Review Spammer Detection System by Using Rating and Review Content Similarity

Chan Myae Aye
Universities of Computer Studies, Yangon.
cmaye84@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 spammer 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]. One approach taken by review site such as Amazon.com is to allow users to label or vote the reviews as helpful or not. Unfortunately, this still demands user efforts and is subject to abuse by spammers.

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 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 language modeling based computational model is effective for the detection of untruthful reviews. Moreover, the Support Vector Machine also called SVM -based method is also effective for the detection of non-reviews. The empirical study found that around 2% of the consumer reviews posted to a large e-Commerce Web site is spam.

Methods that take advantage of the phenomenon of review plagiarism to control for the effects of text in opinion evaluation and a simple and natural mathematical model are presented in [8]. The analysis also allows to distinguish among the predictions of competing theories from sociology and social psychology, and to discover unexpected differences in the collective opinion-evaluation behavior of user populations from different countries. The study is however conducted at the collection level and does not provide evidence to link between spam and helpfulness votes [8].

In this paper the system presents a method to detect spammer using rating behavior by averaging rating they give to the products. Rating similarity is also importance to detect spammer because some rating of the spammer is same with the others rating. The system can calculate and guess users' spam score by this rating similarity by averaging rating. And then the system will calculate the similarity of their review content because some spammers are lazy so that they copy and paste their review instead of writing new one. Spamming behaviors are complex and not easily captured. So many researches about review spam detection are required for improving web sites.

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… 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

$$S_{r,i}(u_i) = \frac{s_i}{\max_{u_i' \epsilon U} \ s_i'} \qquad (1)$$

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

$$s_i = \sum_{e_{ij} \in E_{ij}, |E_{ij}|>1} |E_{ij}| \ . \ \mathrm{sim}(E_{ij}) \qquad (2)$$

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

$$sim(E_{ij}) = 1\text{- } \mathrm{avg}_{e_{ij} \,\epsilon\, \mathrm{Eij}} \ e_{ij} \qquad (3)$$

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.

$$\mathrm{sim}(r_i, r_j) = \cos(r_i, r_j) \quad (4)$$

In this equation r_i amd r_j means two review from set of reviews R_i that user u_i write to the product. The system derives a similarity score sim ( ) as follows:

$$\mathrm{sim}(R_{ij}) = \mathrm{avg}_{r_i, r_j \epsilon R_{ij,} |R_{ij}|>1} \mathrm{sim}(r_i, r_j) \quad (5)$$

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

$$S_{v,i}(u_i) = \frac{s_i^{'}}{\max_{u_i^{'} \epsilon U} s_{i'}^{'}} \quad (6)$$

where $s_i^{'}$ can be defined by,

$$s_i^{'} = \sum_{r_{ij} \epsilon R_{ij}, |R_{ij}|>1} |R_{ij}| \,.\, \mathrm{sim}(R_{ij}) \quad (7)$$

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 $S_{c,i}$ be the required rating and review based spam score of user u_i and is defined by:

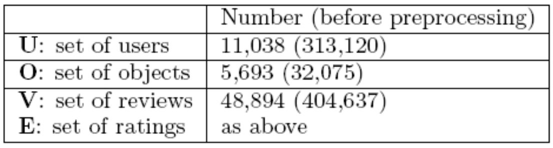
$$S_{c,i}(u_i) = w_r\, S_{r,i} + w_v\, S_{v,i} \quad (8)$$

where w_r and w_v 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



	Number (before preprocessing)
U: set of users	11,038 (313,120)
O: set of objects	5,693 (32,075)
V: set of reviews **E**: set of ratings	48,894 (404,637) as above

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

Other Method		Proposed Method	
User ID	Spam Score	User ID	Spam Score
A1345VRK5MYG7	1	A1345VRK5MYG7	1
A12DP14GPRZF7E	0.481574539	A12DP14GPRZF7E	0.708263069
A108AJDHNBV29M	0.312674484	A108H0RI2LWLUO	0.303541315
A105S56ODHGJEK	0.281092965	A16QODENBJVUI1	0.26306914
A16QODENBJVUI1	0.24427694	A1OL4KPWG2794I	0.19560000
A108AVF62U97NY	0.226130653	A105S56ODHGJEK	0.175379427
A105GWGM7PDAI2	0.190536013	A12DQZKRKTNF5E	0.175379427
A108H0RI2LWLUO	0.186522892	A108AVF62U97NY	0.134907251
A12DQZKRKTNF5E	0.153545505	A105GWGM7PDAI2	0.101180438
A1OL4KPWG2794I	0.132507777	A10UEU8TQU6PKI	0.101180438
A1143SNKOV0ZIT	0.113065327	A1143SNKOV0ZIT	0.101180438
A1087EHT8T5KFI	0.083752094	A10EBAP0TYPDTD	0.053962901
A10DNCYK7YISHU	0.041876047	A11LS12ZU93SV6	0.040472175
A10UEU8TQU6PKI	0.029313233	A1087EHT8T5KFI	0.031478359
A10Y8FHZU6B2X1	0.027917365	A108AJDHNBV29M	0.02698145
A11LS12ZU93SV6	0.027638191	A10EEPCU4SZO06	0.02698145
A1004AX2J2HXGL	0.025125628	A10Y8FHZU6B2X1	0.02698145
A10E139NRMGOMM	0.025125628	A102OZZ31NALGU	0.013490725
A10EBAP0TYPDTD	0.022333892	A10E139NRMGOMM	0.013490725
A1025UG0K6EF5X	0.016750419	A10DNCYK7YISHU	0.006745363

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

[1] A-M. Popescu and O. Etzioni. Extracting Product Features and Opinions from Reviews. *EMNLP'2005*.

[2] A. Ntoulas, M. Najork, M. Manasse & D. Fetterly. Detecting Spam Web Pages through Content Analysis. WWW'2006.

[3] B. Liu. Web Data Mining: Exploring hyperlinks, contents and usage data. Springer, 2007.

[4] B. Pang, L. Lee & S. Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. *EMNLP'2002*.

[5] B. Wu and B. D. Davison. Identifying link farm spam pages. *WWW'06*, 2006.

[6] B. Wu, V. Goel & B. D. Davison. Topical TrustRank: using topicality to combat Web spam. *WWW'2006*.

[7] C. Castillo, D. Donato, L. Becchetti, P. Boldi, S. Leonardi, M. Santini, S. Vigna. A reference collection for web spam, *SIGIR Forum'06*, 2006.

[8] C. Danescu-Niculescu-Mizil, G. Kossinets, J. Kleinberg, and L. Lee. How opinions are received by online communities: a case study on amazon.com helpfulness votes. In WWW, 2009.

[9] C.L. Lai, K.Q. Xu, Raymond Y.K. Lau ,Y. Li and L. Jing, A language modeling approach for consumer review spam detection, IEEE International Conference on E-Business Engineering , 2010.

[10] Ee-Peng Lim, Viet-An Nguyen, Nitin Jindal, Bing Liu, Hady W. Lauw. Detecting Product Review Spammers using Rating Behaviors, CIKM'10, Toronto, Ontario, Canada, October 26–30, 2010.

[11] K. Dave, S. Lawrence & D. Pennock. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. *WWW'2003*.

[12] M. Hu & B. Liu. Mining and summarizing customer reviews. KDD'2004.

[13] N. Jindal and B. Liu. Review spam detection. In WWW (poster), 2007.

[14] Nitin Jindal and Bing Liu . Opinion Spam and Analysis. WSDM'08, Palo Alto, California, USA, February 11-12, 2008.

[15] P.Turney. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *ACL'2002*.

[16] R. Baeza-Yates, C. Castillo & V. Lopez. PageRank increase under different collusion topologies. *AIRWeb'05*, 2005.

[17] Soo-Min Kim, Patrick Pantel, Tim Chklovski, Marco Pennacchiotti. Automatically Assessing Review Helpfulness. Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP 2006), pages 423–430, Sydney, July 2006.

[18] Y. Wang, M. Ma, Y. Niu, H. Chen. Spam Double-Funnel: Connecting Web Spammers with Advertisers. *WWW2007*.

[19] Z. Gyongyi & H. Garcia-Molina. *Web Spam Taxonomy*. Technical Report, Stanford University, 2004.