International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 07 | July 2018 www.irjet.net p-ISSN: 2395-0072

Spam URL Detection and Image Spam Filtering using Machine Learning

Parth Parekh[1], Kajal Parmar[2], Pournima Awate[3]

[1]UG Scholar, Computer Engineering Department, Shah and Anchor Kutchhi Engineering College, Mumbai, Mumbai University, Maharashtra, India
[2]UG Scholar, Information Technology Department, Terna Engineering College, Mumbai, Mumbai University, Maharashtra, India
[3]UG Scholar, Computer Engineering Department, K. J. Somaiya Institute of Engineering and Information Technology, Mumbai, Mumbai University, Maharashtra, India

---

Abstract - *The increasing volume of pernicious content in social media requires automate methods to detect and eliminate such content. This paper describes a superintend machine learning classification model that will be built to detect the distribution of pernicious content in online social networks/medias(ONSs/OMSs). Multisource features have been used to detect social network posts that contain vitriolic Uniform Resource Locators (URLs). These URLs could direct users to websites that contain malignant content, drive-by download attacks, phishing, spam, and scams. For the data collection stage, the Twitter streaming application programming interface (API) was used and Virus Total was used for labeling the dataset. The fraudulent practice of sending emails is a criminal scheme to get the user's personal data and other login and confidential information. It is known as phishing that acquires users private information such as password, bank account detail, credit card number, financial username and password etc. and later it can be mistreat by attacker .We aim to use fundamental visual features of a web page's appearance as the basis of detecting page similarities. We propose a novel solution, to efficiently detect phishing web pages. Note that page layouts and contents are fundamental feature of web pages' appearance. Since the standard way to specify page layouts is through the style sheet (CSS), we develop an algorithm to detect similarities in key elements related to CSS. In this paper, we proposed a system that uses SVM technique along with Image Spam Filtering, spam map reduce archetype to achieve a higher accuracy in detection of the spam urls and image spamming.*

Keywords: Phishing, Map reduce, URL Spam Detection, Image Spam Detection, OCR , SVM, Image Spam Filtering.

1. INTRODUCTION

The main challenges for social network security administrators are not only protecting the social network management system and database, but also protecting OSN users from being exposed to malicious content that is spread over those social networks. 60% of social network users have received or been exposed to malicious content [1] such as spam, scams, and drive-by downloads. A number of OSNs are now developing malicious content detection systems for such attacks e.g. the Facebook Immune System detects suspicious activities such as like-jacking, social bots, and fake content [2].

An identity theft that occurs when a malicious web site masquerades a legitimate one is called Phishing. Such a theft occurs in order to procure sensitive information such as passwords, bank account details, or credit card numbers. Phishing makes use of spoofed emails which look exactly like an authentic email. These emails are send to a bulk of users and appear to be coming from legitimate sources like banks, e-commerce sites, payment gateways etc. The makers of such illegitimate website made them exactly look like a legitimate one so that no user can identify the difference easily. The phishing attackers use different kind of social engineering tactics to lure users for example: giving attractive offers to just visit the site.

Malicious URL is a URL created with malicious purposes, among them, to download any type of malware to the affected computer, which can be contained in spam or phishing messages, or even improve its position in search engines using Black hat SEO techniques.

Smart Malicious Url Detection System is an anti-phishing technique to safeguard our web experiences. Our approach uses the Chinese Image Spamming Lexical features, Host based features and site popularity features of a website to detect any suspicious or phishing website. These features are obtained from the source code by taking URL as input and then these features are fed to the classifier algorithm. The results obtained from our experiment shows that our proposed methodology is very effectual for preventing such attacks and the performance was measured by using Confusion Matrix for all the classifiers.

2. Related Works

The majority of studies in this area aim to find the most predictive features that they can acquire and the best algorithm to develop a classifier model [16]. Researchers in this field focus mainly on finding novel features with high discriminative power in addition to coming up with the most accurate machine learning model [17]. Finding high discriminative features in the area of Internet security and social networks is quite a challenge due to the

variation in attacks and techniques used by spammers. Due to the inventiveness of spammers detection systems are bypassed after some time and the set of features used for spam detection has to be regularly revised [18][19]. Similar to how security researchers study the attacks, spammers and hackers investigate detection systems; therefore, they can change user properties, content or the distribution mechanism to bypass certain restriction or detection rules [20]. For example, a study of detecting spam on Twitter [21] recommended that the number of followers is one of the highest discriminative power features. The feature's discriminative power has been increasingly weakened though by spammers making their accounts more popular. They do this by conducting spam campaigns that make their "fake" accounts connect with other fake accounts, increasing the follower and following numbers [22].

Bur nap et al. [04] used an entirely different method to detect malicious URLs. They deployed a high-interaction honey-net2 to collect system state changes, such as the sending/receiving packets and CPU usage. The training dataset contained 2,000 examples with a 1:1 ratio for spam/non-spam. Ten attributes were used to build a classifier that reflected system status changes after opening the tweet's URL. Bur nap et al. investigated the shortest time required to give a preliminary warning of the existence of malicious content in a particular URL. The best result was reported for Multilayer Perceptron (MLP) using features acquired after 210 seconds (0.723 in the F-measure metric). The features used by Bur nap et al. require complex data analysis; however, they make it difficult for spammer sites to disguise their true nature. Although the recent literature has compared several algorithms, there is a lack of information about important stages in building a machine learning model. In particular, little information is provided about how feature selection methods are managed and how parameter tuning is conducted. We address this issue in section IV.

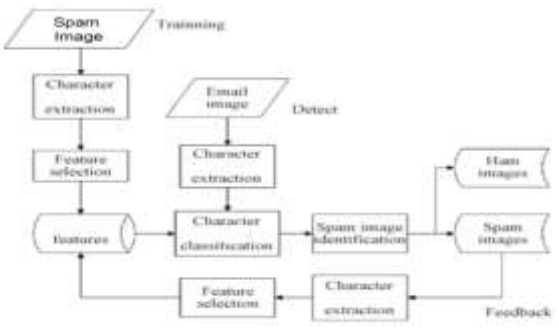
Also In this paper [09], author introduce a method which is combination of fingerprint technique and big data processing to detect the spam emails. Support Vector Machine (SVM) is the machine learning technique that is used for spam filtering. SVM training is a very large process so to deal with this Map Reduce platform for spam filter training was used .In this paper[3] the author used a content based spam filtering. The email classification as spam or ham is based on the data that is present in the content of the mail. So the header section is ignored in case of content based spam filtering. This paper specifically includes the comparison between implementations of Fisher-Robinson Inverse Chi-Square Function, implementation of AdaBoost classifier and KNN classifier.

3. Proposed work/ Proposed Idea

This section describes in detail the main stages of this study, starting with the data collection and labeling of the dataset, followed by a brief comparison of the most common techniques used in related studies. The main purpose of the system is to not only protect the social network management system and database, but also protect OSN users from being exposed to malicious content that is spread over those social networks as many of social network users have received or been exposed to malicious content such as spam, scams, and drive-by downloads.

3.1 Pseudo-OCR for Image Spam Filtering

The image spam manufacturing technology makes image spams more similar to the harm ones, thus more difficult to identify directly from image features without any content information. What's more serious, for some advanced applications, the spam image filtering process actually requires more contextual information than a simple filtering result. Hence we believe, it is essential for a anti-spam system to obtain extent content information of current image which apparently could only be obtained through long-established OCR based methods. While, as the discussed disadvantages mentioned above, traditional OCR is not our best choice. So, the idea of pseudoOCR is proposed to avoid such defects while still be able to extract enough content information. Compared with long-established technology, our proposed pseudo-OCR exhibits the following improvements for Chinese image spam filtering. Firstly, pseudo-OCR has a more approachable requirement for character reader. Determining whether or not a given character feature belongs to spam image rather than recognizing it is sufficient. Secondly, pseudo-OCR can effectively process a much wide range of images, even the ones with complex background and human interferences which are usually difficult to handle for traditional OCR based methods. And last, for Chinese character recognition, the proposed pseudo-OCR generates template features from certain training images instead of a set of standard Chinese characters. The architecture of our proposed system is illustrated in Figure, and as we can see, it consists of the following three sub-systems.



It gives the system learning ability to keep a proper high performance for a long period. It is well known for anti-spam communities that spammers tends to modify their image spam templates over time, which would result in an inevitable degradation of performance for near-duplicate based methods. Although our proposed methods is not strictly near-duplicate based, it adopts the similar methodology for extracting template character features from some known spam images. To handle such foreseeable defect, feedback mechanism is introduced in our system. By using detected spam as an additional source of template characters, it is very much possible to replace the obsoleted template character features with new ones, therefore to sustain a better performance.

3.2 Key-Point Based Character Feature

To meet the requirements of pseudo-OCR, the Chinese character feature extracted should also be modified. Concerning only certain key-points of a character, we devised a novel character feature. Which probably fails to be used for traditional character recognition yet sufficient to reserve enough content information for pseudo-OCR. The core of extracting such feature is a two-phase procedure. During the first phase the key-points and their connectivity information are extracted and stored as adjacency matrix using a DFS based algorithm, then the actual feature is calculated from this adjacency matrix in the second phase. For identifying image spams using this feature, every character feature extracted from a given image is compared with the template ones to determine its category information first, then the distribution of all these characters' category information is used for the final judgment.

3.3 Image spam filtering

From feature extraction described above, any input image will be converted into a set of 20-dimension key-point based character features. To use such features for image spam filtering, whose category information has to be obtained first. For a given character feature, the minimal L1 distance of those between it and all template features is calculated and compared with a certain threshold to determine its category. Here, this threshold is named category threshold to distinguish which with the following predefined threshold. Given all the category information of character features of an image, the distribution of such is used to make the final judgement. Because all the template character features in our implemented system fall into two categories, spam or ham. Then, by comparing the spam feature proportion with a predefine threshold calculated during the training process to choose the minimal spam image feature ratio of all training spam images, we are able to determine whether or not it is a spam image. In our system, a 0.25 minimal threshold is picked out of total 82 training spam images. Experiment results show that our proposed Chinese image spam filtering system using pseudo-OCR usually achieves a better performance when compared with traditional OCR based method.

3.4 Detecting Spam URL using SVM algorithm

An identity theft that occurs when a malicious web site masquerades a legitimate one is called Phishing. Such a theft occurs in order to procure sensitive information such as passwords, bank account details, or credit card numbers. Phishing makes use of spoofed emails which look exactly like an authentic email. These emails are send to a bulk of users and appear to be coming from legitimate sources like banks, e-commerce sites, payment gateways etc. Machine learning is a subset of artificial intelligence in the field of computer science that often uses statistical techniques to give computers the ability to learn. Therefore it turns out to be most critical to build up a quick and exact phishing recognition tool. Statistics about phishing activity and phishtank usage The GUI of our phishing detection system engages end users and provides them an environment for detecting malicious sites The whole discussion is to provide a user-friendly and effective, efficient way to prevent the internet users from phishing attacks and protect them from malicious sites.

We propose a url based phishing detection system using lexical features ,site popularity features and host based features by using the algorithms like ANN ,K Nearest Neighbors Classifier, Support vector machine (SVM) classifier ,Logistic Regression, Decision Tree , Bagging classifier , Random forest, Gradient Boosting Classifier.

We have implemented a Malicious URL based phishing detection system for end user where the GUI of our system engages end users and provide user friendly experience. The System Analyzes Uniform Resource Locator (URL) itself without accessing of Web sites. Also there exists no runtime idleness.

For our future work we aim to develop a browser plug-in which can work online. Besides, we aim to incorporate different parts of web based learning and assembling information to see the new patterns in phishing exercises.

Proposed Algorithm x ObURL Detection Algorithm

1. Input: Content

2. Output: Prevent the user if URLs seems defraud.

3. Caution User: Possible Phishing

4. Safe User: No Phishing

5. DB: Database

6. If Input from Content then

Alert User;

End For each iframe //get the content of frame

For each frame

Source do

If input form found then

Caution User;

End For each hyperlink in content's iframe source

Do

// perform the test 1 to 6

End For each hyperlink found in content

and iframe source URL

Do

7. Test 1: DNS Test

If hypertext! = Anchor text

Then

Caution User;

8. Test 2: IP Address Test

If IP address found in hyperlink

Then

If IP address found in White list DB then

Safe User;

Else Alert User;

// IP Address found in blacklist DB

9. Test 3: Encoded Test

If hyperlink found encoded

Then

Decode hyperlink;

Inform User; 10.Test 4: Shorten URL Test

If URL is shorten

Then

Alert User;

11. Test 5:hyperlink white list and blacklist test

If URL found in whitelist DB

Then

Safe User;

Else

Alert User;

// URL Found in Blacklist DB

12. Test 6: Pattern Matching Test

If hypertext and anchor text pattern is matching

then

Alert User.

4.Conclusion

In order to extract enough content information and avoid the defects of traditional OCR based methods, we propose the idea of pseudo-OCR, which reserves the structure of traditional OCR yet with a looser recognition requirement, the ability to process a wider range of input images and a more data oriented template feature generating mechanism. Furthermore, a pseudo-OCR based Chinese image spam filtering system with automatic learning ability is proposed in this paper. In the implementation part, we also create a novel Chinese key-points based character feature suitable for pseudo-OCR. Which, we believe, could also been used in applications like image spam clustering. By measuring comprehensive performance of proposed system concerning precision, recall and false positive rate, analysis of influence caused by two important perimeters is also conducted. Not only is optimization of our implemented system accessible, but also such analysis could provide instructive information if application circumstance changes. And by comparing with a traditional OCR based image spam filtering method, the experiment shows that our proposed method obtains a much better performance than the traditional OCR based methods and has the potential for practical use.

We also propose a url based phishing detection system using lexical features ,site popularity features and host based features by using the algorithms like ANN ,K Nearest Neighbors Classifier, Support vector machine (SVM) classifier ,Logistic Regression, Decision Tree , Bagging classifier , Random forest, Gradient Boosting Classifier.

We will also implement a Malicious URL based phishing detection system for end user where the GUI of our system engages end users and provide user friendly experience. The System Analyzes Uniform Resource Locator (URL) itself without accessing of Web sites. Also there exists no runtime idleness. For our future work we

aim to develop a browser plug-in which can work online. Besides, we aim to incorporate different parts of web based learning and assembling information to see the new patterns in phishing exercises.

5. REFERENCES

[1] A novel approach to protect against phishing attacks at client side using auto-updated white-list (IEEE 2016)

[2] Adulghani Ali Ahmed, N. A. A.: 2016, Real time detection of phishing websites, IEEE

[3] A.Bavani, D.Aarthi2, V. C.: 2017, Detecting phishing websites on real time using anti-phishing framework, Department of Information Technology (UG) 1, 2, 3, Assistant Professor4 Kingston Engineering College, India

[4] P. Burnap, A. Javed, O. F. Rana, and M. S. Awan, "Real-time classification of malicious URLs on Twitter using machine activity data," in Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2015, 2015, pp. 970–977.

[5] "A lightweight anti-phishing scheme for mobile phones" Lonfei WC IEEE, 2016

[6] Godwin Caruana, Maozhen Li, Yang Liu, An ontology enhanced parallel SVM for scalable spamlter training, Neuro computing Elsevier, vol. 108, pp.45-57, 2013.

[7] L. Zhang, J. Zhu, T. Yao, An evaluation of statistical spamltering techniques , ACM Transaction on Asian Language Information Process, vol. 3, pp.243269,2004 [5] Afroz, S. and Greenstadt, R.: 2015, Detecting phishing websites by looking at them, IEEE Communications Society .

[8] Z. Wang, W. Josephson, Q. Lv, M. Charikar and K. Li, "Filtering image spam with near-duplicate detection", Proceedings of the Fourth Conference on Email and Anti-Spam, Mountain View, California, USA, 2007.

[9] Puch-Tran Ho , HEE Su Kin, Application of Sim Hash Algorithm and Big Data Analysis in Spam Email Detection System, International Journal of Computer Applications (0975 8887) Volume 39 No.6, February 2014.C.

[10] G. Fumera, I. Pillai and F. Roli, "Spam filtering based on the analysis of text information embedded into images", The Journal of Machine Learning Research, Vol.7, No.6, pp.2699–2720, 2006.

[11] G. Fumera, I. Pillai, F. Roli, et al., "Image spam filtering using textual and visual information", Proceedings of MIT Spam Conference, Boston, MA, USA, 2007.

[12] H.B. Aradhye, G.K. Myers and J.A. Herson, "Image analysis for efficient categorization of image-based spam Email", Proceedings of the Eighth International Conference on Document Analysis and Recognition, Seoul, Korea, pp.914–918, 2005.

[13] N.C. Woods, O.B. Longe and A.B.C. Roberts, "A sobel edge detection algorithm based system for analyzing and classifying image based spam", Journal of Emerging Trends in Computing and Information Sciences, Vol.3, No.4 pp.506–512, 2012.

[14] C.T. Wu, et al., "Using visual features for anti-spam filtering", Proceedings of the IEEE International Conference on Image Processing, Genoa, Italy, pp.501–504, 2005

[15] M. Dredze, R. Gevaryahu and A. Elias, "Learning fast classifiers for image spam", Proceedings of the Fourth Conference on Email and Anti-Spam, Mountain View, California, USA, pp.487–493, 2007.

[16] C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou, and G. Min, "Statistical Features-Based Real-Time Detection of Drifted Twitter Spam," IEEE Trans. Inf. Forensics Secur., vol. 12, no. 4, pp. 914–925, 2017. [17] G. Canfora and C. A. Visaggio, "A set of features to detect web security threats," J. Comput. Virol. Hacking Tech., pp. 1–19, 2016.

[18] S. Liu, Y. Wang, C. Chen, and Y. Xiang, "An Ensemble Learning Approach for Addressing the Class Imbalance Problem in Twitter Spam Detection," in Information Security and Privacy: 21st Australasian Conference, ACISP 2016, Melbourne, VIC, Australia, July 4-6, 2016, Proceedings, Part I, J. K. Liu and R. Steinfeld, Eds. Cham: Springer International Publishing, 2016, pp. 215–228.

[19] S. K. Trivedi and S. Dey, "Effect of feature selection methods on machine learning classifiers for detecting email spams," Proc. 2013 Res. Adapt. Converg. Syst. - RACS '13, no. August 2016, pp. 35–40, 2013.

[20] A. Bollinger et al., "Adversarial machine learning," in Proceedings of the 4th ACM workshop on Security and artificial intelligence, 2011, vol. 92, no. 1, pp. 43–58.

[21] C. Chen, J. Zhang, X. Chen, Y. Xiang, and W. Zhou, "6 million spam tweets: A large ground truth for timely Twitter spam detection," IEEE Int. Conf. Commun., vol. 2015–Septe, pp. 7065–7070, 2015.

[22] H. Shen and X. Liu, "Detecting Spammers on Twitter Based on Content and Social Interaction," Proc. - 2015 Int. Conf. Netw. Inf. Syst. Comput. ICNISC 2015, pp. 413–417, 2015.