Spammer Detection and Fake User Identification on Social Networks

V. N. V. L. S. Swathi¹, B. Mouna Shruthi², R. Himavanth Kumar³

¹Assistant Professor, Dept. of Computer Science and Engg., CVR College of Engineering, Hyderabad, India
²³Student, Department of Computer Science and Engineering, CVR College of Engineering, Hyderabad, India

Abstract: Social networking sites engage millions of users around the world. The users’ interactions with these social sites, such as Twitter and Facebook have a tremendous impact and occasionally undesirable repercussions for daily life. The prominent social networking sites have turned into a target platform for the spammers to disperse a huge amount of irrelevant and deleterious information. Twitter, for example, has become one of the most extravagantly used platforms of all times and therefore allows an unreasonable amount of spam. Fake users send undesired tweets to users to promote services or websites that not only affect legitimate users but also disrupt resource consumption. Nowadays, a big part of people relies on available content in social media in their decisions for example, reviews and feedback on a topic or product. The possibility that anybody can leave a review provides a golden opportunity for spammers to write spam reviews about products and services for different interests. Identifying these spammers and the spam content is a hot topic of research and although a considerable number of studies have been done recently toward this end, but so far, the methodologies put forth still barely detect spam reviews, and none of them show the importance of each extracted feature type.

Keywords: Classification, Fake user detection, Online social network, Spammer’s identification.

1. Introduction

The term social media has taken the Internet by storm since more than a decade. It is the fastest growing technology since its inception. It is a medium which allows interaction. Web 2.0 is an example of social media. Other widely used social media are the Facebook, Twitter, Blogs, Wikis, LinkedIn, and Instagram. The concept of social media is top of the agenda for many business executives today [1]. The general concept of our 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 a 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. We propose NetSpam framework that is a novel network-based approach which models review networks as heterogeneous information networks. The classification step uses different metapath types which are innovative in the spam detection domain. 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.

NetSpam 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 result in detecting fake reviews easier with less time complexity.

2. Related work

A. Literature review

In recent years, review spam detection has received significant attention in both business and academia due to the potential impact fake reviews can have on consumer behaviour and purchasing decisions. This survey covers machine learning techniques and approaches that have been proposed for the detection of online spam reviews. Supervised learning is the most frequent machine learning approach for performing review spam detection; however, obtaining labelled reviews for training is difficult and manual identification of fake reviews has poor accuracy. This has led to many experiments using synthetic or small datasets. Features extracted from review text (e.g., bag of words, POS tags) are often used to train spam detection classifiers. An alternative approach is to extract features related to the metadata of the review, or features associated with the behaviour of users who write the reviews. Disparities in performance of classifiers on different datasets may indicate the review spam detection may benefit from additional cross domain experiments to help develop more robust classifiers. Multiple experiments have shown that incorporating multiple types of features can result in higher classifier performance than using any single type of feature.

One of the most notable observations of current research is that experiments should use real world data if possible. Despite being used in many studies; synthetic or artificially generated datasets have been shown to give a poor indication of performance on real word data. As it is difficult to procure accurately labelled real-world datasets, unsupervised and semi-supervised methods are of interest. While unsupervised and semi-supervised methods are currently unable to match the
performance of supervised learning methods, research is limited and results are inconclusive, warranting further investigation. A possibility for a less labour-intensive means of generating labelled training data is to find and label duplicate reviews as Multiple studies have shown duplication, or near duplication, of review content is a strong indicator of review spam. Another data related concern is that real world data may be highly class imbalanced, as there are currently many more truthful than fake reviews online. This could be addressed through data sampling and ensemble learning techniques. A final concern related to quality of data is the presence of noise, particularly class noise due to mislabelled instances. Ensemble methods, and experiments with different levels of class noise, could be used to evaluate the impact of noise on performance and how its effects may be reduced.

As review text is an important source of information and tens of thousands of text features can easily be generated based on this text, high dimensionality can be an issue. Additionally, millions of reviews are available to be used to train classifiers, and training classifiers from a large, highly dimensional dataset is computationally expensive and potentially impractical. Despite this, feature selection techniques have received little attention. Many experiments have avoided the issue by extracting only a small number of features, avoiding the use of n-grams, or by limiting the number of features through alternative means such as using term frequencies to determine what n-grams are included as features. Further work needs to be conducted to establish how many features are required and what types of features are the most beneficial. Feature selection should not be considered optional when training a classifier in a big data domain with potential for high feature dimensionality. Additionally, we could find no studies that incorporated distributed or streaming implementations for learning from Big Data into their spam detection frameworks.

3. Problem statement

The profile data in social networks consist of two main parts, static and dynamic. Former is about the information which is set by the user statically, while the latter is observed by the system and is the result of users’ activity on the social network. The static data typically includes users’ demographics and interests, and dynamic data relates to user activities and position in the social network. Most of the existing research solutions depend on both static and dynamic data, which is inapplicable to other social networks, where it has merely a smaller number of visible static profiles and no dynamic profile details to the public. Due to its privacy policies and very restricted information visibility, none of the existing practical and theoretical means of fake profile detections are feasible to apply. Therefore, in this research our goal is to identify an approach to determine the spammers and fake profiles in Social Networks.

4. Proposed method

The user can login using the username and password, if the new User wants access then he needs to register and login. After user login the user has My Profile to view his or her profile and can search friends and friend requests. The user can search any product and view the details of the product. After viewing the product details, the user has access to post his review regarding or describing the product. The review of the product viewed can be positive or Negative. Admin has to Authorize the User and can view the post reviews. If he reports the negative reviews and the user is identified as a spammer and thus the processing happens.

Inputs:
The users input the details of the following details.
The user logsins using his details for the profile information and to have access to the database.

Processing:
The Database is created and updated using MySQL the table of the user, admin and others are stored.

Output:
The admin encounters negative reviews and reports spammers.

5. System architecture

A. Admin Module

In this module, the admin has to login by using valid user name and password. After login successful he can do some operations such as adding Categories, Adding Products for that Categories, Viewing and authorizing users, View Spam accounts details, viewing friend request & response, all recommended posts, all posts with all reviews, all positive and negative reviews, removing products, viewing all purchased

6. System analysis

A. Admin Module

In this module, the admin has to login by using valid user name and password. After login successful he can do some operations such as adding Categories, Adding Products for that Categories, Viewing and authorizing users, View Spam accounts details, viewing friend request & response, all recommended posts, all posts with all reviews, all positive and negative reviews, removing products, viewing all purchased
products, viewing positive and negative reviews chart on products.

The admin adds the category details such as category name. These details will be stored into the database. He adds Product posts for categories which include details such as, product image, product name, cost, description and uses of that product. These details will be stored into the database. These details will be further searched and accessed by the users in order to recommend to their friends and to buy products.

In the users module, the admin can view the list of users who all registered. In this, the admin can view the users’ details such as, user name, email, address, phone number and authorize the users. The admin can view all the friend requests and responses. Here all the requests and responses will be displayed with their tags such as Id, requested user image, requested user name, username request to, status and time & date. If the user accepts the request then the status will change to accepted or else the status will remain as waiting. The admin can view all the recommended products. If any recommendations happened for particular products, those details will be shown along with products. Details include product name, recommended user name, user recommended to name and the date.

In this, the admin can view all posts with their Positive and Negative Comments posted by users based on their opinions.

Positive: If the user comment contains at least one of the words which is listed in positive words, then that comment will be treated as a positive comment.

Negative: If the user comment contains at least one of the words which is listed in negative words, then that comment will be treated as a negative comment.

The comments of all posts will be displayed. Comments include Positive, Negative, Non-Positive and Non-Negative. It includes details such as, commented user name, comment and date, the products which are purchased by users will be displayed. It includes details such as, purchased user name, purchased products, price of the products and the date of purchase. The number of positive Reviews got by the particular product will be displayed in a chart. The number of negative Reviews got by the particular product will be displayed in a chart. In this module, the products which have got the negative comments from more than five users will be listed and removed by the admin.

B. User module

In this module, there are n number of users present. Users should register before doing any operations. Once a user registers, their details will be stored to the database. After registration is successful, he has to login by using authorized user name and password. Once login is successful user will do some operations like viewing their profile account details like spam or normal, search users and send friend requests, viewing friend requests, searching posts and recommendations to friends and viewing all product recommendations sent to him by his friends, commenting on posts, purchasing products and viewing their product search history.

The user can search the users based on names and the server will give responses to the user like User name, user image, Email id, phone number and date of birth. If you wish to send a friend request to a particular user then click on the “request” button, then request will be sent to that particular user.

In this, the user searches for products based on the products description. The user can recommend searched products to his friends, comment on posts and he can add the products to cart to buy those added products later by using their created account.

The user can view the friend requests which are sent by other users. Which includes sending user details with their tags such as user name, user image, date of birth, Email ID, phone number and Address and user can accept the request by clicking on the “waiting” link. The user can view all the products which are recommended by his friends.

This includes recommended user name and his image, recommended products details.

He can view all the searched products names and categories, the keywords which he used to search the products. This includes details such as, searched product, used keyword and date of search. He can create his bank account by providing details such as, account number, branch, address, email id. Later he can add money to his account and can view his account details.

7. Limitations of existing work

1. 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’ opinions. These misleading reviews are then multiplied by the sharing function of social media and propagation over the web.

2. Previous works also aimed to address the importance of features mainly in term of obtain accuracy, but not as a build-in function in their framework (i.e., their approach is dependent to ground truth for determining each feature importance).

3. Many aspects have been missed or remained unsolved.

8. Conclusion

In this report, we have contemplated a solution for detecting spammers and fake reviews on Online Social Networks (OSNs). This study introduces a novel framework namely Nestsmap based approach Addressing the problem of spam detection in such networks can be considered as a new research line in this field. Thus, the review of techniques is performed to detect spammers on social networks. False news identification on social media networks is an issue that needs to be explored because of the serious repercussions of such news at individual as well as collective level. Another associated topic that is worth investigating is the identification of rumor sources on social media. Although a few studies based on statistical methods have already been conducted to detect the sources of rumors, more sophisticated approaches, e.g., social network-
based approaches, can be applied because of their proven effectiveness.

References