Detecting Product Review Spammers using Activity Model

Bo Jiang, Renhao Cao and Bi Chen
School of Computer Science and Information Engineering, Zhejiang Gongshang University, Hangzhou 310018, China
nancybjiang@mail.zjgsu.edu.cn, caorenhao0213@163.com, cb.chenbi@gmail.com

Abstract - This paper aims to improve the accuracy of the original detection model on the product review spammers. The original detection model has three activities of review spammers. Now, we add the new two activities to improve the accuracy. In this paper, we first introduce the three existing models. Then, we give our new models. At last, we propose scoring methods to ensure that the new activity models have the same effects on the prediction model based on an Amazon review dataset. Our results show that our proposed two models have achieved the desired result.

Index Terms - Spam reviewers, activity.

I. Introduction

Nowadays, the e-commerce occupy a more and more important position in the economic, the online sales grows by leaps and bounds every year. At the same time, people's attention has increasingly focused on the online shopping environment. Reference [1] shows that due to the internet's openness, there is no limit on the reviews. So, when people buy the products on the web site, they don't know which review they can believe. From Reference [2], we know that these review spams have caused very bad influence to the e-commerce shopping environment. Thus, how to detect has been a hot research problem.

Reference [3] shows review spams are divided into two broad categories: one is the positive comments for the merchants sell products and the other one is the negative ones for the merchants slander competitors. Some of businessmen hire a lot of people making many positive reviews to lure users buy their products for improving their sales. And also someone hire people publishing many negative reviews to damage the reputation of competitors and lure users cancel the plan. The users can't touch the products, so they make the shopping decisions only by the reviews. The merchants use the weak point to cheat them. Then most of people lose the confidence and the online shopping environment is becoming a pandemonium place because of the review spams.

Now the research has focused on two directions: one is analyzing the review texts, one is analyzing the reviewers.

Reference [5] shows 10 to 15% of reviews are influenced by the previous review spams. It was found that there was almost no difference between the review spams and the normal reviews in [6]. Therefore the method based on content or opinion extracting to identify review spams will be very difficult. Reference [7] found there was a great relationship between the deviation of ratings for products and the usefulness of reviews. So, we can detect the spams by finding the review spammers from the perspective of user rating behavior.

This paper builds the detection model to find the review spammers through the analysis of user behavior. On the basis of reference [8], we add two models for the Chinese users' behavior to improve the correct and adaptive of the detection model.

II. Behavior Analysis and Symbol Definition

A . *Behavior Analysis*

The reviewers' all behavior of reviews, ratings, buying reflects their specific purpose. We assume the spams have the following five kinds of behaviors patterns by analysis the reviewers' behaviors characteristics in the shopping web site: The three kinds of existing one: (1)The pattern that the same user publish several reviews and ratings on the same product; (2)The pattern that the same user publish several reviews and rating on the same product group; (3)The pattern that the rating has a deviation from other reviewers; The two kinds that this paper has coming up with:(1)The pattern that a user would continuous publish some reviews in a short time; (2)The pattern that the number of purchase is far less than the comments. We can build models according to the different behavior patterns. We use the score as a standard to measure the spams in every model. The user who gets a high score may be the review spammer.

B . *Symbol Definition*

We define the following symbol for the paper:

$O = \{o_i\}$: set of reviewers;

$C = \{c_j\}$: set of comment objects;

$T = \{t_k\}$: set of texts for reviews;

$G = \{g_k\}$: set of grades for products;

$P_a = \{p_a\}$: set of products have buy;

$T_a = \{t_a\}$: set of texts for the products have bought;

$T_{ij} = \{t_k | o(t_k) = o_i \wedge c(t_k) = c_j\}$: set of texts for user o_i to product c_i ;

$G_{ij} = \{g_k | o(g_k) = c(g_k) = c_j\}$: set of grades for user o_i to product c_i ;

$G_{i*} = \cup_j G_{ij}$: set of all grades;

$G_{*j} = \cup_i G_{ij}$: set of all grades for product c_i ;

III . Detection Model

A . Target Products

The review spammers would have rating on one product many times, and the scores are very likely. Or he has published reviews several times, and the texts are most likely. So, we can find the spam one by counting the times of rating and commenting.

(1) Rating Detection

$$C_{p,e}\left(o_i\right) = \frac{s_i}{Max_{o_i^{'} \in U} s_{i^{'}}} \cdot \qquad (1)$$

where $s_i = \sum_{e_{ij} \in G_{ij,|G_{ij}|\rangle 1}} \left|G_{ij}\right| \bullet sim\left(G_{ij}\right)$.The similarity function $sim(\ )$ can comparing ratings in a given set and is defined as:

$$sim\left(G_{ij}\right) = 1 - Avg_{g_k, g_{k^{'}} \in E_{ij}, k<k^{'}} \left|g_k - g_{k^{'}}\right| \cdot \qquad (2)$$

(2) Review Text Detection

Because of the most of reviews are short one, also full of pinyin, simplified Chinese characters and traditional Chinese characters. We put the title and the comment text together as the comment text to making word segmentation. Then we extract pinyin string from every words, and changing the texts to the vectors according every texts' pinyin string. We define the text similarity between two reviews t_k and $t_{k^{'}}$:

$$sim\left(t_k, t_{k^{'}}\right) = \cos\left(t_k, t_{k^{'}}\right). \qquad (3)$$

where $\cos\left(t_k, t_{k^{'}}\right)$ represents the cosine similarity of the TF-IDF vectors of Because of t_k and $t_{k^{'}}$. Given a set of reviews $T_{ij}(T_{ij}\rangle 1)$, we can define a similarity score for the review texts :

$$sim\left(T_{ij}\right) = Avg_{t_k, t_{k^{'}} \in T_{ij}} sim\left(t_k, t_{k^{'}}\right). \qquad (4)$$

Thus, we can define the review spammer's score based on the review texts as:

$$c_{p,t}\left(o_i\right) = \frac{S_i^{'}}{Max_{o_i^{'} \in 0} S_i^{'}} \cdot \qquad (5)$$

where $S_i^{'} = \sum_{t_{ij} \in T_{ij,|T_{ij}|>1}} \left|T_{ij}\right| \cdot sim\left(T_{ij}\right) \cdot$

(3) Composite Scores

According to the above two models, we define the review spammer's score based on the target products as：

$$c_p\left(o_i\right) = \frac{1}{2}\left(c_{p,e}\left(o_i\right) + c_{p,t}\left(o_i\right)\right). \qquad (6)$$

B . Target Products Group

The products group is a range of products have the same property. A products group may be a publisher or a brand in the amazon's review sets. The review spammer would give rating or comment on a certain brand's products in a short time. The rating may be very high or low. Thus, we define two detection models based on the product group. One is the high rating, and another is the low one.

(1) High Rating Model

To model rating behavior that involves very high ratings on products sharing the same attribute by the same user within a short span of time, we divided time into a continuous fixed size of time windows, and made the clustering analysis according to high grade. We define high rating cluster that the user o_i to a product group b_k in a time window W as:

$$G_{ik}^{H}\left(W\right) = \left\{g_{ij} \in G_{i*} \middle| c_j \in b_k \wedge t\left(g_{ij}\right) \in W \wedge g_{ij} \in HScore\right\} \qquad (7)$$

where $HSore$ represents a set of high ratings. At amazon site, a score of 5 is the high score, so we set up $HSore = \{1\}$.

We think only sufficiently large $G_{ik}^{H}\left(W\right)$ can be attributed to the spams, then we define $\min^{H}$ as the minimum threshold of $G_{ik}^{H}\left(W\right)$. Only the groupings that greater than $\min^{H}$ can be counting in the review spammers' score. The time window W should be greater than the minimum size of time, and it can capture the user's large ratings on a certain product group in a short time. We set the time window W to the day granularity and $\min^{H} = 3$ according to the experience. Then we can get those groupings' score:

$$c_i^{H} = \cup_{k,w} \left\{G_{ik}^{H}\left(W\right) \middle\| G_{ik}^{H}\left(W\right)\right| \geq \min^{H} \right\} \cdot \qquad (8)$$

Thus, we can define the review spammers' score based on the high rating of product group as:

$$c_{g,H}\left(o_i\right) = \frac{c_i^{H}}{Max_{o_i^{'} \in \cup o_i^{H}}} \cdot \qquad (9)$$

(2) Low Rating Model

The same as the above high rating model, we define low rating cluster that the user o_i to a product group b_k in a time window W as:

$$G_{ik}^{L}\left(W\right) = \left\{g_{ij} \in G_{i*} \middle| c_j \in b_k \wedge t\left(g_{ij}\right) \in W \wedge g_{ij} \in LScore\right\} \qquad (10)$$

where $LSore$ represents a set of low ratings. At amazon site, a score of 1 or 2 is the low score, so we set up $LSore = \{0, 0.25\}$.

Thus, the rating set that the user o_i conforms to the minimum threshold can be defined as:

$$c_i^{L} = \cup_{k,w} \left\{G_{ik}^{L}\left(W\right) \middle\| G_{ik}^{L}\left(W\right)\right| \geq \min^{L} \right\} \qquad (11)$$

We set the time window W to the day granularity and $\min^{L} = 2$ according to the experience. Because of the number of users give the score of 5 is more than the score of 1 or 2, we set the number of $\min^{L}$ is less than the $\min^{H}$ one.

Thus, we can define the review spammers' score based on the low rating of product group as:

$$c_{g,L}\left(o_i\right) = \frac{c_i^L}{Max_{o_i^{'} \in \cup o_i^L}}. \qquad (12)$$

(3) Composite Scores

According to the above two models, we define the review spammers' score based on the target products group as:

$$c_g\left(o_i\right) = \frac{1}{2}\left(c_{g,H}\left(o_i\right) + c_{g,L}\left(o_i\right)\right). \qquad (13)$$

C . Score Deviation

The review spammers would give a score with a deviation from other reviewers to promoting or demoting a product. Then we define the deviation of a score d_{ij} as its difference from the other's average score on a product:

$$d_{ij} = g_{ij} - Avg_{g \in G_{*j}} g. \qquad (14)$$

Thus, the spam score based on the deviation can be defined as:

$$c_d\left(o_i\right) = Avg_{g_{ij} \in G_{*j}} \left|d_{ij}\right|. \qquad (15)$$

D . Review Frequency

In order to finishing his work, the review spammers would continuous publish some reviews in a short time. Then we use the frequency of the reviews as indicators to measure the behavior of publishing review spammers.

And it has two forms: one of that is the reviews are all published in one month. And another one is the reviews are published over a period of time, but every month have a large dissimilar on the frequency. We define $c_r\left(o_i\right)$ as the spam score based on the frequency of the reviews.

(1) All the reviews are published in a short time

If the number of reviews is more than five in one month, then $c_r\left(o_i\right)$ will be 1. And if the number less than five, then $c_r\left(o_i\right)$ will be 0.5.

(2) All the reviews are published in a long time

$$c_r\left(o_i\right) = \frac{\left|t_i - Avg_{t_{ij} \in T_{*j}} t\right|}{\sum t_{ij}}. \qquad (16)$$

where $Avg_{t_{ij} \in T_{*j}} t$ represents the average of every month's reviews, and $\sum t_{ij}$ represents the total reviews.

E . Purchasing Behavior

The review spammers would buy only a few products what he have commented or never once. Then we define $c_b\left(o_i\right)$ as the spam score based on the purchasing behavior:

$$c_b\left(o_i\right) = 1 - \frac{b_i}{Max_{o_i^{'} \in O^{ri^{'}}}}. \qquad (17)$$

where $b_i = \frac{\left|G_b\right|}{\left|G\right|}$ represents the average purchasing time on publishing one review.

IV . Evaluation

A . Data Set

We used the dataset Amazon Product Review Data (Huge) from www.datatang.com. We choose 10 020 reviews as the data set. It had 291 reviewers, 9 384 products and 1 221 product brands. Among reviews, It had 1 034 one more than 200 words, almost 10.34% in the total reviews, so the most reviews is the short one in the web site.

B . Model evaluation

At first, we should sure that which one is the review spammers, so that we can make the effectiveness evaluation for the proposed model. But the amazon doesn't have label, we could get the result by voting.

We can regard the vote as the standard, for each of the review, if it gets a valid vote then its total votes could plus one; in the opposite it gets a invalid vote then its total valid one could plus zero and the total one plus one. We can count the review spammers' score by using the total valid votes devided by the total votes.

First, we can choose a small collection data set to voting. In order to determine the proposed five models' validity, we count the users' scores according to the above models, then we make a descending sorting. We pick out the top ten users who maybe the review spammers and the last ten users who maybe not in each of the models, then we put all the users into one data set.

Second, we use three people to tag the dataset by manual. The process is independent, each people don't know others' result, and we record the results.

According the tag result, we use NDCG (Nor-malized Discounted Cumulative Gain) [9] evaluate the method. We define the DCG as:

$$D_{DCG} = \sum_{p=1}^{50} \frac{2^{f(i_p)} - 1}{\log_2\left(1 + p\right)}. \qquad (18)$$

where $f\left(i_p\right)$ represents the votes of the user $o\left(i_p\right)$.

Then we can define N_{NDCG} as :

$$N_{NDCG} = \frac{D_{DCG}}{DCG_{idez1}}. \qquad (19)$$

where DCG_{ideal} represents the DCG in the ideal state. We can see which model closed to the ideal state.

C . Result

Table 1 and 2 shows the number of review spammers and normal reviewers labeled by the three people. The off-diagonal represents the same number of review spammers and normal reviewers labeled by the corresponding two people.

TABLE I Numbers of Review Spammers Labeled

	Tag people 1	Tag people 2	Tag people 3
Tag people 1	25	20	22
Tag people 2		22	20
Tag people 3			24

TABLE II Numbers of Normal Reviewers Labeled

	Tag people 1	Tag people 2	Tag people 3
Tag people 1	25	23	24
Tag people 2		28	25
Tag people 3			26

According to the principle of most of the votes, we can label the users which one is the review spammer on the basis of the labeled results. If a user is labeled a review spammer by more than two people includes two, we can consider the user will be the spam one. In the end, we have 26 review spammers. We pick out the top ten users who are the review spammers and the last ten users who are not. In table 3, BL is Base Line, P is Products, PG is Product Groups, SD is Score Deviation, RF is Review Frequency, PB is Purchasing Behavior. Table 3 shows that the original model based on the products group has the best result. Also we can see the new models based on the purchasing behavior is the same like it, and the other new models based on the review frequency is same like other original models.

TABLE III Numbers of Top 10 and Last 10

	BL	P	PG	SD	RF	PB	All
Top	6	4	8	3	4	6	10
Last	6	7	6	4	6	6	10

Then we count the result of N_{NDCG} on every model. Figure 1 shows the two models that we have suggested have a better result than the original one. Then we can know if the prediction model increases the two new models will add the accuracy greatly.

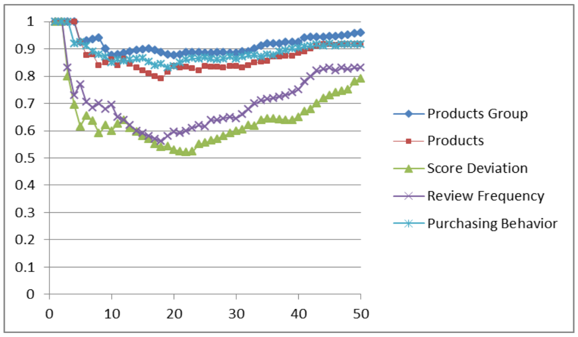

Fig. 1 NDCG Results

V. Conclusions

This paper improves the prediction model that using the behavioral approach to detect review spammers who try to manipulate review ratings on some target product or product groups. The result shows that the idea is successful. We add the two different indicators to avoid missing the spams. The detection to review spammer is a new subject. But there is only a little work on this research. So we should do more work to improve the level of this field. In future, we will increase the accuracy through a variety of ways.

References

[1] Theodoros Lappas.: Fake Reviews: The Malicious Perspective. Natural Language Processing and Information Systems Lecture Notes in Computer Science Volume 7337, 2012.

[2] Hankin, Lisa.: The effects of user reviews on online purchasing behavior across multiple product categories. PhD thesis (2007).

[3] Jindal N, Liu B. Analyzing and Detecting Review Spam. In: Proceeding of the 7th IEEE International Conference on Data Mining (ICDM' 07), Omaha, Nebraska, USA. Washington, DC, USA: IEEE Computer Society, 2007: 547-552.

[4] Jindal N, Liu B. Review Spam Detection. In: Proceeding of the 16th International Conference on World Wide Web, Banff, Alberta, Canada. New York, NY, USA: ACM, 2007:1189-1190.

[5] E. Gilbert and K. Karahalios.: Understanding Deja Reviewers. In CSCW, 2010.

[6] 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.

[7] Jindal, N., Liu, B.: Opinion spam and analysis. In: WSDM 2008. ACM, New York (2008)

[8] Lim E P, Nguyen V A, Jindal N, et al.: Detecting Product Review Spammers Using Rating Behaviors //Proc. of the 19th ACM International Conference on Information and Knowledge Management. New York, USA: [s. n.], 2010.

[9] P. Ravikumar, A. Tewari, E. Yang.: On NDCG Consistency of Listwise Ranking Methods. Proceedings of the 14th International Conference on Articial Intelligence and Statistics (AISTATS) 2011, Fort Lauderdale, FL, USA.