MEDWELL PUBLICATIONS SINCE 2000
ASIAN JOURNAL OF
INFORMATION TECHNOLOGY
Journal Home Page: https://medwelljournals.com/journalhome.php?jid=1682-3915

Survey on Detecting Real Spammers and Consequences of Cyber Attacks in Social Networks

J.S. Harilakshmanraj and S. Rathi
Department of Computer Science and Engineering, Government College of Technology, Anna University, 641038 Coimbatore, Tamil Nadu, India

Key words: Cybernetics, social networking, spam, cyber-bullying, cyber-stalking, systematic

Corresponding Author:
J.S. Harilakshmanraj
Department of Computer Science and Engineering, Government College of Technology, Anna University, 641038 Coimbatore, Tamil Nadu, India

Page No: 164-172
Volume: 18, Issue 6, 2019
ISSN: 1682-3915
Asian Journal of Information Technology
Copy Right: Medwell Publications

Abstract: Online social network is one of the widely used information-sharing channels. Around, billions of users share their contents, views and upload millions of multimedia contents to share with others every day. Cybercrimes like cyber-bullying and cyber-stalking activities are major threats to society. Social spam is another major issue present over OSNs environment. In this study, we present a systematic review over different cyber crimes happens over OSNs which threaten OSN users in general as well as children in particular. In addition, we reviewed over social spam and its impacts. Existing solutions for cybercrimes like cyberbullying and cyberstalking have been reviewed. Some spam detection strategies and methodologies have been discussed. This study also discusses and reviews the solutions for cyberbullying, cyberstalking and spam detection.

INTRODUCTION

The usage of online social networks have been increasing widely and these networks are interlinked with every people’s life to share their thoughts, view and for other communication purpose (Chin *et al*., 2015). They have become extremely popular in the last few years. Every person spends huge amounts of time in OSNs making friends with people who they are known with or interested in forming some forum for further interaction. Twitter which was founded in 2006 has become one of the most popular microblogging service site. Nowadays 200 million Twitter users generate over 400 million new tweets per day.

Due to huge growth of social network sites, cyber security has become one of the major things to concern for users and enterprises alike. While communication technologies have changed their pattern of communication, it also provides way to cybercriminals with some strategies and techniques to be used for illegal purposes such as the spreading of offensive messages and threatening content (De Vel *et al*., 2001), sending spam messages, phishing attack, cyberbullying, viruses, harassment and cyberstalking (Reynolds *et al*., 2011). Cyberbullying and cyberstalking affect large number of individuals unlike many other cybercrimes (Bollen *et al*., 2011).

Cyberbullying has grown as a social threat, it majorly affects children as well as young adult. Majority of the workplaces and some corporate firms have also been affected by cyberbullying (Vandebosch and Cleemput, 2009). According to recent studies, almost 43% of teenagers are victims of cyberbullying during various scenario (Baer, 2010). According to the statement of American Academy of Child and Adolescent Psychiatry, cyberbullying creates emotional psychological suffering among OSNs users. Many of these victims are ended in suicides due to critical nature of this problem.

Existing OSNs suffer from offensive behavior of the users who are able to use OSNs to disallow, interrupt, degrade and delude others in various occasions having a non-negligible impact over various services as well as in government sector. Consequently, Twitter has recently introduced some changes on its user policy in an attempt to settle issue of abuse over content posted by the user. Various studies have stated that cyber-bullying is a major threat to harass another person through any form of digital communications. This behavior is intended to harm the

Table 1: Types of cyber-bullying exist in online social networks

Issues	Description
Flooding	Process of sending abuse messages, comments frequently by the predator in order to distrait the victim and not allowing the victim to participate in the online forum and social groups
Masquerade	Predators encapsulate their original identification and impersonate them as other person. But in reality they are not been unique to identify (Lenhart *et al*., 2010)
Flaming	Sending and posting offensive and violate text electronically to one or more than one user either publicly or privately (Bhat and Abulaish, 2013)
Trolling	It is otherwise known as baiting, Poster have certain motive to publish comments, messages which disagree with other user comments or messages and intended to annoy an argument of the victim (Bhat and Abulaish, 2013)
Harassment	Sending illegal and violence Text_Content electronically to someone continuously (Lenhart *et al*., 2010)
Cyber stalking	It is a fact of sharing offensive messages and comments rapidly, which harm someone physically (Lenhart *et al*., 2010)
Exclusion	In this victim is consciously excluded from an online forum as well as social groups (Benevenuto *et al*., 2010)

dignity or image of the target victim (Hadjidj *et al*., 2009; Vandebosch and Van Cleemput, 2008). An internet “trolling” or cyber-troll is someone who according to Langos (2012) is member of an online forum, posts offensive comments at worst or absonant information at the best to create controversy.

Social network platform provides huge opportunities for spammers who spread malicious messages. During first half of 2013, growth of spam rate was 35% which is much faster than messages that have been sent via various social media. Social spams are sent by followers and recipient’s friends. OSNs provide little support to prevent spam messages on user walls. For example, Facebook allows users to decide who is allowed to post content in their walls (i.e., friends, friends of friends or defined groups of friends). However, no content-based preferences are supported and therefore it is not possible to prevent unwanted messages. Therefore, spams have been spread across various social networking sites and some strategies are being followed to trace and control malicious content.

This study reviews about existing techniques, methodologies and strategies used for controlling cyberbullying and cyberstalking as well as detecting and classifying spam content spread via branded social network sites.

Major issues in online social networks: Cyber bullying differs from traditional bullying in fact. Cyberbullying extends from physical limits of public places like schools, parks, etc., with the victim often experiencing suspension from it (Robert and Doyle, 2003). In the case of cyber bullying, the culprits have ability to harass the victim without any hesitation as the bullying is done through online and does not need any physical presence of the victim. Another major problem when it comes to cyberbullying is the lack of identifiable parameters which notify any post as a bullying instance. Even after identifying bullying, predicting the harness of the instance is a challengeable task as it can be simple name calling leading to social exclusion or uploading offensive pictures that might result in worst consequences (Ogilvie, 2000).

A victim can be exposed to various instances of cyber bullying over various modes available through online and large-set of audience which can witness these instances makes it even more immoral and unpleasent (Dinakar *et al*., 2012). One of the majorly used forms of cyberbullying is posting of harmful comments about someone in social networks. Identification of cyberbullying activity is one of the main courses of actions to battle with Misbehavior in social networking sites. There are two kinds of entities involved in cyber-bullying.

Cyber predator: It is an individual targets over teenagers with the help of Internet. They victimize and threaten them by means of Text_Message. These predators are best in influencing based on building association among the victim. Their motivation is to fulfill their sexual, personal or financial needs to improve their living style.

Cyber victim: Any person who is hassle over the Internet by means of texts, videos, images etc. This term generally represents teenagers who are cyber bullied in cyber world (Table 1). Cyberstalkers most often utilize an large number of technologies, tools and techniques like chat rooms, bulletin boards, newsgroups, Instant Messaging (IM), Short Message Service (SMS), Multimedia Messaging Service (MMS) and trojans, e-mail are most commonly used methods for cyberstalking activity (Bose and Shin, 2006; Sabella *et al*., 2013). They used to send e-mails, SMS, IM, MMS and chat to threaten, insult, harm or disrupt e-mail communications by swamping victim’s e-mail inbox with malicious mail (Sheridan and Grant, 2007; Bakar, 2013). This creates a new challenge for law enforcement and in digital forensic investigation. Anonymity in communication is one of the major issues victimized by cybercriminals (Langos, 2012). Therefore, cyberstalkers could easily conceal themselves by spoofing email and creating different anonym accounts mostly from free web mail providers. Similarly web based gateways are used to spoof SMS (Ogilvie, 2000) and different anonymous chat IDs are easily created.

Other major issue over online social network is “Spam”. It can be present in multiple forms such as images, text, videos etc. Social Networking Sites (SNS) need to be repudiated for long-term accomplishment. If there is any source page representing a corporate firm or a brand over some social media, it has to be protected or else it will damage their reputation. Spams consist of virus links which could lead to personal or business.

Spammers used to attack widely used social networks site. Finally, it increases the rate of victim of social spam. Social spam is un-rated information over social networks, it is same as email spam and it is unbiased bulk messages source that users do not ask for any permission or to specify to subscribe to it. Such, spam is offensive to people and they try to block them from consuming information that is relevant to user. Individual social networks are capable of filtering a significant amount of the spam they receive. They need some large amounts of resources (e.g., personnel) and incur a delay before finding new types of spam.

Finally, Spam is not a new term and it has been present, since, from traditional e-mail. In traditional e-mail networks, the major form of spamming depends on Random Link Attack (RLA) strategy where a less number of spammers send spam to a large number of randomly selected victim nodes (user present over network). Spammers are most important person to generate spam messages to a socially un-related set of receivers, unlike legitimate senders whose receivers tend to cluster or form communities (Stein *et al*., 2011). Twitter also contributes to the growth of spam. Twitter spam which is referred as unbiased tweets containing malicious links that directs victims to external sites containing malware downloads, phishing, drug sales or scandals, etc. (Ortega, 2013). It has not only contrived number of legitimate users but also affected the whole social networking framework. Many spam/spammer detection methods have been proposed in literature.

MATERIALS AND METHODS

Cyberbullying detection methodologies: Facebook Immune System (FIS) is an automated learning system to find abuse sources. It relies on information from user activity logs to find malicious behavior in OSNs framework. This system found about 20% of the deceitful profiles. It has some significant number of false negatives. To deal with cyberbullying, graph-based research methods is more apparent compared to traditional text based approaches using Natural Language Processing (NLP). Text based approaches uses texture-based methods for leveraging texture features and trained classifiers. At Initial stage, some naive texture features like Strokelets (Reynolds *et al*., 2011), T-HOG (De Vel *et al*., 2001), etc. are retrieved to describe text candidate regions and retrieved texture features are sent to trained classifiers, e.g., neural networks (Baer, 2010), Support Vector Machine (SVM) (Bollen *et al*., 2011), etc., to identify the text candidates. Graph based techniques have been useful for detecting combating dishonest behavior (Cao *et al*., 2012) as well as cyberbullying (Boshmaf *et al*., 2015). It also detect fake accounts in OSNs (Alowibdi *et al*., 2014). These methods use machine learning techniques using social-graph metadata in their feature set. This methods are used for detecting fake accounts details (Ott *et al*., 2012), Gender classification over Twitter profiles (Chin *et al*., 2015). It is used to find misleading profile based upon profile attributes of an user.

Garcia-Recuero (2016) developed an “abuse classifier”. Tweet data-set has been collected based on public twitter API. Using this classifier, Relative Importance (RI) of each of the features within the categories were described. For each instance, the Random Forest (RF) learning algorithm (Minnich *et al*., 2015) have been used to highlight the relative importance of each feature, during the decision making process (classify as abusive or not). It is based on a given threshold which is a cutoff value in prediction probability after which the classifier identifies a tweet as potentially malicious. In order to capture the trade-off between True Positive Rate (TPR) and False Positive Rate (FPR) in a single curve, the Receiver Operating Characteristics (ROC) analysis provides possibility of visualize trade-off results from different threshold values.

Hosseinmardi *et al*. (2015) automatically detect incidents of cyberbullying over images that present over Instagram. Large samples of Instagram dataset were collected. It consist of images and appropriate comments. Label-based strategy has been used to cyberbullying as well as image content using human labelers present at Crowdflower Web site. An analysis of the labeled data includes study of correlations between different features and cyberbullying as well as cyber-aggression. Trust-based metric is been used for classification design. Detailed analysis of the distribution results of the labeling of cyberbullying incidents were presented. It include correlation analysis of cyberbullying with other factors derived from message sources. Naives-Bayes and Linear SVM were been incorporated to identify cyberbullying incidents as well as it improves the detection of cyberbullying incidents.

Dadvar and Niamir (2016) proposed Maximum Entropy method (MaxEnt) to identify the bully (Cyberbullying task) users in YouTube. This method are multi-variant distribution of incidents found over feature-space were been estimated based upon principle of maximum entropy. It has shown best approximation of an unknown source distribution along with maximum entropy (the most spread out) subject to known constraints. These constraints are defined by the expected value of the distribution which have been estimated from set of incidents. MaxEnt method was used for identifying bulled users. Using feature profile-set, calculation of probability of user being bulled is compared with multi-model algorithm namely Generalized Linear Model (GLM), random forest, Support Vector Machine (SVM) along with MaxEnt method also been used. MaxEnt method is more reliable compared to other models. This method can be adapted to other social networks site to find bully user in OSNs environment.

Mishra *et al*. (2015) conducted a survey on adolescents. Relationship among cyber bullying, trust and level of information shared among user in OSNs environment. Main cause of cyber bullying using four phases starting from building a relationship to cyberbullying are been developed. Four phases are connection phase, causal information sharing phase, closer relationship and Identifying cyberbullying. Finally, degree of cyber bullying among adolescents with respect to different level of shared information have been calculated. Survey state that adolescents shared their personal information with trusted ones, i.e. as the level of information consecutively increases. But still they were lot of victims found. Reflection of victimization rate is directly propositional to the level of information shared is very high. Rate of victimization is higher when they share high weighted information to OSNs environment. It indicates that chances of victimization increases with the weight of information shared and it also state that chances of cyber bullying increases when there is increase in trust. Main reason was increase in trust because adolescents share high weighted information to trusted person who can be potential cyber attackers.

Cyberstalking: Spitzberg *et al*. (2007) detected three levels of stalking facts; at first order effects are the facts depends on victim and may include combat on the individual's health issues (fear, loss, suicidal ideation anxiety, shame, disturbances, impaired psychological, well-being depression, sleep) social health issues (decreased trust, increased alienation and isolation, restricted social activities), resource health issues (additional security measures, absenteeism from work), cognitive health issues (maladaptive beliefs, attributions of self-blame, personality adaptation), physical health issues (physical and sexual violence) or resilience. Stalking may also result in behavioural or general disturbance (Spitzberg and Cupach, 2007). It is victims can be exposed to extended periods of stalking who experience highest rates of psychiatric unhealthy, irrespective of the nature of the prior relationship with culprit. Some consequences state that stalking will the nature of cross-sectional design of most stalking studies. It does not enable causal interpretations to OSNs user.

Ghasem *et al*. (2015) propose a framework known as Anti Cyberstalking Text-based System (ACTS). It is the first framework that specializes over automatic detection and evidence documentation of text-based cyberstalking. It make use of various text mining strategies, text mining, statistical analysis, text categorization and machine learning to scrap cyberstalking. It consists of five main modules; Detection, Attacker identification, Personalization, Aggregator and messages and Evidence collection (Table 2).

Table 2: Comparison of different methodologies/algorithms to prevent cyberbullying

Proposal	Methodologies/algorithm	Advantage	Disadvantages	Accuracy
01.	Facebook immune system	It is used to predict abusive behavior completely It is automated system for detecting malicious post over OSNs environment	Abusive users can create fresh accounts in order to start abusing again. These kind of automated or semi-automated methods are not perfect	20% of the deceitful profiles they deployed were actually detected
02.	Graph-based methodology	This method make use of machine learning techniques using social-graph metadata in their feature set. For example to detect fake accounts	The classification problems may cause some conflict while classifying the content. Example it may never be state whether a message is really abusive or fake	It has the higher value of metric in terms of detecting malicious post normalized Discounted Cumulative Gain (nDCG) of 0.588
03.	Naives-bayes Linear SVM classifier	It incorporate multi-classification model that use of variety of features to identify cyberbullying incidents	Enlargement of labelled dataset substantially not suitable. It doesn't incorporate image features	Accuracy of identifying cyberbullying is been 87%
04.	Maximum Entropy methodology (MaxEnt)	It is very robust to limited amount of training data. It is well regularized	This model doesn't suitable for balanced datasets or rare number of target incidents	Accuracy level is 75%. Number of profane words has highest contribution (~ 33%), number of subscription had the least contribution to the model (~ 1%)

This system tries to detect cyberstalking based on a Message_ID list which is automatically updated by the system. Messages whose IDs do not appear in the list are verified by identification, personalisation and detection modules. The results from these three modules are sent through to the aggregator for final verdict.

Text categorization based email detection systems used to detect unwanted e-mail, same like detection module is deployed to detect cyberstalking text based on their source content. The received message is processed by utilizing tokenization, stop-word removal, stemming and presentation. Text mining techniques are been applied to retrieve required patterns from the content. Supervised algorithm like neural network (or) support vector machines are used to detect and categorize message to compute value based on three outputs result such as (00) not cyberstalking, (10) cyberstalking and (01) grey email.

The attacker identification module is deployed to identify whether received source content are sent by cyberstalker or not as well as it detect messages from cyberstalkers where the message does not contain any unwanted contents. For this purpose, cyberstalker's writeprints includes lexical, syntactic and structural and content-specific features will be used. Final module messages and evidence collection module regularly update stylometrics, profiles and related information about cyberstalking message to the database.It uses statistical methods like multivariate Gaussian distribution and PCA to examine writeprint and profiles of cyberstalking and text mining to retrieve similar features about attacker behavioural.Anonymous message and non-anonymous message are been classified. The integrity and authenticity of a cyberstalking message are achieved by using hash functions and asymmetric encryption keys.

RESULTS AND DISCUSSION

Spam detection strategies: There are lot of strategies used to identify spam and spammers (Jindal and Liu, 2008; Rayana and Akoglu, 2015). These methodologies can be classified into different forms namely; linguistic patterns, behavioral patterns. The linguistic patterns (Xu and Zhang, 2015; Breiman, 2001; Viswanath *et al.*, 2014) depends on bigram and unigram. The behavioral patterns are based on features extracted from patterns in user behavior (Xu and Zhang, 2015; Breiman, 2001; Viswanath *et al.*, 2014; Minnich *et al.*, 2015; Akoglu *et al*., 2013; Li *et al*., 2014). This techniques can be used for supervised and un-supervised learning approaches for reviewing class-labels. Graphs and graph-based algorithms are also used in spam identification (Heydari *et al*., 2015; Crawford *et al*., 2015; Jindal and Liu, 2008; Rayana and Akoglu, 2015; Shamash, 1974; Hutton and Friedland, 1975).

Shehnepoor *et al.* (2017) introduced novel spam detection framework known as NetSpam which depends on metapath concept as well as a graph-based method to label reviews. Rank-based labeling approach is used for ranking the reviews and calculating Average Precision (AP) and Area Under the Curve (AUC) based on reviews ranking present in final list. These framework have been evaluated using two real-world labeled datasets of Yelp and Amazon websites. Based on their observation shown that metapath concept can be very effective in identifying spam reviews and leads to have better performance. In addition, they stated that without using trainset, NetSpam can calculate the importance of each feature and it has shown quality performance in the features addition process. The results stated that different supervisions, similar to the semi-supervised method have no predicted effect on finding most of the weighted features found in different datasets.

Santosh *et al.* (2017) proposed ENWalk framework that uses content information to fabricate a random walk of the network and find the latent features embedded in the nodes of the network. This framework produces the biased random walks and uses them to maximize the likelihood of obtaining similar nodes in the neighborhood of the network. They conducted study over twitter content dynamics that could be vital to bias those random walks. They classify spammers as; follow-flood and vigilant. They state that success rate, activity window, fraudulence and mentioning behaviors can be used to compare the equivalence of users in the twitter API. They determined network equivalence using these four behavioral features between pairs of nodes and try to bias the random walks with interaction proximity of the pair of nodes. Finally, experimental results showed that this approach significantly outperforms over existing state-of-the-art approaches for deception detection.

Yu *et al*. (2017) proposed anti-spam research work to combine the message content, user behavior and social network structure to perform social media spammer detection accurately and reliably. They used novel based semi-supervised social media spammer detection approach, depends on the message content and user behavior as well as the social relation information.Initially adaptation over original Constrained NMF-based semi-supervised learning (CNMF) algorithm and Nonnegative Matrix Factorization (NMF) by enforcing label information constrain and sparseness constrain were been done. Second, they state novel CNMF-based integral framework for spammer detection by implementing the collaborative factorization over message content matrix as well as over user behavior and social relation information matrix. They explore iterative Update Rule (IUR) and optimization algorithm for the spammer detection model. Additionally, its correlative convergence is also proven.

Experiments are conducted over dataset which is crawled from SinaWeibo system using developed webcrawlers. Experiment result shows that NMF based model outperforms over conventionally applied supervised classifier.

Zheng *et al*. (2016) proposed Extreme Learning Machine (ELM) based supervised machine algorithm for spammer detection. It use systematic strategy for spam detection. Training datasets are converted into a series of feature vectors that includes set of formulated attribute values. These vectors construct the input value of a supervised machine learning algorithm. After training, classification model is used to differentiate based on user's behavior. Whether specific user belongs to either a normal user or spammer. Because spammers and non-spammers have different social behaviors and capabilities, it is able to distinguish abnormal behaviors from legitimate ones. A set of features is retrieved from message content and user behavior were applied to them, using ELM-based spammer classification algorithm. Experiment and evaluation showed excellent performance with a true positive rate of spammers and non-spammers reaching 99 and 99.95%, consecutively. It also showed that solution could achieve better reliability and feasibility compared with existing SVM-based approaches.

Wu *et al*. (2017) proposed classification method based on deep learning algorithms to address and detect spam in the twitter environment. Initially they collected part of labeled data around (376,206 spam and 73,836 non-spam tweets) from a dataset with more than 600 million real-world tweets. They utilized Word Vector (Word2Vec) technique for pre-processing and they converted them into high-dimension vectors. This deep learning method has quality in finding spam messages.

Gao *et al*. (2012) proposed online spam filtering system that has component of the OSN platform to inspect messages generated by users in real-time scenario. They reconstructed spam messages into group for classification rather than examining them individually. Although, campaign identification have been used for offline spam analysis. They applied this technique to solve online spam detection problem with effective low overhead. This system adopts set of novel features that effectively distinguish spam campaigns. It eliminates messages classified as "spam" before they reach the intended receiver, thus, safeguarding them from various kinds of malicious activities. Experiment were conducted using 187 million wall posts collected from Facebook and 17 million tweets collected from Twitter. In different parameter settings, the true positive rate reaches 80.9% while the false positive rate reaches 0.19% in the best case. It stated accuracy for more than 9 months after the initial training phase.It constantly secure OSNs environment without need of frequent re-training. Finally, they tested over server machine with eight cores (XeonE5520 2.2Ghz) and 16GB memory, system achieves an average throughput of 1580 messages/sec and an average processing latency of 21.5 msec over Facebook dataset.

Dewan and Kumaraguru (2015) characterized dataset of 4.4 million public posts formulated over Facebook during 17 news-making events (natural calamities, terror attacks, etc.) and traced 11,217 malicious posts containing URLs. They found that most of the malicious content which states about Facebook's detection techniques depends upon third party web-based applications while more than half of all legitimate content found over mobile applications. They observed greater participation of Facebook pages in generating malicious content as compared to legitimate content. They proposed an extensive feature set based on entity profile, textual content, metadata and URL features to automatically identify malicious content over Facebook in real time. These features set were used to train multiple Machine Learning (ML) models and achieved accuracy_rate of 86.9%. This model was used to create REST API as well as browser plug-in to detecting malicious Facebook posts in real time scenario (Table 3).

Tan *et al*. (2013) propose a sybil defense based spam detection scheme SD2 that significantly outperforms than other supervised schemes. It make use of social

Table 3: Comparison of algorithm/methodology to control cyberstalking

Proposals	Algorithm/methodologies	Advantages	Disadvantages	Performance metric
01.	Noval based spam detection "NetSpam"	It improves the accuracy less complexity Identify spam review based on number of reviews	This method is not suitable for determining most of the weighted features. It is not suitable approach for heterogeneous datasets	Complexity analysis: Degree of spam = O(e^2m) where m is number of features. In online mode, complexity is O(e). Offline mode -O(em)
02.	ENWalk	This method is used for classification and ranking tasks It is most scalable method for biased learning	Predicting of unbiased learning nodes cannot be determined using this method	It has higher AUC (Area Under Curve) of 0.6335 compare to other models like PageRank and Markov random field models
03.	Novel based semi-supervised social media spammer detection approach	It classification of spammers is achieved using CNMF semi-supervised learning (CNMF) algorithm	It doesn't suite for content and characteristics based classification	It has convergence performance of 93% over spammer detection

Table 3: Continue

Proposals	Algorithm/methodologies	Advantages	Disadvantages	Performance metric
04.	ELM-based supervised machine	High efficiency, easy-implementation, unification of classification, easy to be implemented in social spammer detection field	It is not suitable for automatic feature learning and extraction. Dealing with big data, it has low adaptability and expensive	True positive-99.9% False negative-0.1% False positive-0.05 True negative-99.95
05.	Novel technique based on deep learning techniques	It focus and inspect over shortened URLs inside Tweets. This method is more robust features in order to prevent feature fabrication	Low stability, less performance degradation occur over random sampled dataset	Precision-25% higher than the second place on dataset 2 but 5% less than other datasets. It achieved double f-measure of Naive Bayes (frequencies) on dataset 2 and 4
06.	Online spam filtering system	High accuracy, no need for all campaigns to be present over training set no need for frequent re-training and low latency	This method not suitable for image-based spam detection	Higher true positive rate of 80.8%. Lower true positive rate (38.3%) highest false positive rate 0.04%
07.	Multiple-machine learning models	This method automatically identifies malicious content over Social Networking Sites (SNS). Very accessible, very efficient and usability	It is insufficient technique for crowd sourcing and bias label dataset	Higher true positive rate-97.7%. High false negative rate-61.7%. Accuracy- 86.9%
08.	Unsupervised social network spam detection scheme (UNIK)	It has ability to automatically extract spam signatures. It used to find new patterns of spam content	It is suitable for private network, more complex	To detect spammers with a false positive rate of 0.6% and a false negative rate of 3.7%. Detecting spam post false positive rate is 3.7% and the false negative rate is 1.0%

network relationship for schema formation. Robustness is achieved by increasing level of spam attacks. Unsupervised spam detection scheme known as UNIK is been used. Inspect of detecting spammers directly, UNIK removes non-spammers from the network, leveraging both the social graph and the user-link graph in the network scheme. Justification of UNIK is that spammers constantly change their pattern to escape from detection. Non-spammers doesn’t perform any offensive activity. They have relatively non-volatile pattern.

CONCLUSION

This study discussed cyber issues such as cyberbullying, cyberstalking and spamming. The various techniques and strategies used for detecting and preventing the above issues have been reviewed and the performance of the methodologies are also compared and tabu******lated. The results of this review may be used in developing methodologies to strengthen the OSN user security.

REFERENCES

Akoglu, L., R. Chandy and C. Faloutsos, 2013. Opinion fraud detection in online reviews by network effects. Proceedings of the 7th International AAAI Conference on Weblogs and Social Media (ICWSM’13), June 2013, AAAI, Menlo Park, California, USA., pp: 1-10.

Alowibdi, J.S., U.A. Buy, S.Y. Philip and L. Stenneth, 2014. Detecting deception in online social networks. Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM’14), August 17-20, 2014, IEEE, Beijing, China, pp: 383-390.

Baer, M., 2010. Cyberstalking and the internet landscape we have constructed. Virginia J. Law Technol., Vol. 15, No. 154.

Bakar, H.S.A., 2013. Investigating the emergence themes of Cyberbullying phenomenon: A grounded theory approach. Proceedings of the 2013 IEEE 63rd Annual Conference on International Council for Education Media (ICEM), October 1-4, 2013, IEEE, Singapore, pp: 1-14.

Benevenuto, F., G. Magno, T. Rodrigues and V. Almeida, 2010. Detecting spammers on twitter. Proceedings of the International Conference on Collaboration, Electronic Messaging, Anti-Abuse and Spam (CEAS’10) Vol. 6, July 13-14, 2010, Redmond, Washington, USA., pp: 1-12.

Bhat, S.Y. and M. Abulaish, 2013. Community-based features for identifying spammers in online social networks. Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), August 25-28, 2013, IEEE, New Delhi, India, ISBN: 978-1-4503-2240-9, pp: 100-107.

Bollen, J., H. Mao and X. Zeng, 2011. Twitter mood predicts the stock market. J. Comput. Sci., 2: 1-8.

Bose, A. and K.G. Shin, 2006. On mobile viruses exploiting messaging and bluetooth services. Proceedings of the 2006 Securecomm and Workshops, August 28-September 1, 2006, IEEE, Baltimore, USA., pp: 1-10.

Boshmaf, Y., M. Ripeanu, K. Beznosov and E. Santos-Neto, 2015. Thwarting fake OSN accounts by predicting their victims. Proceedings of the 8th ACM International Workshop on Artificial Intelligence and Security, October 2015, ACM, New York, USA., pp: 81-89.

Breiman, L., 2001. Random forests. Mach. Learn., 45: 5-32.

Chin, C.P.Y., N. Evans and K.K.R. Choo, 2015. Exploring factors influencing the use of enterprise social networks in multinational professional service firms. J. Organizational Comput. Electron. Commerce, 25: 289-315.

Crawford, M., T.M. Khoshgoftaar, J.D. Prusa, A.N. Richter and H. Al Najada, 2015. Survey of review spam detection using machine learning techniques. J. Big Data, Vol. 2, 10.1186/s40537-015-0029-9

Dadvar, M. and A. Niamir, 2016. Adopting maxent to identification of bullying incidents in social networks. Proceedings of the 2016 27th International Workshop on Database and Expert Systems Applications (DEXA'16), September 5-8, 2016, IEEE, Porto, Portugal, pp: 186-189.

De Vel, O., A. Anderson, M. Corney and G. Mohay, 2001. Mining e-mail content for author identification forensics. ACM SIGMOD Rec., 30: 55-64.

Dewan, P. and P. Kumaraguru, 2015. Towards automatic real time identification of malicious posts on Facebook. Proceedings of the 2015 13th Annual International Conference on Privacy, Security and Trust (PST'15), July 21-23, 2015, IEEE, Izmir, Turkey, pp: 85-92.

Dinakar, K., B. Jones, C. Havasi, H. Lieberman and R. Picard, 2012. Common sense reasoning for detection, prevention and mitigation of cyberbullying. ACM. Trans. Interactive Intell. Syst. (TIIS.), 2: 1-30.

Gao, H., Y. Chen, K. Lee, D. Palsetia and A.N. Choudhary, 2012. Towards online spam filtering in social networks. Proceedings of the 19th Annual International Symposium on Network & Distributed System Security (NDSS-2012), February 05-08, 2012, San Diego, California, USA., pp: 1-16.

Garcia-Recuero, A., 2016. Discouraging abusive behavior in privacy-preserving online social networking applications. Proceedings of the 25th International Conference Companion on World Wide Web (WWW'16), April 11-15, 2016, Montreal Quebec Canada, pp: 305-309.

Ghasem, Z., I. Frommholz and C. Maple, 2015. A machine learning framework to detect and document text-based cyberstalking. Proceedings of the LWA 2015 International Workshops on KDML, FGWM, IR and FGDB, October 7-9, 2015, Trier, Germany, pp: 348-355.

Hadjidj, R., M. Debbabi, H. Lounis, F. Iqbal, A. Szporer and D. Benredjem, 2009. Towards an integrated e-mail forensic analysis framework. Digital Investigation, 5: 124-137.

Heydari, A., M. Ali Tavakoli, N. Salim and Z. Heydari, 2015. Detection of review spam: A survey. Expert Syst. Appl., 42: 3634-3642.

Hosseinmardi, H., S.A. Mattson, R.I. Rafiq, R. Han, Q. Lv and S. Mishra, 2015. Detection of cyberbullying incidents on the instagram social network. Social Inf. Networks. Vol. 1,

Hutton, M.F. and B. Friedland, 1975. Routh approximations for reducing order of linear time invariant systems. IEEE Trans. Autom. Control, 20: 329-337.

Jindal, N. and B. Liu, 2008. Opinion spam and analysis. Proceedings of the 2008 International Conference on Web Search and Data Mining, February 11-12, 2008, ACM, Palo Alto, California, ISBN:978-1-59593-927-2, pp: 219-230.

Langos, C., 2012. Cyberbullying: The challenge to define. Cyberpsychology Behav. Social Networking, 15: 285-289.

Lenhart, A., K. Purcell, A. Smith and K. Zickuhr, 2010. Social media & mobile internet use among teens and young adults. Pew Internet & American Life Project, Millennials USA. https://eric.ed.gov/?id=ED525056

Li, H., Z. Chen, B. Liu, X. Wei and J. Shao, 2014. Spotting fake reviews via collective positive-unlabeled learning. Proceedings of the 2014 IEEE International Conference on Data Mining, December 14-17, 2014, IEEE, Shenzhen, China, pp: 899-904.

Minnich, A.J., N. Chavoshi, A. Mueen, S. Luan and M. Faloutsos, 2015. Trueview: Harnessing the power of multiple review sites. Proceedings of the 24th International Conference on World Wide Web (WWW'15), May 18-22, 2015, ACM, At Florence, Italy, pp: 787-797.

Mishra, M.K., S. Kumar, A. Vaish and S. Prakash, 2015. Quantifying degree of cyber bullying using level of information shared and associated trust. Proceedings of the 2015 Annual IEEE India Conference on (INDICON'15), December 17-20, 2015, IEEE, New Delhi, India, pp: 1-6.

Ogilvie, E., 2000. The internet and cyberstalking. Proceedings of Criminal Justice Responses Conference, December 7-8, 2000, Australian Institute of Criminology, Sydney, pp: 1-7.

Ortega, F.J., 2013. Detection of dishonest behaviors in on-line networks using graph-based ranking techniques. AI Commun., 26: 327-329.

Ott, M., C. Cardie and J. Hancock, 2012. Estimating the prevalence of deception in online review communities. Proceedings of the 21st International Conference on World Wide Web, April 16-20, 2012, ACM, Lyon, France, pp: 201-210.

Rayana, S. and L. Akoglu, 2015. Collective opinion spam detection: Bridging review networks and metadata. Proceedings of the 21th ACM Sigkdd International Conference on Knowledge Discovery and Data Mining (KDD'15), August 10-13, 2015, Sydney, Australia, pp: 985-994.

Reynolds, K., A. Kontostathis and L. Edwards, 2011. Using machine learning to detect cyberbullying. Proceedings of the 2011 10th International Conference on Machine Learning and Applications and Workshops Vol. 2, December 18-21, 2011, IEEE, Honolulu, USA., pp: 241-244.

Sabella, R.A., J.W. Patchin and S. Hinduja, 2013. Cyberbullying myths and realities. Comput. Hum. Behav., 29: 2703-2711.

Santosh, K.C., S.K. Maity and A. Mukherjee, 2017. Enwalk: Learning network features for spam detection in twitter. Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation (SBP-BRiMS'17), July 5-8, 2017, Springer, Washington, USA., pp: 90-101.

Shamash, Y., 1974. Stable reduced-order models using Pade-type approximations. IEEE. Trans. Autom. Control, 19: 615-616.

Shehnepoor, S., M. Salehi, R. Farahbakhsh and N. Crespi, 2017. NetSpam: A network-based spam detection framework for reviews in online social media. IEEE. Trans. Inf. Forensic. Secur., 12: 1585-1595.

Sheridan, L.P. and T. Grant, 2007. Is cyberstalking different?. Psychol. Crime Law, 13: 627-640.

Spitzberg, B.H. and W.R. Cupach, 2007. The state of the art of stalking: Taking stock of the emerging literature. Aggression Violent Behav., 12: 64-86.

Stein, T., E. Chen and K. Mangla, 2011. Facebook immune system. Proceedings of the 4th International Workshop on Social Network Systems, April 10, 2011, ACM, New York, USA., pp: 1-8.

Tan, E., L. Guo, S. Chen, X. Zhang and Y. Zhao, 2013. Unik: Unsupervised social network spam detection. Proceedings of the 22nd ACM International Conference on Information and Knowledge Management, October 27-November 01, 2013, ACM, New York, USA., ISBN: 978-1-4503-2263-8, pp: 479-488.

Vandebosch, H. and K. Van Cleemput, 2008. Defining cyberbullying: A qualitative research into the perceptions of youngsters. Cyberpsychol. Behav., 11: 499-503.

Vandebosch, H. and K. van Cleemput, 2009. Cyberbullying among youngsters: Profiles of bullies and victims. New Media Soc., 11: 1349-1371.

Viswanath, B., M.A. Bashir, M. Crovella, S. Guha, K.P. Gummadi, B. Krishnamurthy and A. Mislove, 2014. Towards detecting anomalous user behavior in online social networks. Proceediongs of the 23rd USENIX International Security Symposium (USENIX Security'14), August 20-22, 2014, San Diego, California, USA., pp: 223-238.

Wu, T., S. Liu, J. Zhang and Y. Xiang, 2017. Twitter spam detection based on deep learning. Proceedings of the Australasian Multiconference on Computer Science Week, January 30-February 03, 2017, ACM, New York, USA., pp: 1-8.

Xu, C. and J. Zhang, 2015. Combating product review spam campaigns via multiple heterogeneous pairwise features. Proceedings of the 2015 SIAM International Conference on Data Mining, June 2015, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, USA., pp: 172-180.

Yu, D., N. Chen, F. Jiang, B. Fu and A. Qin, 2017. Constrained NMF-based semi-supervised learning for social media spammer detection. Knowl. Based Syst., 125: 64-73.

Zheng, X., X. Zhang, Y. Yu, T. Kechadi and C. Rong, 2016. ELM-based spammer detection in social networks. J. Supercomputing, 72: 2991-3005.