International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 01 | Jan-2018 www.irjet.net p-ISSN: 2395-0072

A Novel Technic to Notice Spam Reviews On e-Shopping

T.Tejaswi*[1], M. Ganesh[2], Sk. Naseem[3], Gandharba Swain[4]

[1,2,3,4] *Department of computer science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India –522502*

Abstract - The most common mode for consumers to express their level of satisfaction with their purchases is through online ratings, which we can refer as Online Review System. Network analysis has recently gained a lot of attention because of the arrival and the increasing attractiveness of social sites, such as blogs, social networking applications, micro blogging, or customer review sites. The reviews are used by potential customers to find opinions of existing users before purchasing the products. Online review systems plays an important part in affecting consumers' actions and decision making, and therefore attracting many spammers to insert fake feedback or reviews in order to manipulate review content and ratings. Malicious users exploit the review website and post untrustworthy, low quality, or sometimes fake opinions, which are referred as Spam Reviews. In this study, we aim at providing an efficient method to identify spam reviews and to filter out the spam content with the dataset.

Keywords- spam; dataset; bigram; unigram; heterogeneous

I. INTRODUCTION

ONLINE Social Media portals play an influential role in information propagation which is considered as an important source for producers in their advertising campaigns as well as for customers in selecting products and services. In the past years, people rely a lot on the written reviews in their decision-making processes, and positive/negative reviews encouraging/discouraging them in their selection of products and services. In addition, written reviews also help service providers to enhance the quality of their products and services. These reviews thus have become an important factor in success of a business while positive reviews can bring benefits for a company, negative reviews can potentially impact credibility and cause economic losses. 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. The reviews written to change users' perception of how good a product or a service are considered as spam, and are often written in exchange for money.

Despite this great deal of efforts, many aspects have been missed or remained unsolved. One of them is a classifier that can determine feature weights that show each feature's level of importance in determining spam reviews. The general concept of our proposed framework is to model a given review dataset as a Heterogeneous Information Network (HIN) [19] and to map the problem of spam detection into a HIN classification problem. In particular, we model review dataset as a HIN in which reviews are connected through different node types (such as features and users). A weighting algorithm is then employed to calculate each feature's importance (or weight). These weights are used to calculate the final labels for reviews using both unsupervised and supervised approaches.

To evaluate the proposed solution, we used two sample review datasets from Yelp and Amazon websites. Based on our observations, defining two views for features (review-user and behavioural-linguistic), the classified features as review-behavioural have more weights and yield better performance on spotting spam reviews in both semi-supervised and unsupervised approaches. In addition, we demonstrate that using different supervisions such as 1%, 2.5% and 5% or using an unsupervised approach, make no noticeable variation on the performance of our approach. We observed that feature weights can be added or removed for labelling and hence time complexity can be scaled for a specific level of accuracy. As the result of this weighting step, we can use fewer features with more weights to obtain better accuracy with less time complexity. In addition, categorizing features in four major categories (review-behavioural, user-behavioural, review-linguistic, user-linguistic), helps us to understand how much each category of features is contributed to spam detection. In summary, our main contributions are as follows:

(i) We propose Net Spam framework that is a novel network-based approach which models review networks as heterogeneous information networks. The classification step uses different meta path types which are innovative in the spam detection domain.

(ii) A new weighting method for spam features is proposed to determine the relative importance of each feature and shows how effective each of features are in identifying spams from normal reviews. Previous works [12], [20] 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). As we explain in our unsupervised approach, Net Spam is able to find features importance even without ground truth, and only by relying on meta path definition and based on values calculated for each review.

(iii) Net Spam improves the accuracy compared to the state-of-the art in terms of time complexity, which highly depends to the number of features used to identify a spam review; hence, using features with more weights will resulted in detecting fake reviews easier with less time complexity.

II. Related Work

In the past ten years, email spam detection and filtering mechanisms have been widely implemented. The main work could be summarized into two categories: the content-based model and the identity-based model. In the first model, a series of machine learning approaches are implemented for content parsing according to the keywords and patterns that are spam potential. In the identity-based model, the most commonly used approach is that each user maintains a whitelist and a blacklist of email addresses that should and should not be blocked by anti-spam mechanism [5,6]. More recent work is to leverage social network into email spam identification according to the Bayesian probability [7]. The concept is to use social relationship between sender and receiver to decide closeness and trust value, and then increase or decrease Bayesian probability according to these values.

With the rapid development of social networks, social spam has attracted a lot of attention from both industry and academia. In industry, Facebook proposes an Edge Rank algorithm [8] that assigns each post with a score generated from a few feature (e.g., number of likes, number of comments, number of reposts, etc.). Therefore, the higher Edge Rank scores, the less possibility to be a spammer. The disadvantage of this approach is that spammers could join their networks and continuously like and comment each other in order to achieve a high Edge Rank score.

In academia, Yardi et al. [9] studies the behaviour of a small part of spammers in Twitter, and find that the behaviour of spammers is different from legitimate users in the field of posting tweets, followers, following friends and so on. Stringhini et al. [10] further investigates spammer feature via creating a number of honey-profiles in three large social network sites (Facebook, Twitter and Myspace) and identifies five common features (follow-to-follower, URL ratio, message similarity, message sent, friend number, etc.) potential for spammer detection. However, although both of two approaches introduce convincible framework for spammer detection, they lack of detailed approaches specification and prototype evaluation.

Wang [11] proposes a naïve Bayesian based spammer classification algorithm to distinguish suspicious behaviour from normal ones in Twitter, with the precision result (F-measure value) of 89%. Gao et al. [12] adopts a set of novel feature for effectively reconstructing spam messages into campaigns rather than examining them individually (with precision value over 80%). The disadvantage of these two approaches is that they are not precise enough.

III. Existing Method

Different techniques have been used by researchers to find out the spam profiles in various OSNs. We are focussing only on the work that has been done to identify spammers in Twitter as it is not only a social communication media but in fact is used to share and spread information related to trending topics in real time. Table 1 is showing the summary of the papers reviewed regarding the detection of spammers in Twitter.

Table 1. Outline of techniques used for the detection of spammers

Author	Metrics Used	Methodology Used	Dataset Used	Result
Alex Hai Wang[20]	Graph based and Content based	Compared Naive Bayesion , Neural Network ,SVM & Decision tree	Validated on 500 Twitter with 20 recent tweets	Naive Bayesion giving highest accuracy 93.5%
Lee et al.[15]	User Based	Compared Decorate, Simple Logistic, FT, Logi Boost ,RandomsubSpace,Bagging,j48,LibSVM	Validated on 1000 Twitter Users	Decorate giving highest accuracy 88.98%
Benevenoto et.al[21]	User based and Content Based	SVM	Validated on 1065 Twitter Users	Accuracy 87.6% with User Based & Content Based features and Accuracy 84.5(With only user based features
Gee et.al[22]	User Based	Compared Naive Bayesion ,SVM	Validated on 450 Twitter Users with 200 recent tweets	Accuracy 89.6%
McCord et.al[23]	User Based and Content Based	Compared Random Forest,SVM, Naive Bayesion ,KNN	Validated on 1000 Twitter Users with 100 recent tweets	Random forest giving highest accuracy 95.7%
Chakraborty et.al[24]	User Based and Content Based	Compared Random Forest ,SVM, Naive Bayesion ,Decision Tree	Trained on 5000 Twitter Users with 200 recent tweets	SVM giving highest accuracy-89%
X. Zheng et al[25]	User Based and Content Based	SVM	Validated on 30,000 weibo users	SVM giving highest accuracy-99%

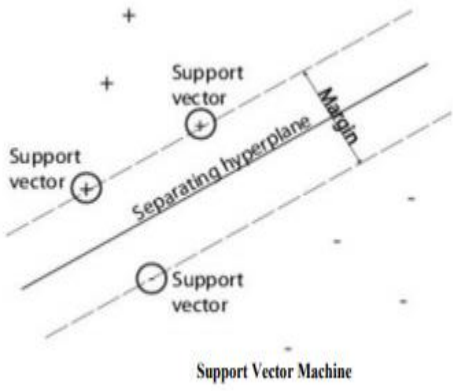
IV. Implementation and Methodology

Classification an extensive number of classification algorithm has been connected to spam recognition region, where support vector machine classification for its decent generalization performance effect Furthermore, exceptionally well known. SVM is an intense method utilized for data classification. Despite the fact that people consider that it is simpler to use than Neural Networks. Each example in the preparation set contains one class marks and a few components. The fundamental point of SVM is to create a model which predicts class labels of information occurrences in the testing set which are given only the features. At the show, the support vector machines have been broadly utilized as a part of content based hostile to spam system. SVM is a splendid solution for the little sample size issue, by developing an isolating hyperplane to finish the classification. As the support vector machine in spam identification in the great execution, the paper utilizes this algorithm to identify spam reviews.

1) Support Vector Machine A support vector machine (SVM) can be utilized when our information has totally two classes. An SVM classifies information by finding the ideal hyperplane that isolates all information purposes of one class from those of alternate class. The hyperplane for an

SVM implies the one with the biggest margin between the two classes. Margin suggests the maximal width of the segment parallel to the hyperplane that has no interior information points.

2) Properties of SVM Support Vector Machine has a place with a group of generalized linear classifiers and it can be deciphered as an extension of the perception. A unique ability is that they all the while limit the exact classification error and maximize the geometric margin; henceforth they are otherwise called maximum margin classifiers.



Support Vector Machine

In this section, we will discuss the proposed methodology for email spam detection technique.

A. Pre-processing

The pre-processing step is used to remove the noises from the email which are irrelevant and need not be present. The pre-processing step includes.

- Removal of Numbers
- Removal of Special Symbol
- Removal of URLS
- Stripping HTML
- Word Stemming

B. Feature Extraction

Feature Extraction is used to extract the important and relevant features from the email body. The feature transforms the email into 2 D vector space having features number. These features are mapped from the vocabulary list.

$$x = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^n$$

V. Conclusion

The Spam is a standout amongst the most irritating and malicious increments to worldwide PC world. In this paper, we propose a novel method for email spam detection which can effectively identify the spam emails from its contents. The spam emails can be blocked by the user and genuine review can be retained by the user. The proposed classifier achieves 98 % accuracy while classifying the series of datasets.'

REFERENCES

[1] J. Donfro. A Whopping 20% of Yelp Reviews are Fake, accessed on Jul. 30, 2015. [Online]. Available: http://www.businessinsider.com/20-percent-of-yelp-reviews-fake-2013-9

[2] M. Ott, C. Cardie, and J. T. Hancock, "Estimating the prevalence of deception in online review communities," inProc. ACM WWW, 2012, pp. 201–210.

[3] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," inProc. ACL, 2011, pp. 309–319.

[4] C. Xu and J. Zhang, "Combating product review spam campaigns via multiple heterogeneous pairwise features," inProc. SIAM Int. Conf. Data Mining, 2014, pp. 172–180.

[5] N. Jindal and B. Liu, "Opinion spam and analysis," in Proc. WSDM, 2008, pp. 219–230.

[6] F. H. Li, M. Huang, Y. Yang, and X. Zhu, "Learning to identify review spam," inProc. 22nd Int. Joint Conf. Artif. Intell. (IJCAI), 2011, pp. 1–6.

[7] G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, "Exploiting burstiness in reviews for review spammer detection," in Proc. ICWSM, 2013, pp. 1–10.

[8] A. J. Minnich, N. Chavoshi, A. Mueen, S. Luan, and M. Faloutsos, "Trueview: Harnessing the power of multiple review sites," in Proc. ACM WWW, 2015, pp. 787–797.

[9] B. Viswanath et al., "Towards detecting anomalous user behavior in online social networks," inProc. USENIX, 2014, pp. 1–16.

[10] H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao, "Spotting fake reviews via collective positive-unlabeled learning," inProc. ICDM, Dec. 2014, pp. 899–904.

[11] L. Akoglu, R. Chandy, and C. Faloutsos, "Opinion fraud detection in online reviews by network effects," inProc. ICWSM, 2013, pp. 1–10.

[12] S. Rayana and L. Akoglu, "Collective opinion spam detection: Bridging review networks and metadata," in Proc. ACM KDD, 2015, pp. 1–10.

[13] S. Feng, R. Banerjee, and Y. Choi, "Syntactic stylometry for decep-tion detection," in Proc. 50th Annu. Meeting Assoc. Comput. Linguis-tics (ACL), 2012, pp. 1–5.

[14] N. Jindal, B. Liu, and E.-P. Lim, "Finding unusual review patterns using unexpected rules," inProc. ACM CIKM, 2012, pp. 1–4.

[15] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers using rating behaviors," inProc. ACM CIKM, 2010, pp. 1–10.

[16] A. Mukherjee et al., "Spotting opinion spammers using behavioural footprints," in Proc. ACM KDD, 2013, pp. 1–9.

[17] S. Xie, G. Wang, S. Lin, and P. S. Yu, "Review spam detection via temporal pattern discovery," in Proc. ACM KDD, 2012, pp. 823–831.