International Journal of Trend in Scientific Research and Development (IJTSRD)

International Open Access Journal

ISSN No: 2456 - 6470 | www.ijtsrd.com | Volume - 2 | Issue – 3

Netspam: An Efficient Approach to Prevent Spam Messages using Support Vector Machine

P. Sai Kiran, K. Prudhvi Chowdary, T. T. Venkata Rayudu, K. Vinay Kumar

Department of Computer Science, Dhanekula Institute of Engineering &Technology, Vijayawada, Andhra Pradesh, India

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 increasing attractiveness of social sites, such as blogs, social networks, 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 play an important part in affecting consumers' actions and decision making, and therefore attracting many spammers to insert fake feedback or reviews to manipulate review content and ratings. Malicious users misuse the review website and post untrustworthy, low quality, or sometimes fake opinions, which are referred as Spam Reviews. In this study, we aim at classifying reviews as positive, negative and spam reviews by creating a social network similar platform and providing communication between users in it.

***Keywords**: NETSPAM (Network Spam); SVM (Support Vector Machine); HIN (Heterogeneous Information Network); OSN (Online Social Network); sending product posts*

INTRODUCTION:

Online Social Media portals play an influentialrole in information propagation. So, this is considered as an important source for producers in their advertising campaigns as well as for customers in selecting products and services. In addition, written reviews also help service providers to enhance the quality of their products and services. These reviews thus became 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.

As shown in, 20% of the reviews in the Yelp website are actually spam reviews. On the other hand, a considerable amount of literature has been published on the techniques used to identify spam and spammers as well as different type of analysis on this topic. These techniques can be classified into different categories; some using linguistic patterns in text which are mostly based on bigram, and unigram, others are based on behavioral patterns that rely on features extracted from patterns in users' behavior which are mostly metadata based and even some techniques using graphs and graph-based algorithms and classifiers.

Despite this great deal of efforts, many aspects have been missed or remained unsolved. One of them is a classifier that can calculate 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) 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 utilized 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 behavioral-linguistic), the classified features as review behavioral 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 labeling 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-behavioral, user-behavioral, 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 IEEE Transactions on Information Forensics and Security, Volume:12, Issue:7, Issue Date: July.2017 2 different metapath 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 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 can find features importance even without ground truth, and only by relying on metapath 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.

SYSTEM DESIGN:

In this section we present the design of our proposed system which detects abnormal spam messages using support vector machine algorithm.

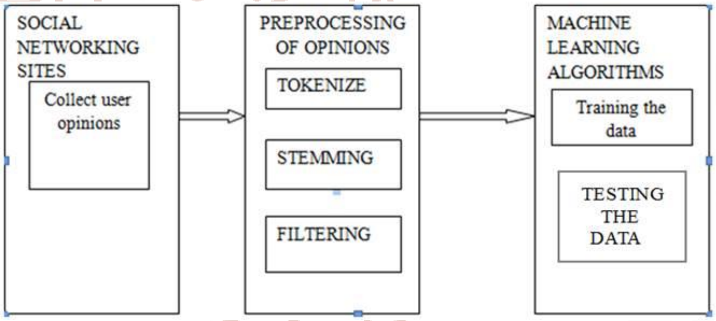


Fig 2: System architecture

We divide the architecture of our system into three phases which are offline phase which includes collecting user opinions, preprocessing of opinions, machine algorithms.

In the user opinions part, we collect the user's opinion from various registered users and the received data can be stored for future scope. In the preprocessing step, we perform various levels of steps in order to frame the data according to the data for testing and from that we can retrieve the results. In the tokenize step we process the data based on the linguistic used by the user according to the parts of speech like using the adjective as a key for detection of spam words. In the second level that is stemming in which we frame the data according to the simplified sentence's that is filtering of unwanted words in the user's collected data. In the final step, we perform the filtering on the collected data so that the data can be processed for further stages. In the final stage we have the different machine algorithms, but we use the support vector machine algorithm for the classification and execution of the data collected and then the results is outputted to the users.

MODULES:

1. Data collection from users:

We collect the data from users through an Net Spam framework which uses support vector machine algorithm as their classifier model. We store the data in the database for processing to it further stages in the execution.

2. Classification of live user's and artificial data:

With the help of classifier, we can calculate 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) 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).

3. Filtering of data using weighting method:

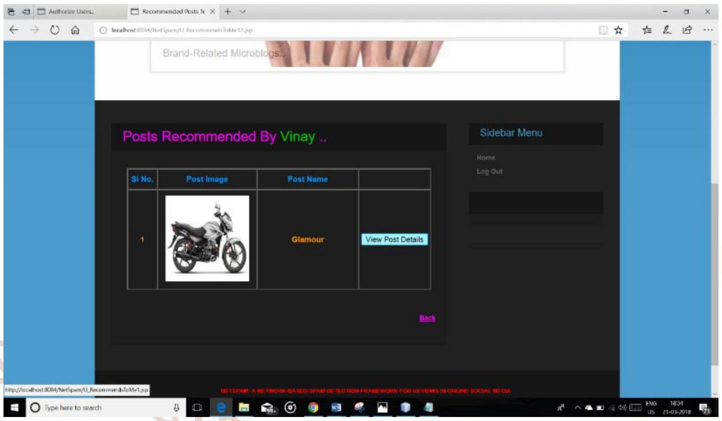
A new weighting method for spam features is proposed to determine the relative importance of each feature and shows how effective each of features are in identifying spams from normal reviews. Previous works also aimed to address the importance of features mainly in term of obtained accuracy, but not 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 can find features importance even without ground truth, and only by relying on metapath definition and based on values calculated for each review.

4. Displaying of resulted data to users into positive, negative and unclassified spam messages:

Finally, the user's data is processed and the results is outputted to the users into positive, negative and even unclassified spam messages to the users so that the users can able to take the decision before buying the online product in social media.

Detecting of spam messages using NetSpam framework:

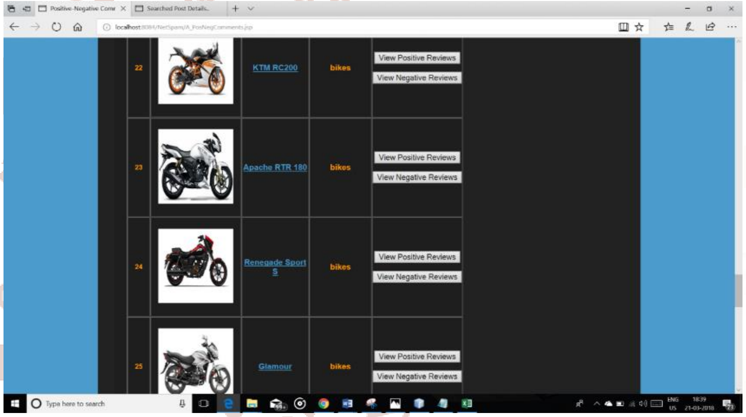
The figure shows the web interface of our application. We have different outputs in our application.


Authorize Users..
Recommended Posts To
localhost:8084/NetSpam/U_RecommendsToMe12.jsp
Brand-Related Microblogs..
Sidebar Menu
Home
Log Out
Back
http://localhost:8084/NetSpam/U_RecommendsToMe1.jsp
NETSPAM: A NETWORK-BASED SPAM DETECTION FRAMEWORK FOR REVIEWS IN ONLINE SOCIAL MEDIA
Type here to search
ENG US 18:34 21-03-2018

Posts Recommended By Vinay ..

Sl No.	Post Image	Post Name	
1		Glamour	View Post Details

Fig 1: User's are posting the reviews about the product.



22		KTM RC200	bikes	View Positive Reviews View Negative Reviews
23		Apache RTR 180	bikes	View Positive Reviews View Negative Reviews
24		Renegade Sport S	bikes	View Positive Reviews View Negative Reviews
25		Glamour	bikes	View Positive Reviews View Negative Reviews

Fig 2: User's are viewing the positive and negative reviews.

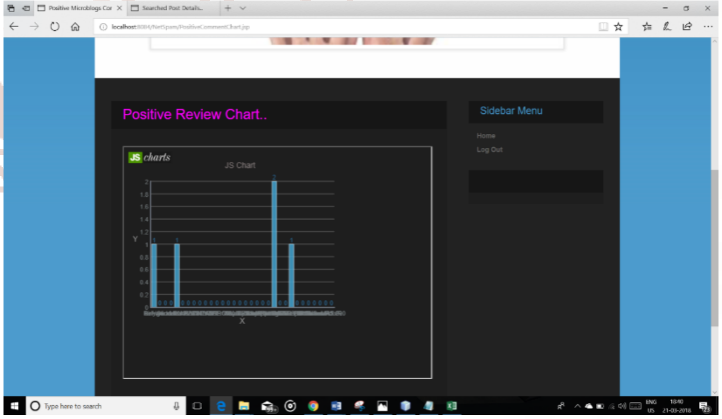


Fig 3: The positive and negative chart of user's data.

CONCLUSION

In this paper we developed aeb framework for detection of spam messages using the support vector machine algorithm as the classifier model for the user's data.

For future work, metapath concept can be applied to other problems in this field. For example, similar framework can be used to find spammer communities. For finding community, reviews can be connected through group spammer features (such as the proposed feature in) and reviews with highest similarity based on metapath concept are known as communities. In addition, utilizing the product features is an interesting future work on this study as we used features more related to spotting spammers and spam reviews. Moreover, while single networks have received considerable attention from various disciplines for over a decade, information diffusion and content sharing in multilayer networks is still a young research. Addressing the problem of spam detection in such networks can be considered as a new research line in this field.

REFERENCES:

1. J. Donfro, A whopping 20 % of yelp reviews are fake. http://www.businessinsider.com/20-percent-of-yelp-reviews-fake-2013-9. Accessed: 2015-07-30.
2. M. Ott, C. Cardie, and J. T. Hancock. Estimating the prevalence of deception in online review communities. In ACM WWW, 2012.
3. M. Ott, Y. Choi, C. Cardie, and J. T. Hancock. Finding deceptive opinion spam by any stretch of the imagination. In ACL, 2011.
4. Ch. Xu and J. Zhang. Combating product review spam campaigns via multiple heterogeneous pairwise features. In SIAM International Conference on Data Mining, 2014.
5. N. Jindal and B. Liu. Opinion spam and analysis. In WSDM, 2008.
6. F. Li, M. Huang, Y. Yang, and X. Zhu. Learning to identify review spam. Proceedings of the 22nd International Joint Conference on Artificial Intelligence; IJCAI, 2011.
7. G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh. Exploiting burstiness in reviews for review spammer detection. In ICWSM, 2013.
8. A. j. Minnich, N. Chavoshi, A. Mueen, S. Luan, and M. Faloutsos. Trueview: Harnessing the power of multiple review sites. In ACM WWW, 2015.
9. B. Viswanath, M. Ahmad Bashir, M. Crovella, S. Guah, K. P. Gummadi, B. Krishnamurthy, and A. Mislove. Towards detecting anomalous user behavior in online social networks. In USENIX, 2014.
10. H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao. Spotting fake reviews via collective PU learning. In ICDM, 2014.
11. L. Akoglu, R. Chandy, and C. Faloutsos. Opinion fraud detection in online reviews by network effects. In ICWSM, 2013.
12. R. Shebuti and L. Akoglu. Collective opinion spam detection: bridging review networks and metadata. In ACM KDD, 2015.
13. S. Feng, R. Banerjee and Y. Choi. Syntactic stylometry for deception detection. Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers; ACL, 2012.
14. N. Jindal, B. Liu, and E.-P. Lim. Finding unusual review patterns using unexpected rules. In ACM CIKM, 2012.
15. E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw. Detecting product review spammers using rating behaviors. In ACM CIKM, 2010.
16. A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh. Spotting opinion spammers using behavioral footprints. In ACM KDD, 2013.
17. S. Xie, G. Wang, S. Lin, and P. S. Yu. Review spam detection via temporal pattern discovery. In ACM KDD, 2012.
18. G. Wang, S. Xie, B. Liu, and P. S. Yu. Review graph based online store review spammer detection. IEEE ICDM, 2011.
19. Y. Sun and J. Han. Mining Heterogeneous Information Networks; Principles and Methodologies, In ICCCE, 2012.
20. A. Mukerjee, V. Venkataraman, B. Liu, and N. Glance. What Yelp Fake Review Filter Might Be Doing? In ICWSM, 2013.