ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

Netspam: A Network-Based Spam Detection

[1] Prathyusha.k, [2] Pallavi.k, [3] Suhas.R , [4] Nagaraj.M
[1][2][3][4] RYMEC

Abstract: **- At present days people are interacted on the information present on social media and their decisions. There is a chances of leaving a reviews on a social media whether is positive or negative by spammers on particular product, organization and their services .by identifying these spammers and spams inorder to know the reviews in the social media, we are introducing a novel framework called Netspam which utilizes spam features for modeling review datasets as heterogeneous information networks to map spam detection procedure into a classification problem in such networks.**

I. INTRODUCTION

Online social media places a major role for producers by providing a advertisements, campaigns etc for the customers due to selection of products whether it is good or bad it is based on the reviews written in online either in positive or negative depend on that the customer is going to select the product.by the help of positive reviews will automatically get success in business this leads to the development in economy ,some people identify a comments as a reviews this is a good oppurtinity for the spammers to produce a fake reviews in social media about the product and their services. The fact that anyone with anyidentity can leave comments as review, provides a tempting opportunity for spammers to write fake reviews designed to mislead users' opinion. This abnormal reviews are multiplied in sharing on a online social media.mainly the reviews are written for the customers how good the product or a services are considered as spam and are often written in exchange for money. Recently a few literatures are published to detect the spammers and spams and also helpful in making the analyses on this particular topic. In this we are using different techniques or algorithms to find the spammers and spam reviews like N-gram technique,comparision algorithm, splitting algorithm etc; and classifiers One of them is a classifier that can calculate feature weights that show each feature's level of importance in determining spam reviews. Previous work:As mentioned earlier, we model the problem as a heterogeneous network where nodes are either real components in a dataset (such as reviews, users and products) or spam features.Index Term:Social Media, Social Network, Spammer, Spam Review, Fake Review, Heterogeneous Information Networks.

II. PROPOSED SOLUTION:

- **Upload Excel File:**

In the Upload Excel File Module, user has to select the file from the client machine and the file content will be sent to the server via URL in the form of multipart ,in the server side servlet receives the file content and write the file content in the folder of the application. From that folder it reads the file content and store the file content in to the database.

- **Fake Review Detection 1:**

In the Fake Review Detection 1 process , Data will be read from the database and checks whether the IP_Address and UserID is fake or not based on the meta data table and insert the fake reviews in to the fake review table . And also it checks whether the number of reviews from the IP_Address are exceeding the threshold limit with in the threshold time limit, if any IP_Address exceeds the threshold limit,then that reviews will be inserted to the fake review table and that IP_Address will be inserted to the Meta Fake IP_Address table and rest of the reviews will be inserted to the Real reviews table.

- **Fake Review Detection 2:**

In the Fake Review Detection 2 process , reviews will be read from the Real reviews table, considering each reviews , in the first level ,unnecessary words and special characters will be removed, in the second level categorize each and every word is noun or adjective , in the third level paring the noun and adjacent adjective , in the fourth level checks whether the adjective which is paired with the noun is negative or positive , in the fifth level checks whether the maximum number of pairs are positive or negative , based on the maximum count of positive or negative, assign the review value as positive or negative ,in the sixth level calculate and insert the two gram and three gram pairs in to the database, in the seventh level calculate the count percentage, positive percentage and n-gram percentage of each user and add all the percentages and get total percentage threshold , if any user exceeds total percentage threshold, consider that user is fake and insert that user in to the meta fake user

table.

III. CONCLUSION

This study introduces a novel spam detection framework namely NetSpam based on a metapath concept.Our observations show that calculated weights by using this metapath concept can be very effective in identifying spam reviews and leads to a better performance. NetSpam can calculate the importance of each feature and it yields better performance in the features addition process, and performs better than previous works, with only a small number of features. Moreover,after defining four main categories for features our observations show that the reviews behavioral category performs better than other categories, in terms of AP, AUC as well as in the calculated weights. The results also confirm that using different supervisions, similar to the semi-supervised method, have no noticeable effect on determining most of the weighted features, just as in different datasets.

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 [19]) and reviews with highest similarity based on metapth 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 has received considerable attention from various disciplines for over a decade, information diffusion and content sharing in multilayer networks is still a young research [27]. Addressing the problem of spam detection in such networks can be considered as a new research line in this field.

REFERENCES

[1] L. Akoglu, R. Chandy, and C. Faloutsos. Opinion fraud detection in online reviews bynetwork effects. In ICWSM, 2013.

[2] R. Shebuti and L. Akoglu. Collective opinion spam detection: bridging review networksand metadata. In ACM KDD, 2015.

[3] 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.

[4] N. Jindal, B. Liu, and E.-P. Lim. Finding unusual review patterns using unexpected rules. In ACM CIKM, 2012.

[5] 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.

[6] 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.

[7] S. Xie, G. Wang, S. Lin, and P. S. Yu. Review spam detection via temporal pattern discovery. In ACM KDD, 2012.

[8] G. Wang, S. Xie, B. Liu, and P. S. Yu. Review graph based online store review spammer detection. IEEE ICDM, 2011.

[9] Y. Sun and J. Han. Mining Heterogeneous Information Networks; Principles and Methodologies, In ICCCE, 2012.

[10] A. Mukerjee, V. Venkataraman, B. Liu, and N. Glance. What Yelp Fake Review Filter Might Be Doing?, In ICWSM, 2013.

[11] S. Feng, L. Xing, A. Gogar, and Y. Choi. Distributional footprints of deceptive product reviews. In ICWSM, 2012.

[12] Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. In VLDB, 2011.

[13] Y. Sun and J. Han. Rankclus: integrating clustering with ranking for heterogeneous information network analysis. In Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology, 2009.

[14] C. Luo, R. Guan, Z. Wang, and C. Lin. HetPathMine: A Novel Transductive Classification Algorithm on Heterogeneous Information Networks. In ECIR, 2014. [15] R. Hassanzadeh. Anomaly Detection in Online Social Networks: Using Datamining Techniques and Fuzzy Logic. Queensland University of Technology, Nov. 2014.

[16] M. Luca and G. Zervas. Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud., SSRN Electronic Journal, 2016.

[17] E. D. Wahyuni and A. Djunaidy. Fake Review Detection From a Product Review Using Modified Method of Iterative Computation Framework. In Proceeding MATEC Web of Conferences. 2016.

[18] M. Crawford, T. M. Khoshgoftaar, and J. D. Prusa. Reducing Feature set Explosion to Faciliate Real-World Review Sapm Detection. In Proceeding of 29th International Florida Artificial Intelligence Research Society Conference. 2016.

[19] A. Mukherjee, B. Liu, and N. Glance. Spotting Fake Reviewer Groups in Consumer Reviews. In ACM WWW, 2012.

[20] A. Heydari, M. A. Tavakoli, N. Salim, and Z. Heydari. Detection of review spam: A survey. Expert Systems with Applicants, Elsevier, 2014.

[21] M. Crawford, T. D. Khoshgoftar, J. N. Prusa, A. Al. Ritcher, and H. Najada. Survey of Review Spam Detection Using Machine Learning Techniques. Journal of Big Data. 2015.

[22] H. Xue, F. Li, H. Seo, and R. Pluretti. Trust-Aware Review Spam Detection. IEEE Trustcom/ISPA . 2015.

[23] C. L. Lai, K. Q. Xu, R. Lau, Y. Li, and L. Jing. Toward a Language Modeling Approach for Consumer Review Spam Detection. In Proceedings of the 7th international conference on e-Business Engineering. 2011.

[24] N. Jindal and B. Liu. Opinion Spam and Analysis. In WSDM, 2008.

[25] S. Mukherjee, S. Dutta, and G. Weikum. Credible Review Detection with Limited Information using Consistency Features, In book: Machine Learning and Knowledge Discovery in Databases, 2016.

[26] K. Weise. A Lie Detector Test for Online Reviewers. http://bloom.bg/1KAxzhK. Accessed: 2016-12-16.

[27] M. Salehi, R. Sharma, M. Marzolla, M. Magnani, P. Siyari, and D. Montesi. Spreading processes in multilayer networks. In IEEE Transactions on Network Science and Engineering. 2(2):65–83, 2015