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

Sayyeda Zeba[1] Zarinabegam K Mundargi[2]
[1,2]Department of Computer Engineering
[1,2]SECAB Institute of Engineering and Technology

Abstract— Now a days, people confide on available content in social media in their decisions (e.g. reviews and feedback 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.

Key words: 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 user's 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]

A. Feature Types

In this I have used Meta path concept. A Meta path 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, and 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. EXISTING SYSTEM

Existing system 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 behavioral patterns that rely on features extracted from patterns in users' behavior which are mostly meta data based and even some techniques using graphs and graph-based algorithms and classifiers.

Existing system can be summarized into three categories: Linguistic-based Methods, Behavior-based Methods and Graph-based Methods.

A. Disadvantages of Existing System

- The fact 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.

III. PROPOSED SYSTEM

The general concept of 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, model review dataset as a HIN in which reviews are connected through different node types (such as features and users).

Net Spam proposes framework that is a novel network based approach which models review networks as heterogeneous information networks. The classification step

uses different Meta path types which are innovative in the spam detection domain.

- A 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.
- Net Spam improves the accuracy compared to the state of- the art in terms of time complexity, which highly depends on 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.

IV. ADVANTAGES OF PROPOSED SYSTEM

- Improved Accuracy
- Easier in detecting fake reviews
- Less time Complexity

Net Spam is able to find features importance even without ground truth, and only by relying on Meta path definition and values are calculated for each review.

No previous method are used to engage importance of features.

V. IMPLEMENTATION

A. Modules

1) Admin

Admin is the main user of our application, after login admin can add/delete products for user shopping and collect the user's reviews. After admin can perform spam detection for using unsupervised models. In unsupervised models admin will perform User Based, Review Based spam detection models. Using supervised approaches admin will perform weight calculations.

2) User

User is the end user of our application, and we can also consider e-commerce user. User can perform search products, buy products, and submit rating and review for products. This data will deliver to admin, and this data will be our dataset to our application.

3) Heterogeneous Information Network (HIN)

This is mapping the problem of spam detection into a HIN classification problem. In particular, the model review dataset as a HIN in which reviews are connected through different node types (such as User Based, Review Based). 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.[3]

VI. SOFTWARE ARCHITECTURE

A. About Project Software's

JAVA, Apache Server, MSQL, EDIT ++

In web Application Development are using one tier architecture as total applicant will be developed in single system with all the three layers of application development like presentation layer where this web technologies are used to make of GUI of the application like HTML, HTML-5, CSS, JS Etc. and in second layer it have to make our business logic or called as implementation of application where this is used java, J2EE and also used JDBC to connect from our Business layer to data base layer and final our data base layer where we develop the Data structure of the application

Fig. 1: Single tire Architecture Project Development

B. System Requirements

1) Project Hardware Requisites

Hardware	:	Pentium
Speed	:	1.1 GHz
RAM	:	1GB
Hard Disk	:	20 GB

2) Min Software requisites For Project Development

Operating System	: Windows Family
Technology	: Java and J2EE
Web Technologies	: Html, JavaScript, CSS
Web Server	: Apache Tomcat 7.0/8.0
Database	: Free Download My SQL 5.5 or Higher
UML's	: Star UML
Java Version	: Open Source JDK 1.7 or 1.8

VII. RESULTS & DISCUSSION

A. Project Home Page

Net Spam from different perspective and compare it with two other approaches, Random approach and SPeaglePlus [8]. 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 Net Spam, 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 especially 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 spam city. 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.

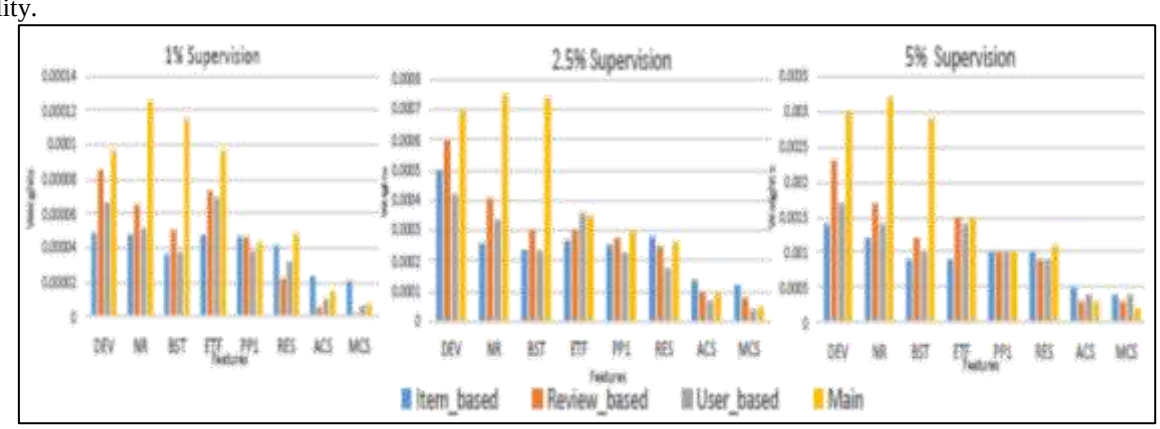

Fig. 2: Features weights for Net Spam Framework on Different Datasets using Different Supervision

VIII. 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.

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

