian classifier to advertise same page during message content observation.

(Lee et al., 2010) dealt with the social spam detection which has become tedious in social media nowadays. Here Social honey pots are deployed after its construction based on the features such as number of friends, text on profile, age, etc. Here, both legitimate and spam profiles have been used as initial training set and Support Vector Machine has been used for classification. An inspector has been assigned to validate the quality of extracted spam candidates using "Learned classifier" and provide feedback to spam classifier for correct prediction in future. In this study three research challenges have been addressed. Initially, it validates whether the honey pot approach is capable of collecting profiles with low false positives, next to that it addresses whether the users are correctly predicted and finally it evaluates the effectiveness of fighting against new and emerging spam attacks. The first challenge is proved using automatic classifier which groups the spammers accurately. The second one considers demographic features for training the classifier using 10-fold cross validation. It has been tested in MySpace using Meta classifier. In twitter it used Bigram model for classification along with the preprocessing steps. Finally, post filters has been used to check the links and remove the spam label by applying "Support Vector Machine" for correct prediction. They also proposed that in future, Clique based social honey pots can be applied with many honey profiles against many social network communities. Next to honey pots, Spammers have been identified in the literature by analyzing content and link based features of web pages.

(Sreenivasan and Lakshmipathi, 2013) has performed spam detection in social media by considering content and link based features. Web spam misleads search engines to provide high page rank to pages of no importance by manipulating link and contents of the page. Here, to identify web spam, Kullback-leiblerence techniques are used to find the difference between source page features (anchor text, page title, and Meta tags) and target page features (recovery degree, incoming links, broken links etc). Therefore, three unsupervised models were considered and compared for web spam detection. As a first one, Hidden Markov model has been used which captures different browsing patterns, jump between pages by typing URL's or by opening multiple windows. The features mentioned above were given as input to HMM and it is not visible to the user. As a result, a link is categorized as spam or non-spam based on how frequently a browser moves from one page to another. Second method uses "Self Organizing maps" a neural model to classify training data (Web links) without human intervention. It classifies each web link as either spam or non-spam link. One more method called Adaptive Resonance Theory has also been used to clarify a link as either spam or not.

(Karthick et al., 2011) has dealt with the detection of link spam through pages linked by hyperlink that are semantically related. Here, Qualified Link Analysis (QLA) has been performed. The relation existing between the source page and target page is calculated by extracting features of those two pages from web link and compared with the contents extracted from these pages. In QLA, the nepotistic links are identified by extracting URL, anchor text and cached page of the analyzed link stored in the search engines. During query generation, once the page is available with search engines, this result has been compared together with the page features for easy prediction of spam and non-spam links. In this study, QLA has been combined with language model detection for better prediction of spams. In Language model detection, the KL divergence technique has been used to calculate the difference between the information of the source pages with the content extracted from the link. Once matched, it is clustered as non-spam and vice versa. Here, the result of LM detection, QLA along with pre trained link and content features lead to accurate classification and detection.

Qureshi et al. (2011) handled the problem of eliminating the existence of irrelevant blogs while searching for a general query in web. The objective is to promote relevancy in ordering of blogs and to remove irrelevant blogs from top search results. The presence of irrelevancy is not because of spam, but is due to inappropriate classification for a topic against a query. This approach uses both content and link structure features for detection. The content features calculate the cosine similarity between blog text and blog post title while searching for a particular blog. It has been proved that a co-relation exist between the above two features with which the spammer activity is detected based on the degree of similarity. This detection achieved a precision of 1.0 and recall of 0.87. The blog link structure feature finds spammer activity by decoupling between two classes (duplicate and unique links) up to three hop counts. Spammers always move within closed group rather than with other blogosphere. The duplicate links are identified and removed.

Wang and Lin (2011) focused on comment spams with hyperlinks. The similarity between the content of page for a post to the link it points to has been compared to identify spam. Here, the collected blogs are preprocessed which finds the stop word ratio that is found to be less in spammers post. The contents are extracted from the post and are sorted where "Jaccard and Dice's "co-efficient is calculated which provides the degree of overlapping between words. The degree of overlapping is used for calculating inter comment similarity for a comment with respect to a post. Analysis of content features like inter comment similarity and post comment similarity along with the non-content features like link number, comment length, stop words showed better results in identifying spam links. Next to this, comment based spams have also been discussed here.

Archana et al. (2009) has dealt with the spam that gets penetrated in the form of comments in Blog. A blog is a type of web content which contains a sequence of periodic user comments and opinions for a particular topic. Here, spam comment is an irrelevant response received for a blog post in the form of a comment. This comments are analyzed using supervised and semi supervised methods. Analysis considers various features to identify spams. They are listed below: The post similarity feature has been used to find the relevancy between the post and the comment. Word Net tool has been used to spot out the word duplication features. Word duplication feature identifies the redundant words in comments and it is found to be higher for spam comments and low for genuine comments. Anchor text feature counts the number of links exists for a comment and predicts that the spammers are the one having higher count. Noun concentration feature has been used to extract comments and part of speech tags from the sentences. In that, the legitimate users have low noun concentration. Stop word ratio feature consider sentences with a finishing point where spammers have less stop word ratio. Number of sentence feature counts the number of sentences exists in a comment and is found to be higher for spammers. Spam similarity feature checks for the presence of spam words listed and categorize it. The words identified as spam after preprocessing were assigned a weightage and the contents which falls above the threshold are detected as spam comments. Here, a supervised learning method (Naïve Baye's classifier) has been used along with pre classified training data for labeling a comment as spam and non-spam. One more unsupervised method directly classifies the comments based on the expert specified threshold. Interestingly in literature, works have been carried out for book spammers also.

Sakkura et al. (2012) deals with bookmark spammers who create bookmark entries for the target web resource which contains advertisements or inappropriate contents thereby creating hyperlinks to increase search result ranking in a search engine system. Spammer may also create many social bookmark accounts to increase the ranking for that web resource. Therefore, user accounts must be clustered based on the similarity between set of bookmarks to a particular website or web resource and not based on the contents. Here in this study, data preprocessing is done by clustering bookmarks by extracting web site URL from the raw URL since spammer may create different bookmark entry for same URL. Here, the similarity based on raw URL (which is the ratio of number of common URL'S to total number of URL'S contained in the bookmarks of two accounts) has been considered and the similarity based on site URL without duplicates (which is the ratio of number of common site URL'S to the total number of all URL's in both the accounts) and the similarity based on site URL with duplicates (Weight of the sites based on the number of bookmarks common to the user accounts) has been calculated. The agglomerative hierarchal clustering of accounts has been made based on one of the above mentioned similarities. The cluster which is large and having higher cohesion is categorized as an intensive bookmark account spammer. This study achieves a precision of 100%. Yang and Chen (2012) this study deals with online detection of SMS spam's using Naïve Bayesian classifier, which considers both content and social network SMS features. The SMS social network is constructed from the historical data collected over a period with the help of telecom operator. The content features are extracted that are presented in vector space and the weights are assigned to the vector obtained using term frequency function. The feature selection methodologies like information gain and odd ratio has been used for selecting words from SMS with which class dependency and class particularity are found for clustering "content based features". Features on social network tries to extract both the sending behavior of mobile users and closeness for categorizing spammer and legitimate user. Bloom filter is used to test the membership between sender and receiver for removing spammer's relationship. Naïve Bayesian classifier has been used for classifying users as legitimate or spam using the above features.

Ravindran et al. (2010) deals with the problem of tag recommendation face which contains popular tags for particular bookmarks based on user feedback and to filter spam posts. In this problem, Spammer may increase the frequency of a particular tag and the system may suggest those tags which have higher frequency to the user. To eliminate this problem, this study uses “frequency move to set “model to choose a set of tags suggested by user for a bookmark. To find whether a tag is popular or not for placing it in the suggestion set, the tag feature like simple vocabulary similarity has been considered. The suggestion set which is kept updated is measured using the stagnation rate and unpopular tags are removed randomly from the set. The decision tree classifier has been used here to classify tags as spam and non-spam. The accuracy obtained in this approach is about 93.57%.

Ariaeinejad and Sadeghian (2011) deals with detecting email spam in an email system by considering plain text alone for categorizing a mail as spam or ham. The common words in spam and ham emails are eliminated and stored in white list. The collected words are parsed by removing unwanted spaces and other signs among the words. The parsed words are compared with white list and common words are eliminated. The cleaned words are checked for making decision using “Jaro-Wrinkler” technique. Here, a fuzzy map is constructed as a two dimension using an interval type and 2 fuzzy methods have been used which represents distance of each word in email with closed similarity in dictionaries as a horizontal vector and represents weight of the words in dictionary as an vertical vector. Third dimension considers importance of a word in an email and its frequency which is identified using term frequency inverse document frequency technique. Here, Email has been categorized into spam, ham and uncertain zone using fuzzy-C means clustering. Later, the words are updated consistently for correct prediction.

Amlan Mohanty, 2011 have described the various amendments made to the IT (Amendment) Act, 2000. He has analyzed the legislative response to cybercrime in India with analysis of the Information Technology (Amendment) Act, 2008. The amendments of new crimes are examined by the author. Kigerl, 2014 have addressed the spammer behavior, spam volume, spam compliance, spam locality with the CAN SPAM Act by analyzing the 5,490,905 samples of spam emails received in the United States from March, 1998 to November, 2013. Govt. of India, 2013 have defined Spam as “the use of E-mail systems to send unsolicited bulk E-mails, especially advertising, indiscriminately. This notification has advice the government servants not to use their official E-mail addresses to subscribe on any unsafe or fake website. Ramasubramanian and Prakash, 2013 have presented a brief report which has discussed the historical growth, spread of spam and Internet abuse which includes telemarketing and mobile spam messaging, in India. They have also addressed the current and proposed Indian law which includes cybercrime.

FTC, 2005 have presented some schemes which include the impact Spam, spyware which try to steals the user’s data and crashes their computers. Also, telemarketing and health claims which hits the weaknesses of users. The FTC in its report has recommended for making proper provisions to fight against these issues. IT Act, 2000, The Indian Government has passed an information Technology law to address the issues related to cybercrime. Information technology this has not address the problem of Spam. Amend_IT ACT, 2008] The amendments made to IT ACT, 2000 includes the issues rated to digital secure electronic signature, protecting data, provision of compensation in case of failure of data are made on 5th February, 2009.

Coello, 2005 has studied the current technical and legislative solutions which are proposed by the governments and private companies to fight the problem of Spam. Robinson et.al, 2011 have presented a comparative study on legislative and non-legislative measures used to combat identity theft and identity-related crime. This comparative study includes the definition and context of identity theft. They have summarized the overall legislation made by different countries.

Karen Ng, 2005 have presented a review on various options for controlling the Spam in context with Canada. He has recommended two important points which should be considered in future Canadian Spam legislations, which government of Canada is competent to control and regulate the problem of Spam and have suggested to cover charter scrutiny under this law.

Moustakas et.al, 2005 have presented an overview of different anti-spam laws including weaknesses of these laws, to fight against problem of Spam and have compared this with US and European Union anti-Spam Legislation.

Hohlfeld et.al, 2012 have presented a work which find out the origins of the spamming process which concentrated on address harvesting on the web. They have suggested that, simple methods of obfuscations are still efficient for protecting addresses from being harvested.

Wang et.al, 2006 have investigated the use of Hill Climbing, Simulated Annealing, and Threshold Accepting optimization techniques as feature selection algorithms. They have compared the performance of these techniques with the Linear Discriminate Analysis. Their results show that all these techniques can be used not only to reduce the dimensions of the E-mail, but also to improve the performance of the classification filter. Among all the strategies, simulated annealing has the best performance which reaches a classification accuracy of 95.5%. Lai, 2007 has carried out systematic experiments of machine learning techniques such as, SVM, k-NN and Naive Bayesian on different parts of an E-mail. His experiments show that, using the header part only one can achieve satisfactory performance and the idea of integrating disparate methods is a promising way to fight spam. He has carried out a comparative study on performance of these machine learning methods in spam filtering.

El-Halees, 2009 have compared six supervised machine learning classifiers which are maximum entropy, decision trees, artificial neural nets, Naive Bayes, support system machines and k-nearest neighbour. He has applied these methods on stemmed Arabic spam words. He showed that, for most cases, classifiers using feature selection techniques can achieve better performance than the filters which do not use it. Romero et.al, 2010 used blog comment corpus as a case study in which they have used 50 pages and 1024 blog comments. The percentage of spam of this corpus is 67%. They have applied and compared the result of four machine learning techniques such as Naïve Bayes, K-nearest neighbour, neural networks and the support vector to classify blog comments as spam and non-spam. Khan et.al, 2010 have reviewed important techniques and methodologies that are employed in text documents classification based on existing literature. They have focused mainly on text representation and machine learning techniques.

Caruana and Li, 2012 have reviewed emerging approaches to spam filtering built on recent developments in computing technologies which includes peer-to-peer computing, grid computing, semantic Web, and social networks. They have addressed a number of perspectives related to personalization and privacy in spam filtering. Finally they have concluded that, important advancements have been made in spam filtering in recent years, high performance approaches remain to be explored due to the large scale of the problem.Aski, 2013 have proposed an algorithm to classify E-mails and minimize spam using nearest neighbor classifier. This approach involves low computational load in relatively high rate relying on a Hash table as well as a flag varying in the range of {1, 0}.Uysal, 2013 have investigated the impact of several feature extraction and feature selection approaches on filtering of short message service (SMS) spam messages in two different languages Turkish and English. The entire feature set of filtering framework consists of the features originated from the bag-of-words (BoW) model along with the ensemble of structural features (SF) specific to spam problem. The distinctive BoW features are identified using information theoretic feature selection methods. Various combinations of the BoW and SF are then fed into widely used pattern classification algorithms to classify SMS messages. They have evaluated filtering framework on both Turkish and English SMS message datasets which is Turkish SMS message collection. They have demonstrated that, the combinations of BoW and SFs, rather than BoW features alone, provide better classification performance on both Turkish and English message datasets.

Pawlak,1982 have presented an alternative to fuzzy set theory and tolerance theory using approximate operation on sets, equality of sets and on inclusion of sets have analyzed spam filtering technology by carrying out detailed study of Naive Bayes algorithm, and proposed the improved Naive Bayesian mail. Saab et.al, 2014 have presented a survey of some popular filtering algorithms that rely on text classification to decide whether an email is unsolicited or not. This algorithm is executed on the Spam base dataset to identify the best classification algorithm in terms of accuracy, computational time, and precision and recall rates. Peace et.al, 2015 have carried out the comparative analysis of prediction success using rapid miner tool. In this process of prediction, they have applied machine learning classifiers such as k-NN and ANN on English Premier League dataset. They have proved that, ANN outperforms k-NN with prediction success of70%. Elavarasi et.al, 2014 have presented a comprehensive survey of semantic similarity measures with various approaches including path based measures, information based measures, feature based measures and hybrid measures. All these methods are discussed with their advantages, disadvantages, features and other issue. Kim et.al, 2007 have presented a new approach different from content based method which uses user preferences for constructing an anti-spam mail system. They have constructed user preference ontology using important concept. Using inference engine they have proved that their method gives good performance. Kiamarzpour et.al, 2013 have proposed new method for classifying the spam Email. They have showed high accuracy of E-mail classification by using the several decision trees in combination with ontology.

IV. Existing System

Image spam is a kind of E-mail spam where the message text of the spam is presented as a picture in an image file. The basic rationale behind image spam is that it is difficult to detect using text spam filtering methods designed to detect patterns in text in the plain-text E-mail body or attachments. A new trend in email spam is the emergence of image spam. Although current anti-spam technologies are quite successful in filtering text-based spam emails, the new image spams are substantially more difficult to detect, as they employ a variety of image creation and randomization algorithms. Many anti-spam systems also use a combination of whitelists, blacklists, and so-called greylists that force legitimate clients to re-send messages since spammers often do not bother doing so. Other common techniques include block lists distributed via DNS that identify addresses assigned to dialup users or known open relays and challenge-response systems that automatically build whitelists. Most systems such as Mail Avenger, Spam Assassin, and SpamGuru use multiple techniques, including multiple classifiers, to identify spam. Filters for text-based spam, including plain text and HTML e-mail, have employed a variety of statistical techniques, particularly Bayesian inference; these statistical filters appear to classify text-based e-mail well. The situation has significantly frustrated end-users as many image spam messages are able to defeat the commonly deployed anti-spam systems. In order to reduce the impact of spam, it is crucial to understand how to effectively and efficiently filter out image spam messages. Spammers have recently begun developing image-based spam methods to circumvent current anti-spam technologies since existing anti-spam methods have proved quite successful at filtering text-based spam email messages. Once spammers have applied an image creation algorithm to make a message difficult to detect with computer vision algorithms, they apply further randomization to construct a batch of images for delivery. The result is that current image spam methods present serious challenges for anti-spam systems. We believe that an effective image spam detection system should satisfy several requirements. First, it should be accurate, detecting most image spams while maintaining a low false positive rate. Second, it should be efficient, parsing incoming emails with images at modern WAN speeds. Third, it should be extensible, allowing new image spam filtering methods to be added to deal with quickly evolving image spam techniques.

Disadvantages:

1. Text classification method - Fails for image spam detection.
2. Negative reviews can potentially impact credibility and cause economic losses.
3. Spammers to write fake reviews designed to mislead users' opinion.
4. The fact that anyone with any identity can leave comments as review, provides a tempting opportunity for spammers to write fake reviews designed to mislead users' opinion. These misleading reviews are then multiplied by the sharing function of social media and propagation over the web.

5. Many aspects have been missed or remained unsolved.
6. Previous works also aimed to address the importance of features mainly in term of obtained accuracy, but not as a build-in function in their framework (i.e., their approach is dependent to ground truth for determining each feature importance).

Hardware Specification:

Processor	: Intel I3
Speed	: 1 MHz
RAM capacity	: 2GB MIN
Hard disk drive	: 2TB MIN
Monitor	: 15 inch color monitor
Keyboard	: HP 104 Keys
Mouse	: HP Optical

Software Specification:

Front-end used	: JAVA
Back-end used	: MY SQL
Operating System	: Windows XP

V. Proposed System

An application programming interface (API) is a set of procedures, protocols, and tools; it is used for construct various applications and software. Social network platforms offer APIs to users to develop various new web applications. That will benefit its programming structure for outside groups to utilize and create new features to their websites [21]. An API usually consist of an operating system, a web-based system, or a database tool, and always based on a specific programming language. It is useful for developing applications for the different system. APIs can work as the GUI components, or to access computer hardware or database like the hard disk driver. Through various APIs, third parties and researchers have access to the instant data, user activities, celebrities' actions and the most popular topics in the world. In this section, we will introduce the background information about Facebook API and Twitter API, and the datasets collected during the research and then classify research goal before we analyse the datasets.

Advantages:

1. Written reviews also help service providers to enhance the quality of their products and services. Spam deteriorates the quality of search result and deprive legitimate websites of revenue.
2. Spam have economic impact since a high ranking provides large free advertising and so an increase in web traffic volume.
3. The trust of a user in search engine provider which is especially tangible issue due to zero cost of switching from one search provider to another. Spam websites are means of malware and adult content dissemination and Phishing attack. Spam forces a search engine company to waste a significant amount of computation and storage resource.
4. To identify spam and spammers as well as different type of analysis on this topic.

Contributions:

Our research's major contributions are as follows

1. We propose a new perspective of the spam detection in online social networks. Traditional detection methods are focused on only one social network. However, our work concentrates on spam similarities in different OSNs to analyse and detect such activities.
2. We collected two datasets from Twitter and Facebook through their APIs, each of them contains spam and non-spam contents.

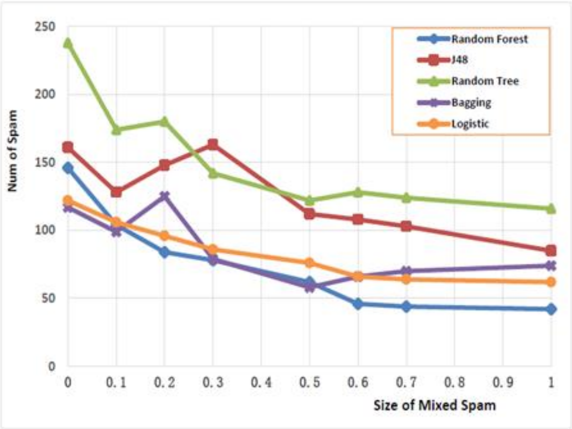

Fig. 1. Size of Spam vs. No. of Spam

VI. Conclusion

Online social networks spam detection and classification have been a popular topic in the science and technology areas. Scientists and researchers pay lots of attention in it to build a more developed and convenient visual world to human beings. This research makes a new progress in this topic and proposes a new point of spam detection.

In this whole research, we introduce a new perspective to distinguish between spam and legitimate contents in Twitter and Facebook, the top two most popular social networks in the world. For research convenience, we collected two new datasets through their APIs via similar topics and users group. When collected Twitter dataset which the keyword was set as "Taylor Swift" on Twitter from Sep 2017 to August 2018, and then get the Twitter Spam Dataset (TSD). After our labeling and normalizing, we got this dataset that consists of 1937 spam tweets and 10942 ham tweets. For Facebook, We collected data from the open public group on Facebook, which was named as "World of Taylor Swift" from Sep.

References

[1] Wikipedia, "Social network service," 2016. [Online].Available:https://en.wikipedia.org/wiki/Social_networking_service. [Accessed: 27-July-2016]

[2] B. Logan, "Publishing to twitter from facebook pages," 2009. [Online].Available:https://www.facebook.com/notes/facebook/publishing-to-twitter-from-facebook pages/123006872130. [Accessed: 27-July-2016]

[3] N. Ungerleider, "Almost 10% of Twitter is spam," 2014. [Online].Available:http://www.fastcompany.com/3044485/almost-10-of-twitter-is-spam. [Accessed: 27-July-2016]

[4] Avoiding Social Spam Hackers on Facebook and Twitter. [Online]. Available: http://www.sileo.com/social-spam/. [Accessed: 27-July-2016]

[5] C. Nerney, "5 top social media security threats," 2011. [Online].Available:http://www.networkworld.com/article/2177520/collaboration-social/5-top-social-media-security-threats.html. [Accessed: 27-July-2016]

[6] X. Hu, J. Tang, and H. Liu, "Online social spammer detection," *AAAI'14 Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, pp. 59-65, Québec City, Québec, Canada, July 27 - 31, 2014.

[7] X. Jin, C. Lin, J. Luo, and J. Han, "A data mining-based spam detection system for social media networks," *Proceedings of the VLDB Endowment*, vol. 4, no. 12, pp. 1458-1461, 2011.

[8] B. Markines, C. Cattuto, and F. Menczer, "Social spam detection," *In Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web*, 2009, pp. 41-48.

[9] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao, "Detecting and characterizing social spam campaigns," *In Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, 2010, pp. 35-47.

[10] K. Thomas, F. Li, C. Grier, and V. Paxson. "Consequences of connectivity: Characterizing account hijacking on Twitter," *In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, 2014, pp. 489-500.