International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 03 Issue: 09 | Sept -2019 ISSN: 2590-1892

AN IMPROVED NETWORK-BASED SPAM DETECTION FRAMEWORK FOR REVIEWS IN ONLINE SOCIAL MEDIA

N SWAPNA (PG SCHOLAR)
Dept. of COMPUTER SCIENCE AND ENGINEERING.
SREEYAS INSTITUTE OF ENGINEERING&TECHNOLOGY

Dr. ROHIT RAJA (PROFESSOR)
Dept. of COMPUTER SCIENCE AND ENGINEERING.
SREEYAS INSTITUTE OF ENGINEERING&TECHNOLOGY

ABSTRACT: Todays, a major part of everyone trusts on content in social media like opinions and feedbacks of a topicor a product. The liability that anyone can take off a survey give a brilliant chance to spammers to compose spamsurveys about products and services for various interests. Recognizing these spammers and the spam content is a wildlydebated issue of research and in spite of the fact that an impressive number of studies have been done as of late towardthis end, yet so far the procedures set forth still scarcely distinguish spam reviews, and none of them demonstrate thesignificance of each extracted feature type. In this investigation, we propose a novel structure, named Net Spam, whichuses spam highlights for demonstrating review datasets as heterogeneous information networks to design spamdetection method into a classification issue in such networks. Utilizing the significance of spam features help us toacquire better outcomes regarding different metrics on review datasets. The outcomes demonstrate that Net Spam resultsthe existing methods and among four categories of features; including review-behavioral, user-behavioral,review linguistic, user-linguistic, the first type of features performs better than the other categories. The contribution work iswhen user search query it will display all top-k products as well as recommendation of the product.

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

INTRODUCTION

Online Social Media portals play an influential role in information propagation which is considered as an importantsource for producers in their advertising campaigns as well as for customers in selecting products and services. In thepast years, people rely a lot on the written reviews in their decision-making processes, and positive/negativereviewsencouraging/discouraging them in their selection of productsand services. In addition, written reviews also help serviceproviders to enhance the quality of their products and services. These reviews thus have become an important factor insuccessof a business while positive reviews can bring benefits for a company, negative reviews canpotentially

impactcredibility and cause economic losses. The fact that anyone with any identity can leave comments as review, provides atempting opportunity for spammers to write fake reviews designed to mislead users' opinion. These misleading reviewsare then multiplied by the sharing function of social media and propagation over theweb.

EXISTING SYSTEM APPROACH

Online Social Media websites play a main role in information propagation which is considered as an important sourcefor producers in their advertising operations as well as for customers in selecting products and services. People mostlybelieve on the written reviews in their decision-making processes, and positive/negative reviewsencouraging/discouraging them in their selection of products and services. These reviews thus have become animportant factor in success of a business while positive reviews can bring benefits for a company, negative reviews canpotentially impact credibility and cause economic losses. The fact that anyone with any identity can leave comments asreviews provides a tempting opportunity for spammers to write fake reviews designed to mislead users' opinion. Thesemisleading reviews are then multiplied by the sharing function of social media and propagation over the web. Thereviews written to change users' perception of how good a product or a service are considered as spam, and are oftenwritten in exchange formoney.

Disadvantages:

1. There is no information filtering concept in online socialnetwork.
2. People believe on the written reviews in their decision-making processes, and positive/negative reviewsencouraging/discouraging them in their selection of products andservices.
3. Anyone create registration and gives comments as reviews for spammers to write fake reviews designedto misguide users' opinion.
4. Lessaccuracy.
5. More timecomplexity.

PROPOSED SYSTEM APPROACH

The proposed framework is to model a given review dataset as a Heterogeneous Information Network (HIN) and tomap the problem of spam detection into a HIN classification problem. In particular, we model review dataset as a HINn which reviews are connected through different node types (such as features and users). A weighting algorithm is thenemployed to calculate each feature's importance (or weight). These weights are utilized to calculate the final labels forreviews using both unsupervised and supervised approaches. Based on our observations, defining two views forfeatures (review-user and behavioral-linguistic), the classified features as review behavioral have more weights andyield better performance on spotting spam reviews in both semi-supervised and unsupervised approaches. The featureweights 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 howmuch each category of features is contributed to spam

detection.

1. Net Spam framework that is a novel network based approach which models review networks as heterogeneousinformationnetworks.
2. A new weighting method for spam features is proposed to determine therelative importance of each featureand shows how effective each of features are in identifying spams from normalreviews.
3. Net Spam improves the accuracy compared to the state-of-the art in terms of time complexity, which highlydepends to the number of features used to identify a spamreview.

The general concept of our proposed framework is to model a given review dataset as a Heterogeneous InformationNetwork and to map the problem of spam detection into a HIN classification problem. In particular, we model reviewdataset 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 calculatehe final labels for reviews using both unsupervised and supervised approaches. Based on our observations definingtwo views for Features.

Advantages:

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

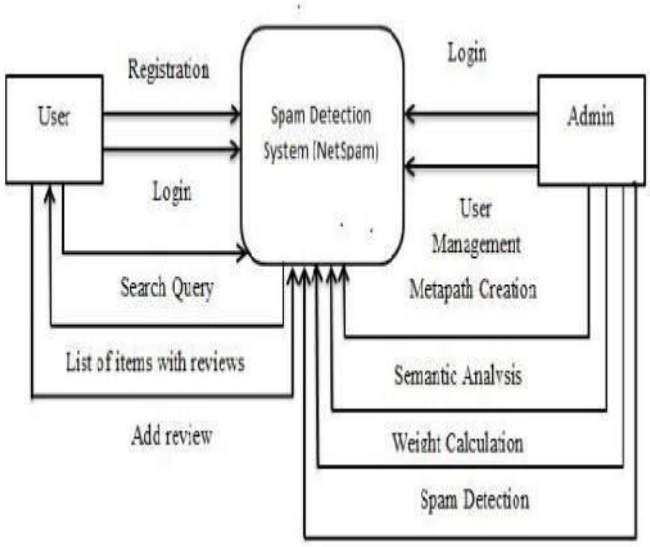

Fig.1: System Architecture

MATHEMATICAL MODEL FOR THE LOW-LEVEL DESIGN (MODULE-WISE)

Spam Features:

User-Behavioral (UB) based features: Burstiness: Spammers, usually write their spam reviews in short period of time for two reasons: first, because they wantto impact readers and other users, and second because they are temporal users, they have to write as much as reviewsthey can in shorttime**.**

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

User-Linguistic (UL) based features: Average Content Similarity, Maximum Content Similarity: Spammers, often write their reviews with same templateand they prefer not to waste their time to write an original review. In result, they have similar reviews. Users have closecalculated 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 topreviews which other users visit them sooner.

Review-Linguistic (RL) based features: Number of first Person Pronouns, Ratio of Exclamation Sentences containing '!': First, studies show Reviews are close toeach other based on their calculated value, take same values (in [0;1]).

CONCLUSION

This investigation presents a novel spam detection system in particular Net Spam in view of a met path idea andanother graph based strategy to name reviews depending on a rank-based naming methodology. The execution of theproposed structure is assessed by utilizing review datasets. Our perceptions demonstrate that ascertained weights byutilizing this met path idea can be exceptionally powerful in recognizing spam surveys and prompts a superiorexecution. Furthermore, we found that even without a prepare set, Net Spam can figure the significance of each elementand it yields better execution in the highlights'expansion procedure, and performs superior to anything past works, withjust few highlights. In addition, in the wake of characterizing four fundamental classifications for highlights ourperceptions demonstrate that the review behavioral classification performs superior to anything different classifications,regarding AP, AUC and in the ascertained weights. The outcomes likewise affirm that utilizing diverse supervisions,like the semi- administered strategy, have no detectable impact on deciding the vast majority of the weighted highlights,similarly as in various datasets. Contribution part in this project, for user when searches query he will get the top-khotel lists as well as one recommendation that spammers usesecondpersonal pron hotel by using personalized recommendationalgorithm.

REFERENCES

[1] Ch. Xu and J. Zhang. Combating product review spam campaigns via multiple heterogeneous pairwise features. In SIAM InternationalConference on Data Mining, 2014.

[2] G.Fei,A.Mukherjee,B.Liu,M.Hsu, M. Castellanos, and R. Ghosh.Exploitingburstiness in reviews for review spammer detection.InICWSM, 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 anomaloususer behavior in online social networks.In USENIX,2014.