International Journal on Applications in Engineering and Technology
Volume 9 : Issue 2 : May 2023, pp 65 – 68 www.aetsjournal.com ISSN (Online) : 2455 - 0523

Identification of an Account as Spammer or Non-Spammer with Accuracy

Dr K.Nagarajan, Mr. V. Vinod Kumar , Mr. M. Mohammedkasim , Mr. T Prabu

Abstract— Online reviews have a huge impact on the customer's choice for purchase. These reviews are being read and affects a lot of user decision whether to buy a product or not. There are a lot of reviews based on a product majority of which turns out to be fake written by people to defame a product, which is often created by the competitors of a particular product. The present research concentrates on spam detection and categorizing them from these reviews. We have found that the studies can be classified into three groups that focus on methods to detect fake reviews, individual spammers and group fraudsters. Different techniques have been used for the detection of fake reviews, and block them. In this study a novel framework to improve Netspam method is proposed. The reviews are analysed by the admin and if found spam, the user is blocked from access his account. The results shows that it is more efficient in preventing people from creating fake reviews, whereas existing methods defines which is a fake review and which is not.

Keyword- Fake reviews, Reviewers, E-Commerce, Products, and Items.

I. INTRODUCTION

Twenty first Century is the era of technology, where even a baby is born directly towards the face of a camera in Skype call or IMO. Technology and devices have made a huge space for itself in our daily life. With the invention of applications and websites evolving day by day people find it easier to access devices to get information rather than moving out. There the e-commerce sector has grown tremendously to our lives.

Online shopping sites are increasing day by day with advent of new materials made available online. Every sites and applications are adding new features to it, tempting the consumers to buy their products. Reviews hold an important part in a customer's decision to buy a product, if the reviews are found to be positive, chances that a user purchases it is high but in other case if the review turns out to be negative it has reverse impact on the users choice.

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. 20% of the reviews in the Yelp website are actually spam reviews.

The widespread impact of customer reviews has raised spam (review) issues on the websites containing customer reviews. As anyone can post anything on the review box, people are carefree to type anything in the review box irrespective of the product. Some reviews are based on the usage of the product, some are based on the features of the product, some can be even based on the comparison of the products, and some may be based on nothing related to them.

Researchers have developed various detection techniques for fake reviews in past years to preserve the accuracy of opinion-mining results, real customers and true vendors. The important task is distinguishing between fake reviews and real reviews.To identify literature concerning "fake review",background study on related journals and conference papers were done. The terms used for searching the related works were spam detection, fake reviews, and opinion mining, fraudulent reviews. The papers were selected on the basis of its relation with the base paper related to Net Spam.

Online social media entries are not filtered, anyone can post anything in fraction of seconds it spreads widely until or unless a report is filed against it, it is not verified. The amounts of abusive comments are also to be considered and the user needs to be blocked to type certain words in their typing window. The websites like shtyle have a feature which prevents the user from typing certain dataset so that even if the user types the word it either is seen in the screen as * or dot, it doesn't get printed in the same manner as the user have typed.

The websites like Amazon provides reliability on a comment by providing "Amazon Verified Purchase" tag and Yelp.com removes the suspicious review by using heuristic rules. There are still ways how a fake reviewer can escape the eyes of detection system. Some of the reviewers are trained in writing fake reviews which go undetected; to get them to the spotlight an alternative method used is the socializing behaviour of the user. The websites allow users to put trusted tag on a particular user based on the review, helpfulness tags etc. are also used by websites. Some users tend to put reviews and not socialize much and some may be socially inactive and still post reviews.

The lazy users are few who comment and socialize not frequently but once in a while. The activity logs of a user is

Dr K. Nagarajan, Assistant Professor, Department of Electronics And Communication Engineering, Nehru institute of Engineering and Technology , Coimbatore , India . (Email Id : naguambani@gmail.com)

Mr.Vinod Kumar V, Assistant Professor, Department of Electronics And Communication Engineering, Nehru institute of Engineering and Technology , Coimbatore , India . (Email Id: vinodvijayan0289@gmail.com)

Mr. Mohammedkasim M, Assistant Professor, Department of Electronics And Communication Engineering, Nehru institute of Engineering and Technology , Coimbatore , India . (Email Id : mohammedkasim1983@gmail.com)

Mr. T Prabu, Assistant Professor (SG) , Department of Electronics And Communication Engineering, Nehru institute of Engineering and Technology , Coimbatore , India . (Email Id : tprabu19@gmail.com)

useful is analysing the social behaviour of the user and ensure if he or she can be trusted. The reviewers who write genuine reviews may even be socially interactive with the fake reviewers. This makes it difficult to analyse the social nature of genuine and deceptive users.

II. PROBLEM DEFINITION

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 metapath types which are innovative in the spam detection domain.

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 termof 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 metapath definition and based on values calculated for each review.

Net Spam improves the accuracy compared to the stateof-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.

III. .EXISTING SYSTEM

In Existing work, it only depend on the detecting the spam reviews and spammers. None of them show the importance of each extracted feature type. 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 behavioural patterns that rely on features extracted from patterns in users' behaviour which are mostly metadata based.These work not enough to classify the spam network.

- Lack of work to detect spam features.
- Negative reviews can potentially impact credibility and cause economic losses.
- Spammers to write fake reviews designed to mislead users' opinion.
- It requires more execution time for identify spam in Twitter Data and that methods provide the less Accuracy.

IV. PROPOSED SYSTEM

Spamming is an undesirable activity in OSNs and effective mechanisms need to be developed to detect it and thereafter take remedial steps. In this Net Spam framework, we have applied two algorithms namely k-NN, Conflict detection algorithms. In k- Nearest Neighbour approach, user accounts were classified by calculating the probability of the given account to be Spammer/Non-Spammer, given the feature values of that account. Bayes theorem was used to calculate this probability. On the basis of similar feature values, Conflict detection could classify the entire set of accounts into two classes (Spammers and Non-Spammers) user based and content based.

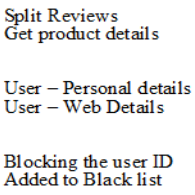

Fig 1: Process Diagram

To classify the dataset with highest accuracy combine the content, user based and behaviour based features. Our algorithm was able to maintain the high accuracy of conflict detection algorithm in detecting non-spam and at the same time, retain the accuracy of Naive Bayes in detecting Spammers accounts thereby, increasing the overall accuracy.

A. Data Classification

The user uploads the comments regarding an item that can be of anything a dress, phone, hotel, or places. The comments that user type will be positive or negative. The reviewer may know exactly about the item or may be a random comment. The review words are collected and the stop words are removed. The stop words are those words that are more like pronouns for example like, this that, who, me, you etc. These words are removed and the remaining words are analysed. The comment or review is then classified into relevant or fake review using k- Nearest Neighbour method. The kNN method determines the link between the data updated and the previously available data in the database. It checks

The dataset includes the reviewers' impressions and comments about the quality, and other aspects related to products, restaurants (or hotels). The dataset also contains labelled reviews as ground truth (so-called near ground.-truth), which indicates whether a review is spam or not.The process of this module is containing that field of processing in data processing, which is after file uploaded in these method for that data incurred of the network.The data set was pre-processed by extracting spam features from the features data in the data set. Other attributes in the dataset are rate of reviewers, the date of the written review, and date of actual visit, as well as the user's and the product id (name).

B. Metapath Identification

Metapath is defined by a sequence of relations in the network schema. For metapath creation, we define an extended version of the metapath concept considering different levels of spam certainty. In particular, two reviews are connected to each other if they share same value.

In particular, given a review, the levels of spam certainty for metapath (i.e., feature) is calculated as the number of levels. we need enough spam and non-spam reviews for each step, with fewer number of reviews connected to each other for every step, the spam probability of reviews take uniform distribution, but with lower value of s we have enough reviews to calculate final spamicity for each review.

C. Spam labelling and User Blocking

When we sort spam probabilities for reviews, all of the reviews with spam labels are located on top of the list and ranked as the first reviews. It is worth to note that in creating the HIN, as much as the number of links between a review and other reviews increase, its probability to have a label similar to them increase too, because it assumes that a node relation to other nodes show their similarity. In particular, more links between a node and other non-spam reviews, more probability for a review to be non-spam and vice versa. In other words, if a review has lots of links with non-spam reviews, it means that it shares features with other reviews with low spamicity and hence its probability to be a non-spam review increases.

The review after updating is found to be fake, then the time user details are considered. The user log activities are accessed and assessed to verify the genuine nature of his logs. The user profile activity is viewed such as the time or duration of his log session and the time taken to post the content after logging in to the account is also noted. If the time taken to type comment was short than normal time taken to type and post the comment, then it is sure that the person have copied and pasted the comment. The time which the user have taken to update the comment is available in the admin database. Once the data upon being verified and spam then the user account is blocked.

V.IMPLEMENTATION

The user registers with the system providing the credentials like name, email, password, mobile number, aadhar number. The user after registration can login to the system using the username and password. The system here does not use a training dataset. The data uploaded by the user is taken as dataset. The system is designed in a way that once the user uploads a comment it is extracted and classified using kNN algorithm. The time of the user login and data upload is recorded in the database to which the admin has access. The comment uploaded if then classified as spam and the report is generated for the admin to check and confirm.

The k-NN classifier finds out the nearest distance i.e. the similarity between the uploaded data and the previously existing data using cosine similarity of the contents, the user time frame comparison is done using the Euclidean distance calculation.

The comment if found to be spam then the system doesn't allow the user to post it. There is a file in database which has predefined set of words considered as spam. There is another file which states the stop words in general and the uploaded comment is compared with the stopwords. These stopwords are removed from the comment and the remaining set of phrase is forwarded to verify its credibility. There is a specific set of depressed sentences which are genuinely considered inappropriate in the reviews.

The reviews after removing stopwords is then compared with the depressed sentences to ensure the credibility, once there is a similarity between these sentences, then a report is generated in the admin database regarding the same. The next type of report is prepared upon the duplication of sentences, i.e. if there is a comment similar to the one recently updated by different users then the comment is considered to be spam.

The final report is when the user upload comments about the same brand more than once. The comment and the posted time and user name is recorded in the database and the administrator is responsible to verify the credibility of the result.

The user log is also recorded in the database and the in time and out time is recorded to ensure the usage of the reviewer. This record ensure if the reviewer is a frequent user or just an alias for the reviews in he record.

The reviews that occur in all the three reports is more likely to be spam. The report is viewed by the administrator and the time frame is verified. If the time frame is short and the review is large which doesn't match the minimum amount of time required to type the sentences then the review is confirmed as fake. The user is the blocked.

The user can't login using the same credentials once locked and the same credentials can't be used again to create account.

VI. RESULT

The results obtained as reports in the admin database. The results are obtained depending upon the classification algorithm KNN which is used to divide the contents of spam depending upon its features. The result is generated by checking if the data uploaded is similar to the previously uploaded data, the comment from the user is about the same product and if the comment consists of the spam words and depressed sentences.

Fig 2: System Home Page

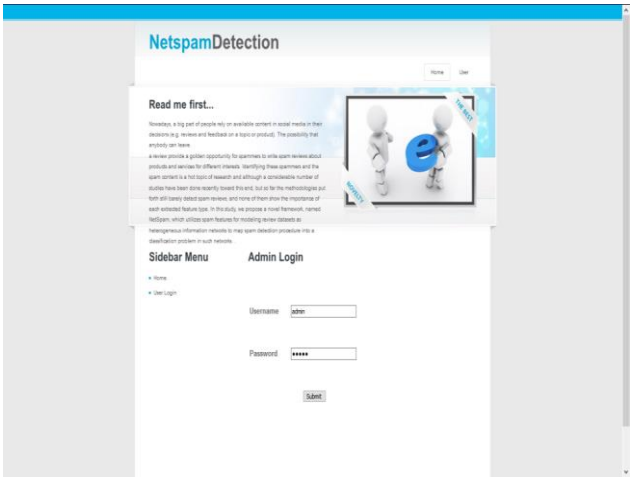

Fig 3: Admin Page

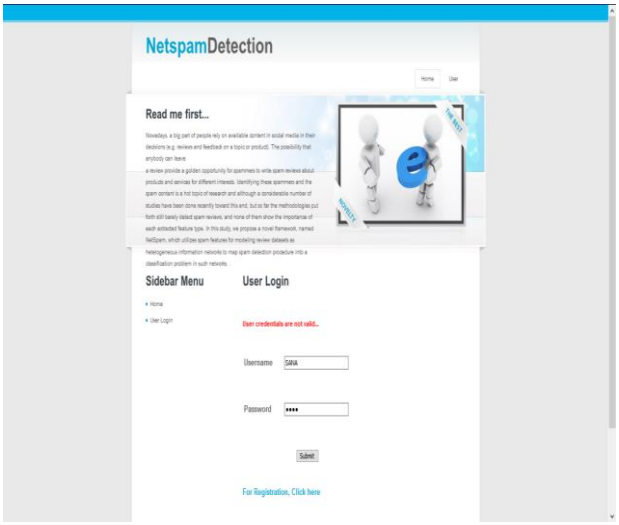

Fig 4: User Acess Denial

The result obtained by this method is more accurate than the existing system as it only obtains the classification data based on the review and user linguistic behavioural nature and the algorithm used to classify the fake and genuine contents vary. The system yields result 70% success rate more than that of the existing system.

VII. CONCLUSION

Online reviews provide valuable information about products and services to consumers. However, spammers are joining the community trying to mislead readers by writing fake reviews. Previous attempts for spammer detection used reviewers' behaviours, text similarity, linguistics features and rating patterns. In the work, we proposed algorithms such as Naive Bayes, conflict detection based are used. Although, each of these approaches can be solely used to classify user accounts, but in order to increase the accuracy, the integrated algorithm is used by combining these two approaches.

It is evident that the proposed algorithm was able to successfully identify an account as spammer or non-spammer with accuracy. The performance of the proposed framework is evaluated by using two real-world labelled datasets of Yelp and Amazon websites. 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.

VIII. FUTURE WORK

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 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 youngresearch. Addressing the problem of spam detection in such networks can be considered as a new research line in this field.

REFERENCES

1) A.J. Minnich, N. Chavoshi, A. Mueen, S. Luan, and M. Faloutsos. (2015) “Trueview: Harnessing the power of multiple review sites”, In ACM WWW.

2) A.Heydari, M. A. Tavakoli, N. Salim, Z. Heydari, (2015) "Detection of review spam: A survey", Expert Syst. Appl., vol. 42, no. 7, pp. 3634-3642.

3) A.Mukherjee, B. Liu, N. Glance. (2012) "Spotting fake reviewer groups in consumer reviews", Proc. ACM WWW, pp. 1-10.

4) B. Viswanath, M. Ahmad Bashir, M. Crovella, S. Guah, K. P. Gummadi, B. Krishnamurthy, and A. Mislove. (2014) “Towards detecting anomalous user behaviour in online social networks”, In USENIX.

5) Bing Liu. (2012) “Sentiment Analysis and Opinion Mining”, Morgan & Claypool Publishers.

6) Ch. Xu and J. Zhang. (2014) “Combating product review spam campaigns via multiple heterogeneous pairwise features”, In SIAM International Conference on Data Mining.

7) E. D. Wahyuni, A. Djunaidy. (2016) "Fake review detection from a product review using modified method of iterative computation framework", Proc. MATEC Web Conf., pp. 1-7.

8) F. H. Li, M. Huang, Y. Yang, X. Zhu. (2011) "Learning to identify review spam", Proc. 22nd Int. Joint Conf. Artif. Intell. (IJCAI), pp. 1-6.

9) G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh. (2013) “Exploiting burstiness in reviews for review spammer detection”, In ICWSM.

10) H. Xue, F. Li, H. Seo, R. Pluretti, (2015) "Trust-aware review spam detection", Proc. IEEE Trustcom/BigDataSE/ISPA, pp. 726-733.

11) 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.