

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

A Survey on Spam Review Detection Using NetSpam Framework

Srushti Ghatage, Kiran Hinge, Gayatri Holkar, Prayag Kamble

B. E Students, Department of Computer Engineering, Sinhgad Institute of Technology & Science, Narhe, Pune, India

ABSTRACT: Major Society of people using internet trust the contents of net. The liability that anyone can take off a survey give a brilliant chance to spammers to compose spam surveys about hotels and services for various interests. Recognizing these spammers and the spam content is a widely debated issue of research and in spite of the fact that an impressive number of studies have been done as of late towards this end, yet so far the procedures set forth still scarcely distinguish spam reviews, and none of them demonstrate the significance of each extracted feature type. In this application, we use a novel structure, named NetSpam, which proposes spam features for demonstrating hotel review datasets as heterogeneous information networks to design spam review detection method into a classification issue in such networks. Utilizing the significance of spam features helps us to acquire better outcomes regarding different metrics on review datasets. The outcomes represent that NetSpam results with the previous methods and encompassed by four categories of features; involving review-behavioral, user-behavioral, review linguistic, user-linguistic, the first type of features performs better than the other categories. The contribution work is when user will search query it will display all top products as well as there is recommendation of the product.

KEYWORDS: Social Media, Social Network, Spammer, Spam Review, Fake Review, Heterogeneous Information Networks.

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.

II. MOTIVATION

NetSpam is able to find features' importance relying on metapath definition and based on values calculated for each review. NetSpam improves the accuracy and reduces time complexity. It highly depends to the number of features used to identify spam reviews. Thus using features with more weights will result in detecting spam reviews easier with lesser time complexity.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

III.LITERATURE SURVEY

The pair wise features are first explicitly utilized to detect group colluders in online product review spam campaigns, which can reveal collusions in spam campaigns from a more fine-grained perspective. A novel detecting framework [1] named Fraud Informer is proposed to cooperate with the pair wise features which are intuitive and unsupervised. Advantages are: Pair wise features can be more robust model for correlating colluders to manipulate perceived reputations of the targets for their best interests to rank all the reviewers in the website globally so that top-ranked ones are more likely to be colluders. Disadvantage is difficult problem to automate.

The paper [2] proposes to build a network of reviewers appearing in different bursts and model reviewers and their co-occurrence in bursts as a Markov Random Field (MRF) and apply the Loopy Belief Propagation (LBP) method to induce whether a reviewer is a spammer or not in the graph. A novel assessment method to evaluate the detected spammers automatically using supervised classification of their reviews. Advantages are: High accuracy, the proposed method is effective. To detect review spammers in review bursts. To detect spammers automatically. Disadvantage is: a generic framework is not used for detect spammers.

In [3] paper, the challenges are: The detection of fraudulent behaviors, determining the trustworthiness of review sites, since some may have strategies that enable misbehavior, and creating effective review aggregation solutions. The TrueView score, in three different variants, as a proof of concept that the synthesis of multi-site views can provide important and usable information to the end user. Advantages are: develop novel features capable of finding cross-site discrepancies effectively, a hotel identity-matching method with 93% accuracy. Enable the site owner to detect misbehaving hotels. Enable the end user to trusted reviews. Disadvantage is difficult problem to automate.

In [4] paper describes unsupervised anomaly detection techniques over user behavior to distinguish probably bad behavior from normal behavior. To find diverse attacker schemes fake, compromised, and colluding Facebook identities with no a priori labeling while maintaining low false positive rates. Anomaly detection technique to forcefully identify anomalous likes on Facebook ads. Achieves a detection rate of over 66% (covering more than 94% of misbehavior) with less than 0.3% false positives. The attacker is trying to drain the budget of some advertiser by clicking on ads of that advertiser.

In [5] paper, a grouped classification algorithm called Multi-typed Heterogeneous Collective Classification (MHCC) and then extends it to Collective Positive and Unlabeled learning (CPU). The proposed models can markedly increase the F1 scores of strong baselines in both PU and non-PU learning environment. Advantages are: Proposed models can markedly increase the F1 scores of strong baselines in both PU and non-PU learning settings. Models only use language self-contained features; they can be smoothly generalized to other languages. Detects a large number of potential fake reviews hidden in the unlabeled set. Fake reviews hiding in the unlabeled reviews that Dianping's algorithm did not capture. The ad-hoc labels of users and IPs used in MHCC may not be very specific as they are computed from labels of neighboring reviews.

IV.OPEN ISSUES

Online Social Media websites play a main role in information propagation which is considered as an important source for producers in their advertising operations as well as for customers in selecting products and services. People mostly believe on the written reviews in their decision-making processes, and positive/negative reviews encouraging/discouraging them in their selection of 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 reviews 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.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

Disadvantages:

- There is no information filtering concept in online social network.
- People believe on the written reviews in their decision-making processes, and positive/negative reviews encouraging/discouraging them in their selection of products and services.
- Anyone create registration and gives comments as reviews for spammers to write fake reviews designed to misguide users' opinion.
- Less accuracy.
- More time complexity.

V.SYSTEM OVERVIEW

The 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. 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. The feature weights can be added or removed for labeling and hence time-complexity can be scaled for a specific level of accuracy. 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.

1. NetSpam framework that is a novel network based approach which models review networks as heterogeneous information networks.
2. 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.
3. NetSpam 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.

The general concept of our proposed framework is to model a given review dataset as a Heterogeneous Information Network and to map the problem of spam detection into a HIN classification problem. In particular, we model review dataset as in which reviews are connected through different node types.

A weighting algorithm is then employed to calculate each feature's importance. These weights are utilized to calculate the final labels for reviews using both unsupervised and supervised approaches. Based on our observations defining two views for features.

Advantages:

1. To identify spam and spammers as well as different type of analysis on this topic.
2. Written reviews also help service providers to enhance the quality of their products and services.
3. To identify the spam user using positive and negative reviews in online social media.
4. To display only trusted reviews to the users.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

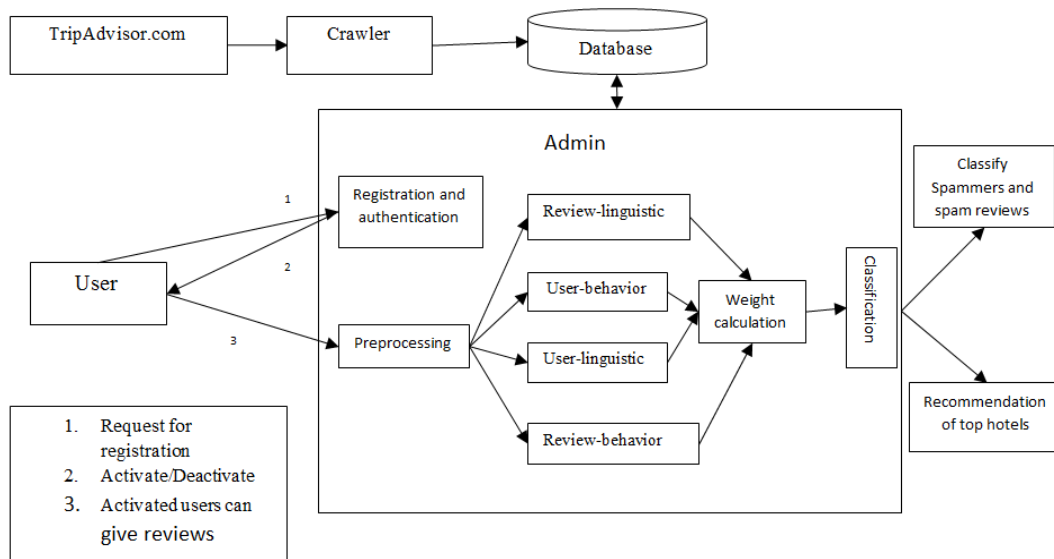


Fig No 01 System Architecture

MATHEMATICAL MODEL

User-Behavioral (UB) based features:

Burstiness: Spammers, usually write their spamreviews in short period of time for two reasons: first,because they want to impact readers and other users,and second because they are temporal users, they haveto write as much as reviews they can in short time.

$$x_{BST}(i) = \begin{cases} 0 & (L_i - F_i) \notin (0, \tau) \\ 1 - \frac{L_t - F_t}{\tau} & (L_i - F_i) \in (0, \tau) \end{cases} \quad (1)$$

Where,

$L_i - F_i$ describes days between last and firstreview for $\tau = 28$.

Users with calculated value greaterthan 0.5 take value 1 and others take 0.

User-Linguistic (UL) based features:

Average Content Similarity, Maximum Content Similarity: Spammers, often write their reviews with same template and they prefer not to waste their time to write an original review. In result, they have similar reviews. Users have close calculated values take same values (in [0; 1]).

Review-Behavioral (RB) based features:

- Early Time Frame: Spammers try to write their reviews a.s.a.p., in order to keep their review in the top reviews which other users visit them sooner.

$$x_{ETF}(i) = \begin{cases} 0 & (L_i - F_i) \notin (0, \delta) \\ 1 - \frac{L_t - F_t}{\delta} & (L_i - F_i) \in (0, \delta) \end{cases} \quad (2)$$

Where,

$L_i - F_i$ denotes days specified written review and first written review for a specific business. We have also $\delta = 7$.

Users with calculated value greater than 0.5 takes value 1 and others take 0.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

- Rate Deviation using threshold: Spammers, also tend to promote businesses they have contract with, so they rate these businesses with high scores. In result, there is high diversity in their given scores to different businesses which is the reason they have high variance and deviation.

$$x_{DEV}(i) = \begin{cases} 0 & \text{Otherwise} \\ 1 - \frac{r_{ij} - \text{avg}_{e \in E^*} r(e)}{4} & > \beta_1 \end{cases} \quad (3)$$

Where,

β_1 is some threshold determined by recursive minimal entropy partitioning. Reviews are close to each other based on their calculated value, takesame values (in [0; 1)).

Review-Linguistic (RL) based features:

Number of first Person Pronouns, Ratio of Exclamation Sentences containing '!': First, studies show that spammers use second personal pronouns much morethan first personal pronouns. In addition, spammers put '!' in their sentences asmuch as they can to increase impression on users and highlight their reviewsamong other ones. Reviews are close to each other based on their calculatedvalue, take same values (in [0; 1]).

VI. CONCLUSION

We use novel spam detection framework named NetSpam based on a metapath creation as well as new graph-based method for labeling reviews relying on a rank-based labeling approach. The calculated weights by utilizing this metapath concept can be very impressive in identifying spam reviews and spammers leads to a better performance. In extension, we found that even without a train set, NetSpam can calculate the consequence of each feature and it yields better performance in the features' addition process, and performs better than existing works, with only a small number of features. Moreover, after defining four main categories for features our conclusions show that the reviews behavioral category performs better than other categories.

REFERENCES

- [1] Ch. Xu and J. Zhang. Combating product review spam campaigns via multiple heterogeneous pairwise features. In SIAM International Conference on Data Mining, 2014.
- [2] G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh. Exploiting burstiness in reviews for review spammer detection. In ICWSM, 2013.
- [3] A. J. Minnich, N. Chavoshi, A. Mueen, S. Luan, and M. Faloutsos. Trueview: Harnessing the power of multiple review sites. In ACM WWW, 2015.
- [4] 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.
- [5] H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao. Spotting fake reviews via collective PU learning. In ICDM, 2014.