Review Spammer Detection System by Using Rating and Review Content Similarity

Chan Myae Aye
Universities of Computer Studies, Yangon.
cmave84@gmail.com

Abstract

Assessing the trustworthiness of reviews is a key issue for the maintainers of opinion such as Amazon.com. Opinion reviews on products and services are used by potential customers before deciding to purchase a product. An important issue that has been neglected so far is opinion spam or trustworthiness of online opinions. To the best of our knowledge, there is still little published study on this topic, although Web spam and email spam have been investigated extensively. This paper presents spamer detection techniques to calculate spam score of each user based on their rating and review similarity on the target products. The experiment showed that the presented technique has comparatively spammer detection with less computation than others.

Key words: Spam detection, rating, similarity

1. Introduction

Product reviews are an increasingly important type of user-generated content as they provide a valuable source of information to help customers make good purchasing decisions [13]. Typically, these reviews consist of an overall product score (often in the form of a star-rating) and some free-form review text to allow the reviewer to describe their experience with the product or service in question. Web user can post products reviews at merchant sites to express their views and interact with other users via blogs and forums. It is now well recognized that the user generated content contains valuable information that can be exploited for many applications [3, 12].

In the past few years, there was a growing interest in mining opinions in reviews from both academia and industry. The existing work has been mainly focused on extracting and summarizing opinions from reviews using natural language processing and data mining techniques [1, 4, 11, 12 and 15]. Opinions trustworthiness is not studied till now in reviews. Due to the fact that the quality is not control, anyone can write anything on the Web. This results in many low quality reviews, and worse still review spam.

Review spam is similar to Web page spam. In the context of Web search, due to the economic and/or publicity value of the rank position of a page returned by a search engine, Web page spam is widespread. [5, 6, 7, 16, 18, 19] Web page spam refers to the use of “illegitimate means” to boost the rank positions of some target pages in search engines [2, 19]. In the context of reviews, the problem is similar, but also quite different. There are generally two types of spam reviews. The first type consists of those that deliberately mislead readers or automated opinion mining systems by undeserving positive opinions to some target products in order to promote them and/or by giving unjust or malicious negative reviews to some other products in order to damage their reputation. The second type consists of non-reviews (e.g., ads) which contain no opinions on the product. Detecting review spam is a challenging task as no one knows exactly the amount of spam in existence. Due to the openness of product review sites, spammers can pose as different users (known as sock puppeting”) contributing spammed reviews making them harder to eradicate completely. Spam reviews usually look perfectly normal until one compares them with other reviews of the same products to identify review comments not consistent with the latter. The efforts of additional comparisons by the users make the detection task tedious and non-trivial [10].

The system presents three techniques to detect review spam on the basis of spammer behavior that target on the products are presented in this paper. One should focus on detecting spammers based on their spamming behaviors instead of detecting spam reviews only. In fact, the more spamming behaviors the system can detect for a reviewer, the more likely the reviewer is a spammer. Subsequently, spammers’ review can be removed to accelerate the interests of other review users.

The rest of the paper is organized as follows. Section 2 covers some related works. Section 3 described rating based spam score, review based spam score and combined spam score. Evaluation experiment is described in Section 4 and Section 5 is devoted to conclusions.

2. Related Work

Review spam detection is a relatively new research problem which has not yet been well studied. A preliminary study was reported in [13]. A more in-depth investigation was given in [14].

A spam activities analyzing and a spam detection methods are presented in [14]. Three types of spam review such as untruthful opinion, review on brand only and non-review (e.g. question and answer
and random texts) are also discussed.

The scoring methods to measure the degree of spam for each reviewer is presented in [10] and applied on an Amazon review dataset. Then highly suspicious reviewers is selected for further scrutiny by user evaluators with the help of a web based spammer evaluation software specially developed for user evaluation experiments. The results showed that the ranking and supervised methods are effective in discovering spammers and outperform other baseline method based on helpfulness votes alone. To assign an overall numeric spam score to each user, the spam scores of the user's different spam behaviors are combined by using linear weighted combination. The weights on the different component spam scores can be empirically defined or learnt automatically.

A language modeling approach for consumer review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9].

Moreover, the Support Vector Machine also called SVM is effective for the detection of untruthful reviews. A simple and random texts) are also discussed.

The scoring methods to measure the degree of spam for each reviewer is presented in [10] and applied on an Amazon review dataset. Then highly suspicious reviewers is selected for further scrutiny by user evaluators with the help of a web based spammer evaluation software specially developed for user evaluation experiments. The results showed that the ranking and supervised methods are effective in discovering spammers and outperform other baseline method based on helpfulness votes alone. To assign an overall numeric spam score to each user, the spam scores of the user's different spamming behaviors are combined by using linear weighted combination. The weights on the different component spam scores can be empirically defined or learnt automatically.

A language modeling approach for consumer review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9].

Moreover, the Support Vector Machine also called SVM is effective for the detection of untruthful reviews. A simple and random texts) are also discussed.

The scoring methods to measure the degree of spam for each reviewer is presented in [10] and applied on an Amazon review dataset. Then highly suspicious reviewers is selected for further scrutiny by user evaluators with the help of a web based spammer evaluation software specially developed for user evaluation experiments. The results showed that the ranking and supervised methods are effective in discovering spammers and outperform other baseline method based on helpfulness votes alone. To assign an overall numeric spam score to each user, the spam scores of the user's different spamming behaviors are combined by using linear weighted combination. The weights on the different component spam scores can be empirically defined or learnt automatically.

A language modeling approach for consumer review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9]. They showed that Kullback-Leibler Divergence and the probabilistic review spam detection is presented in [9].

3. Review Spammer Detection System

Spammer usually monitors the targeted products closely and mitigates rating when time is appropriate. Most of the system observes spamming behaviors of each reviewer at first and then derive their respective spam scores. In this paper, three spammer detection techniques such as rating based detection, review based detection and their combination for targeting product is proposed.

Let \( U \) be set of users \( u_1, u_2, \ldots u_i \) and \( E_{ij} \) be set of rating \( e_{ij} \) from user \( u_i \) to product and \( R_{ij} \) be a set of reviews \( r_{ij} \) from user \( u_i \) to product \( o_j \).

The spammer behaviors are different from others. These spammer behaviors can be defined by the spam score. Three spam score techniques based on rating, review and their combination are presented in the next three subsections.

### 3.1 Targeting Products

Spammer promote or victimize a few products or product line and ratings they give to the product are quite difference from others. Reasonable user is expected to give ratings similar to other users on the same product. It's also importance to look at how reviews are equal to another review by the same user. It needs to check spammer behavior on both of their ratings and reviews. Thus, the system is presented spammer detection techniques on both ratings and reviews. The next subsections described about spammer detection techniques based on rating and review.

### 3.2 Rating based Spam Score

Rating based spam score is the users’ spam score to the targeting products by their rating behaviors. Let \( S_{r,i} \) be the rating based spam score of user \( u_i \) and is calculated as:

\[
S_{r,i}(u_i) = \frac{s_i}{\max_{u'_i \in U} s_i}
\]

where \( s_i \) is unnormalized spam score of user \( u_i \) and \( s_i \) can be calculated by,

\[
s_i = \sum_{e_{ij} \in E_{ij}, \mid E_{ij} \mid > 1} \mid E_{ij} \mid \cdot sim(E_{ij})
\]

The function \( sim(\cdot) \) is a similarity function comparing rating in a given set and is defined as:

\[
sim(E_{ij}) = 1 - \text{avg}_{e_{ij} \in E_{ij}} e_{ij}
\]

Based on the above spam score function, reviewers with large proportions of ratings on products are to be assigned high spam scores. By averaging ratings to calculate similarity of the rating, the system finds that it is less computation than the similarity function proposed in [10].

### 3.2 Review based Spam Score

As a user spam a product with multiple ratings, he is also likely to spam the product with multiple review texts. Such review texts are likely to be identical or look similar so as to conserve spamming efforts. The system use cosine similarity computation function to know content similarity of reviews.
sim(r_1, r_2) = \cos (\theta_1, \theta_2) \quad (4)

In this equation, r_1 and r_2 mean two review from set of reviews R, that user u write to the product. The system derives a similarity score sim(·) as follows:

\[
sim(R_{ij}) = \text{avg}_{i \neq j \mid R_{ij} > 1} \text{sim}(r_i, r_j) \quad (5)
\]

Then, define the user spam score function \(S_{\text{c},i}(u_i)\) for user \(u_i\) based on review spamming behavior on some targeted products:

\[
S_{\text{c},i}(u_i) = \frac{s_i}{\max_{u_i \in U} s_i'}
\]

where \(s'_i\) can be defined by,

\[
s'_i = \sum_{i \neq j \mid R_{ij} > 1} |R_{ij}| \cdot \text{sim}(R_{ij}) \quad (7)
\]

3.3 Combined Spam Score

Combined spam score is the users' spam score to the targeting products by their rating ad reviews spamming behaviors. Let \(S_{\text{c},i}\) be the required rating and review based spam score of user \(u_i\) and is defined by:

\[
S_{\text{c},i}(u_i) = w_r S_{r,i} + w_p S_{p,i}
\]

where \(w_r\) and \(w_p\) are the coefficient of rating based spam score and review based spam score respectively.

4. Evaluation

The presented methods are evaluated by using data from amazon.com. The reason for using this data set (http://131.193.40.52/data) is that it is large and covers a very wide range of products. Amazon.com is considered one of the most successful e-commerce Web sites with a relatively long history. This dataset gives the information such as Product ID, Reviewer ID, Rating, Date, Review Title, Review Body, Number of Helpful Feedbacks and Number of feedbacks. The characteristics of the data are presented in the following table.

Table 1. Dataset statistics

<table>
<thead>
<tr>
<th></th>
<th>Number (before preprocessing)</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>set of users</td>
</tr>
<tr>
<td>O</td>
<td>set of objects</td>
</tr>
<tr>
<td>V</td>
<td>set of reviews</td>
</tr>
<tr>
<td>E</td>
<td>set of ratings</td>
</tr>
</tbody>
</table>

In this paper, the results of the top twenty spammers by using rating based spam score are presented in table 2. The results of top twenty users by using [10] are also presented. 90% of spammers are the same with the similarity function used in rating based spam score. Only two spammers among twenty spammers are different and the first two spammer who is the most likely to be spammer is same in both methods.

Table 2. User ID and their respective spam scores

<table>
<thead>
<tr>
<th>Other Method</th>
<th>Proposed Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>User ID</td>
<td>Spam Score</td>
</tr>
<tr>
<td>A1345V2K5MV27</td>
<td>1</td>
</tr>
<tr>
<td>A12D2F4GPRZ27E</td>
<td>0.481574539</td>
</tr>
<tr>
<td>A108A4DHNV59M</td>
<td>0.312674484</td>
</tr>
<tr>
<td>A1085SODIDJHJEK</td>
<td>0.281092695</td>
</tr>
<tr>
<td>A16QODIDJHJEKU1</td>
<td>0.242276404</td>
</tr>
<tr>
<td>A108AVFSZ2U94NY</td>
<td>0.226430653</td>
</tr>
<tr>
<td>A105G2WMJTPDAD2</td>
<td>0.190536013</td>
</tr>
<tr>
<td>A108HDIRZ2U94NY</td>
<td>0.186228892</td>
</tr>
<tr>
<td>A12D2QZKRRKTNFSE</td>
<td>0.135545865</td>
</tr>
<tr>
<td>A10L4KPG2J94T</td>
<td>0.132597777</td>
</tr>
<tr>
<td>A14SNKOVV2ZIT</td>
<td>0.118653272</td>
</tr>
<tr>
<td>A108HE75T5SKF1</td>
<td>0.093752094</td>
</tr>
<tr>
<td>A10DNCYK7Y3SHU</td>
<td>0.081760476</td>
</tr>
<tr>
<td>A104UEUTQ4SPK1</td>
<td>0.029313232</td>
</tr>
<tr>
<td>A10VCFZCHW82X1</td>
<td>0.229737365</td>
</tr>
<tr>
<td>A1SL8JOZ938YV</td>
<td>0.222681919</td>
</tr>
<tr>
<td>A10N8E42J6WZ8X1</td>
<td>0.022125628</td>
</tr>
<tr>
<td>A10E139NRMGG3M</td>
<td>0.022125628</td>
</tr>
<tr>
<td>A108A86E67Y6F3X</td>
<td>0.016780419</td>
</tr>
<tr>
<td>A105GWGM7TPDAD2</td>
<td>0.022125628</td>
</tr>
<tr>
<td>A108A86E67Y6F3X</td>
<td>0.016780419</td>
</tr>
</tbody>
</table>

5. Conclusion

This paper presents review spammer detection methods on target products. The system has studied spammer behaviors and detects this spammer by their rating behaviors which give difference from other. It is presented to calculate spam score by using rating similarity, review content similarity and combine these spam score to get effective detection system. The current studies are only investigation state and need to test and prove how much these methods can detect correctly to spammer than other methods.

References


