International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 3 | ISSN : 2456-3307

NET-SPAM : A Network Based Spam Detection Framework For reviews in online Social Media

Sayyeda Zeba, Zarina Begam K Mundargi
Department of Computer Science Engineering, SECAB Institute of Engineering and Technology,Vijayapura Karnataka, India

ABSTRACT

Now a days, people confide on available content in social media in their decisions (e.g. reviews and feed back on a topic or product).For different interests and services, a spammers which can write spam reviews about their products that can leave a review. So far strategy used to detect spam reviews to show importance of each extracted feature type. A novel structure, named Net spam, which utilizes spam features for modeling review datasets as heterogeneous information networks to map spam detection procedure into classification problems in such networks. with the help of this features it help us to obtain better results for different experimented metrics on real-world review datasets from Amazon websites.Net Spam out performs the existing methods among four categories of features are; review-behavioral, user-behavioral, review linguistic, user-linguistic, review behavioral performs better than other categories.

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

I. INTRODUCTION

Online Social Media play an influential role in information propagation, which is used as important source in advertising campaigns for producers and selecting products and services for customer. People rely on the written reviews in decision-making processes, for selection process of products and services positive/negative reviews encouraging/discouraging reviews are used.Written reviews helps to enhance the quality of products and services for service providers. These reviews have become an important factor in success of a business, for positive reviews bring benefits for a company, negative reviews can potentially impact credibility and cause economic losses. Identity can leave comments as review and provide tempting opportunity for spammers to write fake reviews which is used mislead users opinion. Misleading reviews are 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 [1], and are often written in exchange for money.[2]

Feature Types:

In this I have used metapath concept.A metapath is defined as a path between two nodes,which indicates the connection of two nodes through their shared features. In this work features for users and reviews fall into categories are review-behavioral, review linguistic, user-behavioral, user-linguistic.

Review-Behavioral (RB) based features. This feature type is based on metadata and not the review text itself. The RB category contains two features; Early time frame (ETF) and Threshold rating deviation of review (DEV) [3].

Review-Linguistic (RL) based features. This is based on the review itself and extracted directly from text of the review, two main features are used in RL category; the Ratio of 1st Personal Pronouns (PP1) and the Ratio of exclamation sentences containing '!' (RES) [4].

User-Behavioral (UB) based features. These features are specific to each individual user and they are calculated per user, so these features are used to generalize all of the reviews written by that specific user. This category has two main features; the Burstiness of reviews written by a single user [5], and the average of a users' negative ratio given to different businesses [6].

User-Linguistic (UL) based features. These features are extracted from the users' language and shows how users are describing their feeling or opinion about what they've experienced as a customer of a certain business. This type of features is to understand how a spammer communicates in terms of wording. There are two features engaged for our framework in this category; Average Content Similarity (ACS) and Maximum Content Similarity (MCS). These two features show how much two reviews written by two different users are similar to each other, as spammers tend to write very similar reviews by using template pre-written text [7].

II. LITERATURE SURVEY

Spam campaigns spotted in popular product review websites (e.g., amazon.com), where a group of online posters are hired to collaboratively craft deceptive reviews for some target products.

The goal is to manipulate perceived reputations of the targets for their best interests. Many efforts have been made to detect such colluders by extracting point wise features from individual reviewers/reviewer-groups; however, pair wise features which can potentially capture the underlying correlations among colluders are either ignored or just explored insufficiently in the literature [8]

In this I observed that pairwise features can be more robust to model the relationships among colluders since them, as the ingredients of spam campaigns, are correlated in nature. In this paper, I explore multiple heterogeneous pairwise features in virtue of some collusion signals found in reviewers' rating behaviors and linguistic patterns.

I have used an unsupervised anomaly detection technique to build an Anomaly classifier that learns normal patterns of behavior .Any behavior that deviates significantly from normal is anomalous for learning phase: Input only includes behavior of unlabeled random sample of users this approach has the potential to catch diverse attack strategies[9].

Click-spam on Facebook

Advertisers lose money on spam clicks they might lose confidence in the advertising platform Affects the sustainability of the social networking service Preliminary experiment to understand click-spam in Facebook ads set up bluff ad and a real ad targeting users in USA Heavily instrumented the landing page to capture user activity both bluff and real ad performed nearly identically e.g., similar number of clicks and similar levels of activity on landing page.

Service abuse is a huge problem in social networks today Attackers use diverse strategies and also tend to adapt, so I propose an unsupervised anomaly detection scheme PCA serves as a nice tool to model behavior and detect anomalous ones. So this evaluate our technique on extensive ground-truth data of anomalous behavior to apply our approach to detect click-spam in a social ad platform Sp Eagle collectively utilizes both metadata(review text, timestamp, rating) and the review network(plus available labels, if any) under a unified framework to rank all of users, reviews, and products by spamicity.

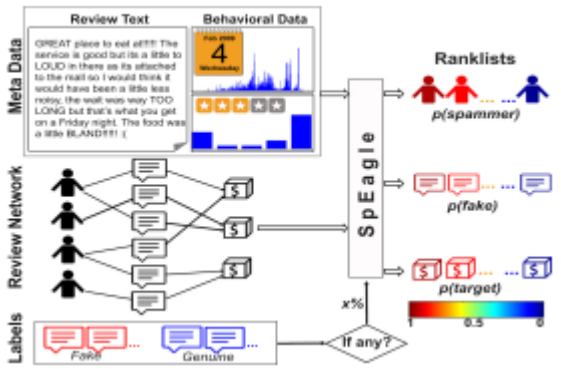

Figure 1. spamicity

Authors SpEagle[10], a new approach for the opinion spam detection problem, which ties together relational data with metadata, i.e., it utilizes all of graph, behavioral, and review content information collectively.

1) Trust worthy Large Scale Social Networks Evaluation

2) Data Privacy Preserving

3) Friend Recommendation

4) Vote Trust In social Networking

5) Trust Based web recommendations

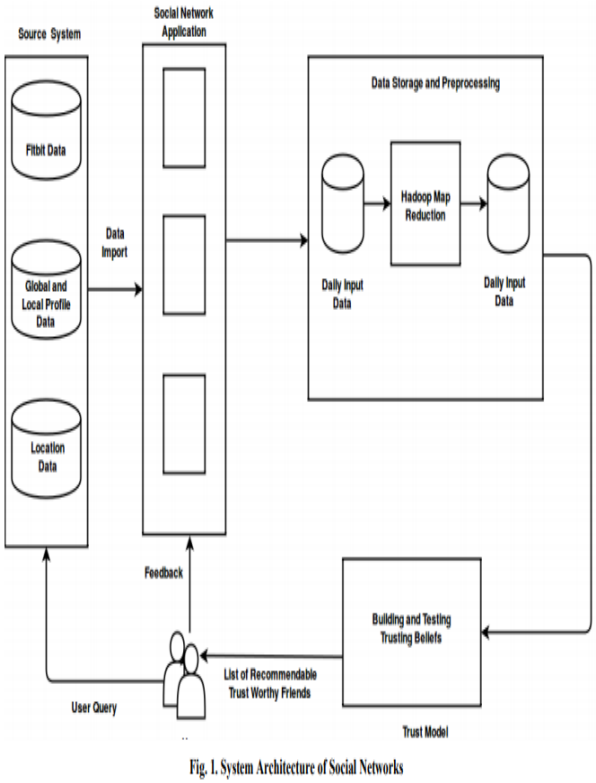

Figure 2. system Architecture of Social Network

It Performs Sentiment classification that determines whether a Review is positive, negative or neural.Featured base-opinion mining that discovers features of a reviewed entity with the intent of acquiring the opinion of a reviewer about that specific feature and providing a spam free content.The propose work discuss modules that will perform the process as Review spam detection, non-Review spam detection,Brand spam detection and filter out the spam content.The recent work related to spam detection is done with classifier, language model and Decision tree, which gives more efficiency and trustworthiness while detecting and filtering the spam content. The results are promising. Supervisors, controllers, organizations can use review spam detection result as an administrative tool to supervise target e-commerce accumulation. The system gives convenience to administrators, flexible settings are available. It provides efficient and trustworthy opinion and feedback[11].

III. RESULTS AND DISCUSSION

NetSpam from different perspective and compare it with two other approaches, Random approach and SPeaglePlus [12]. To compare with the first one, I have developed a network in which reviews are connected to each other randomly. Second approach use a well known graph-based algorithm called as “LBP” to calculate final labels. Our observations show NetSpam, outperforms these existing methods. Then analysis on our observation is performed and finally it will examine the framework in unsupervised mode. Lastly, this investigate time complexity of the proposed framework and the impact of camouflage strategy on its performance.

1) Accuracy: The four datasets NetSpam outperforms SPeaglePlus specially when number of features increase. In addition different supervisions have no considerable effect on the metric values neither on NetSpam nor SPeaglePlus. Results also show the datasets with higher percentage of spam reviews have better performance because when fraction of

spam reviews in a certain dataset increases, probability for a review to be a spam review increases and as a result more spam reviews will be labeled as spam reviews and in the result of AP measure which is highly dependent on spam percentage in a dataset. On the other hand, AUC measure does not fluctuate too much, because this metric is not dependent on spam reviews percentage in dataset, but on the final sorted list which is calculated based on the final spam probability.

2) Feature Weights Analysis: features weights and their involvement to determine spamicity. First it will inspect how much AP and AUC are dependent on variable number of features. Then show these metrics are different for the four feature types explained before (RB,UB, RL and UL). To show how much the work has done on weights calculation is effective, first I have simulated framework on several run with whole features and used most weighted features to find out best combination which gives us the best results.

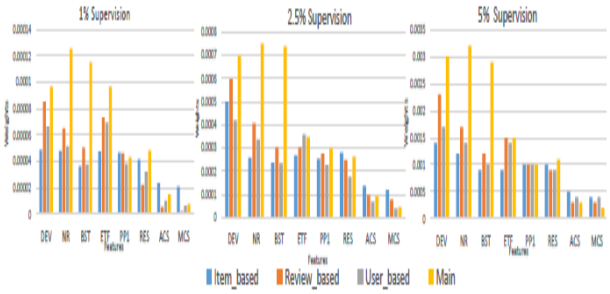

Figure 3. Features weights for Net spam framework on different datasets using different supervision

IV. CONCLUSION

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 meta path concept are known as communities. In addition, utilizing the product features is an interesting future work on this study and these are 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 multi layer networks is still a young research. Addressing the problem of spam detection in such networks can be considered as a new research line in this field.

V. REFERENCES

[1]. 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.

[2]. 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.

[3]. 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.

[4]. F. Li, M. Huang, Y. Yang, and X. Zhu. Learning to identify review spam. Proceedings of the 22nd International Joint Conference on Artificial Intelligence; IJCAI, 2011.

[5]. G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh. Exploiting burstiness in reviews for review spammer detection. In ICWSM, 2013.

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

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

[8]. Survey on "Combating product review spam campaigns via multiple heterogeneous pairwise features" (Ch. Xu and J. Zhang)

[9]. "Towards detecting anomalous userbehavior in online social networks".(B. Viswanath, M.

Ahmad Bashir, M. Crovella, S. Guah, K. P. Gummadi, B. Krishnamurthy, and A. Mislove.)

[10]. "Collective opinion spam detection: bridging review networks and metadata". (R. Shebuti and L. Akoglu.)

[11]. "Trust-Aware Review SpamDetectio"(H. Xue, F. Li, H. Seo, and R. Pluretti.)11]

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