A Network-Based Spam Detection Framework for Reviews in Online Social Media

K. Amar, M. Kameshwara Rao, Ch. Chaitanya, Ravi Kumar Tenali

ABSTRACT: ***Now-a-days, individuals mostly depend on the content in social media in decision making. For instance, they choose to purchase an item depending on the reviews and feedback. Possibility of leaving a review gives a golden chance for spammers to put in writing spam reviews concerning the product and services for various interests. Distinguishing these spammers and accordingly the spam content could be an interesting issue of analysis. Though a substantial range of studies are done recently towards this, till date the methodologies used still barely find the spam reviews. Here, we propose a unique framework called Net-Spam that utilizes spam options for modelling review datasets as heterogeneous data networks to map spam detection procedure into classification. Mistreatment the importance of spam options helps us to get best output in terms of various metrics experimented on real-world review datasets from Yelp and Amazon websites. The results show that Net-Spam outperforms the prevailing ways and among four classes of features; including review-behavioural, user-behavioural, review linguistic, user-linguistic, the primary options performs better than the other categories.***

Index Terms—Social Media, Spammers, Review, Framework, Net-Spam, Heterogeneous data Networks.

I. INTRODUCTION

Social Media portals play vital role in the propagation of information which is an important source for producers and consumers to advertise to select products and services respectively. From the past few years it is observed that people are considering the reviews, be it positive or negative. In terms of business, reviews became an important factor as positive reviews bring benefits whereas negative reviews can cause economic loss. Anyone with any identity can give reviews, this provides an opportunity for spammers to give fake reviews that misleads the user opinion. By sharing, these negative reviews are multiplied over the net. Reviews that are written in exchange of money to change user's perception to buy the product and are considered to be spam. The general idea of our planned framework is to model a given review dataset as a HIN (Heterogeneous Information Network) and to map spam detection into a HIN classification. Specifically, we have a tendency to model review dataset as a HIN in which reviews are connected through different nodes. Weights are calculated and from these weights we calculate the ultimate labels of reviews using unsupervised and supervised approaches. We used two sample datasets of reviews. Based on our observations, shaping two views for options, the classified features as review-behavioural have more weights and yield higher performance on recognising spam reviews in semi-supervisions build no noticeable variation on the performance of our approach. We determined that feature weights are often added or removed for labelling and therefore time complexity is scaled for a specific level of accuracy. We use the features with more weights to get high accuracy with less time complexity. We also classify the features into four major categories that help us to know how much each category of features is contributed to spam detection.

Revised Manuscript Received on April 06, 2019.

K. Amar, His Department is ECM, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P., India.
Ch Chaitanya, His Department is ECM, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P., India.
Ravi Kumar Tenali, His Department is ECM, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P., India.
Dr.M.Kameswara Rao, His Department is ECM, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P., India.

II. THEORETICAL ANALYSIS

In this paper we identify whether the review in the data-set is real or spam. For the better understanding of methodology we first present an overview of some of the concepts in heterogeneous information networks.

A. Definitions:

- ***Heterogeneous Information Network:*** A heterogeneous information network is a graph which is represented as G = (V, E), where each node and each edge belongs to one particular node type and link type respectively.V represents nodes and E represents edge i.e., relationship between two nodes. If two edges belong to the same type, the types of starting node and ending node of those edges are the same.

- ***Network Schema:*** A meta-path with object type mapping and edge mapping is known as network schema. It generally describes about the no.of node types and where the possible edge exists (simply a meta-structure). It is mathematically represented as T = (A, R), where A is object type and relation R.
- ***Meta-path:*** The sequence of relationships in network schema is known as a meta-path. It is generally represented in the form A1(R1)A2(R2)...(R(l−1))Al.It defines a composite relation between two nodes. For convenience, a meta-path can be represented by a sequence of node types when there is no ambiguity, i.e., P = A1A2...Al.There are no edges between two nodes of the same type. The meta-path extends the concept of edge types

- to path types and describes the different relations among node types through indirect links.
- ***Classification problem:*** In the heterogeneous information network, the types of the nodes to be classified. We have some labellednodes and unlabelled nodes. Classification should be done to predict the unlabelled nodes. The nodes are classified into different classes,C1,C2 …Ck. K is the no. Of classes.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

B. *Feature Types:*

As the paper is about spam reviews detection. The data is the written review. The review is written for, rating value of the product. Based on the rate of the review, the business for which the review is written, date of written review it is label as spam or genuine review.The Metapath is defined through their shared features of the reviews between two nodes.In this work features for users and reviews fall into the different categories as follows:

Feature	User	Review
Behavior based feature	This feature depends on the review of single individual user. To generalize all the written reviews, we need to calculate according to the each individual user review. There are two features in user behavior based on which can know whether a review is spam or real. • Burstiness: based on the time of activity of user and time taken to write review and past review time burstiness is calculated. So that based on the burstines value we can know the review is spam or real. Because spam reviews are written fast. • Negative ratio: Basically, the competitors give the ratings as low. Spammers write the reviews to defame the business. So the reviews completely negative with zero rating are spam.	This feature depends on the meta-data. This category has two features. • Early time frame (ETF): The most spam reviews are on the top so that user visit that review first. • Threshold rating deviation of review (DEV): To promote their business the spammers rate high.so that the mean and variance are high based on which we can detect the spam messages.
Linguistic based features	In this feature the opinion and feeling of each individual user are extracted. The spammers generally write the reviews in the same pattern. There are two features in this category.They are average content similarity(ACS)and maximum content similarity(MCS). The spammers don't waste their time in writing original review. They write the same reviews. The values calculated for the similar reviews lie between 0 and 1.	In this feature the opinion and feelings of all reviews are considered. In this category spam reviews are identified based on two attributes. They are the ratio of 1st personal pronouns (PP1) and the Ratio of exclamation sentences (RES). Spammers use second personal pronoun that first personal pronoun and also they use! to impress users. So that the most of the reviews with '!' are noted as spam.

III. METHODOLOGY

A. *Prior Knowledge:*

Initially we should compute prior knowledge that means review u being spam which denoted as y_u.

The framework proposed works in both semi-supervised and unsupervised learning. In the semi-supervisedmethod, if review u is labelled as spam y_u = 1 in thepre-labelled reviews, else y_u = 0. Due to the amount of supervision if the label of thisreview is unknown, we considery_u = 0 i.e., we assume it as a non-spam review. In theunsupervised method, our prior knowledge is realized by using $y_u = (1/L)\sum_{l=1}^{L} f(x_{lu})$ where $f(x_{lu})$ is the probability ofreview u considering an extended version ofthe metapath concept. If they share same value, two reviews are connected to each other. It isbetter to use fuzzy logic for determining a review's label as a spam or non-spam. We have spamcertainty in different levels. We use a step function, to determine these different levels. After computing for all reviews and metapaths, two reviews with thesame metapath values areconnected to each other through that metapath and being spam according to feature l and L is the number of all the used features.

B. *Network Schema definition:*

After prior knowledge we define network schema based on a given spam features which determines the features engaged in spam detection. These Schemas are the general definitions of metapaths
and show different connections of network components.

C. *Meta path definition and creation:*

A Metapath is the sequence of relations in the schema.For the creation of metapath, we define different levels of spamcertainty by create onelink of review network. Using a higher value increases the number ofmetapaths and reviews would beconnected to each other through these features.Since we need enough spam and non-spam reviews for each step, with fewer numbers of reviews connected to each other for every step, the spam probabilities of reviews take uniform distribution, but with lower

value we have enough reviews to calculate final spamicityfor each review. Therefore, accuracy for lower levelsdecreases because of the bipolar problem and it decades forhigher values, because they take uniform distribution.

D. *Algorithm:NETSPAM()*

Inputs: The input for this perticular problem are review dataset, spam feature list, pre-labeled reviews (each review dataset is labeled as spam or real).

Output: features importance(W) and spamcity probability(Pr).

u,v-reviews.

#n- number of reviews.

L-number of features.

#$m_{u,v}^{pl}$-level of spam certainity

#priori-knowledge

1.if semi-supervised mode

if u ∈ prelabeled-reviews

$$y_u = labeled(u)$$

else

$$y_u = 0$$

determining network schema

2. *schema*= determine schema based on spam_feature_list

#metapath creation

3. for $p_l \in schema$

do: for $u,v \in review\_dataset$

do: $m_u^{pl} = \frac{|s \times f(x_{lu})|}{s}$

$$m_v^{pl} = \frac{|s \times f(x_{lu})|}{s}$$

if $m_u^{pl} = m_v^{pl}$

do: $m_{u,v}^{pl} = m_u^{pl}$

else

do: $m_{u,v}^{pl} = 0$

#classification: calculation of weight

4. for $p_l \in schema$

do: $W_{pl} = \frac{\sum_{r=1}^{n}\sum_{s=1}^{n} \boldsymbol{mp_{r,s}^{pl}} \times \boldsymbol{y_r} \times \boldsymbol{y_s}}{\sum_{r=1}^{n}\sum_{s=1}^{n} \boldsymbol{mp_{r,s}^{pl}}}$

#classification: labeling

5.for $u,v \in review\_dataset$

do: $Pr_{u,v} = 1 - \coprod_{pl=1}^{L} 1 - mp_{u,v}^{pl} \times W_{pi}$

$$Pr_u = avg(Pr_{u,1}, Pr_{u,2}, \dots\dots, Pr_{u,n}$$

6. return (W,Pr)

E. *Classification:*

The classification part of NetSpam includes two steps (i) weight calculation(ii) Labelling

i. *Weight calculation:*

This step computes the weight of each metapath. Based on their relations with other nodes in the network it is assumed that the classification of nodes is done linked nodes of high probability may have the same labels. The relation in a HIN includes the direct link and the path that can be measured by using the concept of metapath. Therefore, we use the metapaths defined in the previous step, which represent heterogeneous relations among nodes. Moreover, this step will be able to compute the weight of each relation path that will be used to estimate the label of each unlabelled review in the next step.

ii. *Labelling:*

It is worth to note that in creating the HIN, as much as the number of links between a review and other reviews increase the probability of having a label that is similar, because it assumes that a node is related to other nodes showing similarity. In other words, if a review haslots of links with non-spam reviews, it means features of reviews are shared with other reviews having low spamicity increasing its probability to be a non-spam review.

IV. EXPERIMENTAL EVALUATION

In this section, we discuss about the obtained results based on the dataset and also evaluating the results i.e., whether the proposed approach detects the spam reviews with high accuracy or not. It is determined by the metrics.

A. *Data Set:*

We have collected the data from the Yelp. Yelp is the site where there are reviewsfor the different restaurants, hotels, dentists etc. It also recommends which are the best. We considered almost 608,600 reviews written by the customer for restaurants and hotels in the city New York. In this dataset some of the reviews are labelled as spam or real. The label is given according to the yelp algorithm. It is done by yelp recommender, it is not sure that those labels are perfect but they are trustable. The reviews in the data set contain impression and comments on the quality of the item. The other attributes are the customer id, restaurant name, rate given by the user, date and time when the review is written and when the user visited. This is main dataset. This dataset is manipulated into three other dataset as follows:

- Review based dataset: 20% of the data collected randomly from the main dataset
- Item based dataset: 20 % of the data is collected from the main dataset with the same item.
- User based dataset: minimal set of reviews of the same user are collected.

In addition to these 4 datasets, we have taken the real-world amazon problem. We have taken another dataset from the amazon i.e., Reviews on the amazon site.

Dataset	Reviews (Spam %)	Users	Business (resto & hotels)
Main	608,600(11%)	300,270	6132
Review-based	93980(11%)	58,212	3,827
User-based	94670(35%)	60,342	4.673
Item-based	110,730(20%)	160293	4,623
Amazon	9,000	8325	356

B. *Results:*

We have used Net spam algorithm to detect the spam reviews. The results of net

spam algorithm are compared with other two approaches random approach and Speagle Plus. The accuracy of each approach is compared. Also compared by the feature weights which was discussed in theoretical analysis. This framework is examined in unsupervised mode finally the time complexities are compared.

- ***Accuracy:***To compare those three approaches we have used the average precision (AP) and Area under the Curve(AUC). AUC measures the accuracy based in the True Positive Ration(TPR) against False Positive Ratio(FPR).True positive ratio means real reviews by positive reviews. Fig 2 represents the AUG values for the different dataset and for different approaches. Whereas Fig 1 represents the AP for different datasets.in calculating AP we need to sort the spam reviews in the top of the list. The higher index should be spam review. Then AP is calculated as follow

AP=$\sum_{i=1}^{n} \frac{i}{I(i)}$ where I is index and I is list i.e., dataset.

By the fig.1 and 2 we observe that NetSpam gives the highest accuracy in detecting the spam review. There no effect of supervision on the NetSpam and SPealePlus. The AP value depends on the spam percentage in the dataset whereas AUC values do not change.

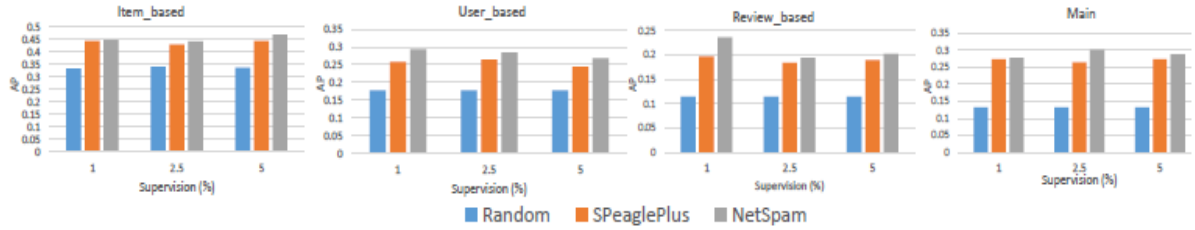

Fig. 1: Average precision (AP) for Random, SPeaglePlus and Net Spam approaches in different datasets and supervisions (1%, 2.5% and 5%)

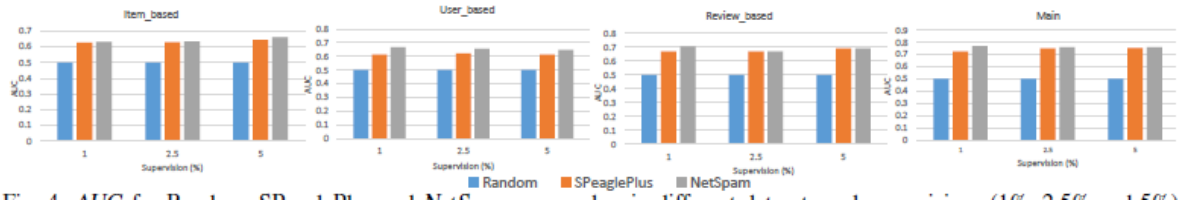

Fig. 2: AUC for Random, SPeaglePlus and Net Spam approaches in different datasets and supervisions (1%, 2.5% and 5%).

- **Feature weight analysis:** This deal with the comparison of features of the dataset. So that we can know by which feature we can detect spam reviews with high accuracy. From the figure 5 we observe that the result of the main dataset is ranked first because it contains all the features .For the feature NR recognise the highest for the every supervision.

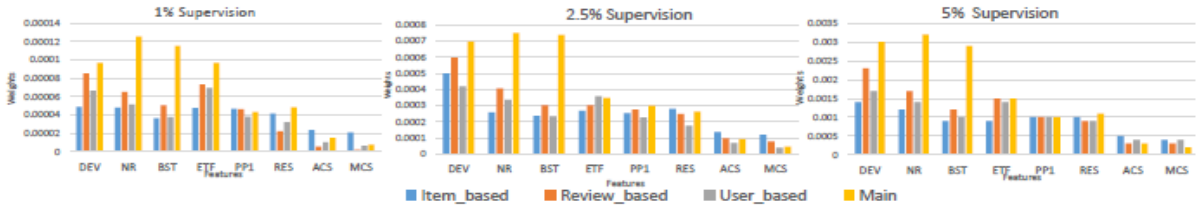

Fig. 3: Features weights for Net Spam framework on different datasets using different supervisions (1%, 2.5% and 5%)

- **Unsupervised method:** In unsupervised approach special process is used to calculate basic labels and next these labels are used to calculate the feature weight and finally review labels.We observe that there is a good correlation between the Main dataset in which for Net Spam it is equal to 0.79 (p-value=0.0207) and for SPeaglePlus this value reach 0.91 (p=0.0022).
- **Time complexity:** let us consider the main dataset as the input dataset. Then the time taken to detect spam reviews in the offline mode is O (e^2m). E is the no.of edges and m is the number of features. Where as in online mode it takes less time approximately O (em) because in online mode there is no need to repeat the process for the every feature like offline mode.

V. CONCLUSION

This study introduces a novel spam detection framework called NetSpam based on the concepts of metapath andgraph-based method of labelling reviews depending on rank-based approach of

labelling. The performance of the framework is evaluated by using real-world labelled datasets of Yelp and Amazon websites. By our observations, we showthat calculated weights by using the concept of meta path can bevery effective in identifying spam reviews and leads to a better performance. In addition, we found that even without a trained set, Net Spam can calculate each feature's importance and yields better performance in the process of features addition, and performs better than previous works, with only asmall number of features. Moreover, after defining four maincategories for features our observations show that the reviews behavioural category performs better than other categories. The result confirms that using various supervisions that are similarto the semi-supervised method will not have effect on
determining features that are most weighted as in various datasets.

In future,concept of Metapath can be applied to otherproblems. For example, to find spammer communities frameworks that are similar can be used. For community finding,we connect
reviews through features of group spammer and reviews with highest similarity based on metapath concept are known ascommunities.Using the product features is aninteresting future work as we use features that arerelated for spotting spammers and spam reviews. Moreover,when single network receives attention fromdifferent disciplines for a decade, information diffusionand content sharing in multilayer networks is still a youngresearch.

REFERENCES

1. Sk. Naseem, A technique to notice spam reviews on e-shopping
2. A Vishal Dyandeo et al., Survey in online social media on Nov-2017 in IRJET
3. B Harshith Reddy, Prof. Geetha B Net spam: A fake review detector International journal of advanced in computer science 2010
4. Saradha. R et al., Detecting spam reviews using user behaviours and unusual review pattern, International journal of engineering and techniques
5. Jeff Hancock et al., Estimating prevalence of deception in online review communities2012 International WWW conference
6. Bling Liu, J Nitin CSE, Opinion spam and analysis
7. Meichun Hsu, Malu C et al., Exploiting burstiness in reviews for review spammer detection, Seventh international AAAI conference, 2013.
8. W Xiaokai et al. Spotting fake reviews via collective positive-unlabeled learning, IEEE conference 2014
9. A leman, Chandy Rishi, F christos, Opinion fraud detection in online reviews by network effects ICWSM 2013
10. Xu Chang et al., Collusive opinion fraud detection in online reviews:A probabilistic modeling approach, Volume 11, 2017 Sep
11. C Yejin, B Ritwik, S Feng, Syntactic stylometry for deception detection, ACL Annual meeting July 2012
12. Guan W et al., Review graph based online store review spammer detection, 11th conference IEEE December 2011
13. Tianyi Wu, C Hong et al., RankClus: integrating clustering with ranking for HIN analysis.International conference on extending DB tech, 2009
14. D Arif, E W Dyar Detection of fake review from product review using iterative computationframework modified method, MATEC 2016 january.
15. Kalyani Adhav, Gawali S.Z, Ravindra Murumkar, Survey on online spam review detection methodsIJCSIT 2014