INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, e-ISSN 1841-9844, 14(6), 759-772, December 2019.

Identification of Opinion Spammers using Reviewer Reputation and Clustering Analysis

M.J. Zhong, L.Tan, X.L. Qu

Minjuan Zhong*
School of Information Technology
Hunan University of Finance and Economics
Changsha 410205, China
*Corresponding author: lucyzmj@sina.com

Liang Tan
School of Information Management
Jiangxi University of Finance and Economics
Nanchang 33013, China
ghosticy@foxmail.com

Xilong Qu
School of Information Technology
Hunan University of Finance and Economics
Changsha 410205, China
quxilong@126.com

Abstract: Online reviews have increasingly become a very important resource before making a purchasing decisions. Unfortunately, malicious sellers try to game the system by hiring a person or team (which is called spammers) to fabricate fake reviews to improve their reputation.Existing methods mainly take the problem as a general binary classification or focus on some heuristic rules. However, supervised learning methods relies heavily on a large number of labeled examples of deceptive and truthful opinions by domain experts, and most of features mentioned in the heuristic strategy ignore the characteristic of the group organization among spammers.

In this paper, an effective method of identifying opinion spammers is proposed. Firstly, suspected spammers are detected by means of unsupervised learning based on reviewer's reputation. We believe that the reviewer's reputation has a direct relation with the quality of reviews. Generally, review written by user with lower reputation, shows lower quality and higher possibility to be fake. Therefore, the model assigns reputation score to each reviewer wherein the content based factors and activeness of reviewers are employed efficiently. On basis of all suspected spammers, k-center clustering algorithm is performed to further spot the spammers based on the observation of burst of review release time.

Experimental results on Amazon's dataset are encouraging and indicate that our approach poses high accuracy and recall, and good performance is achieved.

Keywords: opinion spammer, fake review, reviewer reputation, clustering analysis.

1 Introduction

The rapid development of online shopping has led to a large number of user reviews for various products or services. More and more potential consumers tend to rely on them to assess

the quality of goods or services before buying. Therefore, driven by the desire for profit or competition, producer and retailers generate motivations and behaviors to manipulate reviews, maliciously posting deceptive or fake reviews to deliberately mislead potential consumers and make their risky purchasing decisions. Moreover, for the purpose of high satisfaction rate and reputation, manufactures may hire a person(known as an individual spammer) or team(known as spammer group) to post glamorized positive on their product or harmful negative reviews on their competitor as consumers. Hu Nan et al. [7]conducted research on the website reviews of two major US bookstores, Amazon.com and Bames & Noble, and found that a large number of comments on the website were manipulated by publishers and vendors; Michael Luca and Georgios Zervas analyzed [14] that 16% of the comments on the Yelp website are deceptive; China's e-commerce websites, such as Taobao and Jingdong Mall also acknowledge that there are a large number of fake reviews on their e-commerce platforms.

Studies have shown that the discrimination of such deceptive reviews by ordinary consumers is relatively low and the task of opinion spam or spammers detection is a very challenging problem. For example, consider figure 1 (a) and (b), which illustrates two reviews written by two users for a product from Amazon.

If we only look at the two reviewers individually, they all appear genuine. However, when we collect them together, we could find easily that both reviews are duplicated. From the publication time of the review, it seems likely that the first user copies the review made by the second user. Remarkably, there are ten other users to read and make feedback on this review written by the first user, and two people thought it was a useful review. Obviously, genuine users are unable to judge whether a review is a fake review or not, and whether a reviewer is a spammer or not.

[A2HP9COIPKXQ1I][August 20, 2003][2][10][1.0][Not as good as the Scorpions.] Yanni fans and others have come to expect from him "music" that provokes landscape images, religious symbolism, and cultural...ZZZZZ.ZZZZZZZZZZ. Anyway, the fact is his music has never been the same since his departure from german rock band, the Scorpions. His presence was felt on a few early recordings and fans may have noticed the difference upon its absence. Yanni has completely left behind all previous styles and taken on a new identity in the world music genre. The Album is over the top and over long and will definitely take you into another realm - one in which you may not want to be - So Yanni heads will appreciate it. The only really unique aspect of Acropolis is the out of place appearance of Rob Halford (Judas Priest). I would really like to know how he agreed to appear on it. The unknown Twisted Sister cover song is absent once again - It may not even exist on bootleg. Yanni fans will like this but if you know his metal roots its a little hard to take in.

(a) The first candidate fake reviews

[AOCJ3A12C91U5][May 23, 2001][1][7][1.0][Not as good as the Scorpions.] Yanni fans and others have come to expect from him "music" that provokes landscape images, religious symbolism, and cultural...ZZZZZ.ZZZZZZZZZZ. Anyway, the fact is his music has never been the same since his departure from german rock band, the Scorpions. His presence was felt on a few early recordings and fans may have noticed the difference upon its absence. Yanni has completely left behind all previous styles and taken on a new identity in the world music genre. The Album is over the top and over long and will definitely take you into another realm - one in which you may not want to be - So Yanni heads will appreciate it. The only really unique aspect of Acropolis is the out of place appearance of Rob Halford (Judas Priest). I would really like to know how he agreed to appear on it. The unknown Twisted Sister cover song is absent once again - It may not even exist on bootleg. Yanni fans will like this but if you know his metal roots its a little hard to take in.

(b) The second candidate fake reviews

Figure 1: Two product reviews from Amazon.com

In fact, the opinion spam and spammers are closely related.The reviews written by the

spammers are fake reviews to a large extent, and correspondingly, the author of a fake review must be a spammer. Therefore, many researchers have developed detection techniques combination the fake reviews with spammers together and solved them by means of binary classification problem(spam vs. non-spam, spammer vs non-spammer). As we all known, the classification algorithm is a supervised learning method that requires large sets of pre-labeling instances. However, the task of pre-labeling both classes, deceptive and truthful opinions is difficult and time-consuming. The biggest problem is the authenticity of the labeled opinions.

In this paper, we propose an unsupervised method for detecting opinion spammers. This research makes the following main contributions:

(1)It introduces the concept of reputation and proposes a reviewers' reputation model to detecting spammers. The model can capture the suspected spammers with low reputation value by analyzing both content-based characteristics(context similarity, opinion sentiment, review length and helpful feedback) and behavior-based characteristics (authors' activeness, review deviation and rating) without requiring a large sets of labeled instances.

(2)Clustering technique is further performed to spot accurately spammers. Generally, spammers will post their reviews intensively. This behavior shows the characteristics of abruptness in posting time interval. Consequently, users' posting time interval is proposed to measure the similarity and on basis of all suspected spammers, k-center clustering algorithm is implemented to filter spammers.

(3)We conduce the experiments on the Amazon review dataset. The experimental results are encouraging and indicate that our approach poses high accuracy and recall, and good performance is achieved.

The rest of the paper is organized as follows: the next section discuses some relevant literature on opinion spammers detection. Section 3 describes our user reputation model to the task of suspected spammers detection. Section 4 presents the clustering algorithm for the spammers spotting. Experimental results and analysis are discussed in section 5. Finally, section 6 presents our conclusions and discusses some future work directions.

2 Related work

Detection of opinion spam was first introduced by Jindal N. and Liu B. They conduced a series of studies for automatically detecting review spam [8], spammers [11] and spammer group [15] by machine learning, pattern recognition, graph theory and other techniques.Yuanchao [13] summarized previous detection approach as supervised or unsupervised learning.

Most existing supervised learning methods, which is mainly based on text content, regard the detection of opinion spam as classification process. Combing psychology and computational linguistics to extract the linguistic content cues of the reviews, supervised learning methods applied neural network [16], decision tree [4] and other classifiers to establish a statistical model and then predict the unknown reviews [6].

In text content, repeated review is considered to be important clue for the identification of fake review. In order to find duplicate and near duplicate reviews, Lin et al. [12] defined three similarities, including similarity of reviews written by the same reviewer, similarity between reviews in the target product, and similarity of reviews between different categories of products.

Lau [9] established language model to identify conceptual duplicate reviews.

However, this kind of opinion spam only represents a small percentage of the opinion spam. So, other linguistic features have also received much attention [2].

Rupesh et al. [3] proposed new lexical and syntactic features including type of punctuation mark, Part-of–Speech (POS) etc and applying supervised algorithms,such as SMO, Decision

Tree, and Naive Bayes for performing classification on fake reviews dataset. The final results give promising accuracy 91.51% for detecting fake reviews.

Banerjee and Chua [1] distinguished between deceptive and genuine reviews from following aspects: the readability of a review, the richness of describing the objective information in review, the writing style of reviews, and review genre. On basis of these features, they built logistic regression model to detect opinion spams.

Wen et al. [19] proposed a two-view collaborative filtering approach based on SVM benchmark to identify opinion spams. In their work, two different expressions are used for each review. One is the lexical terms derived from the textual content of the reviews and the other is the PCFG rules derived from a deep syntax analysis of the reviews. The statistical differences between deceptive and truthful opinions are observed from above two different perspectives. On basis of them, CoSpa-C and CoSpa-U strategies are proposed for classifier training to identify fake and true reviews.

The challenge for the existing supervised classification algorithms is that manually marking massive ground truth spam reviews is difficult and time-consuming. Therefore, another detection approach, unsupervised learning, has been proposed and tended to focus on the opinion spammers or group by relying on behavioral reviewers features. For example, Savage Z. et al [17] find that the reviews of opinion spammers tend to give extreme evaluation scores.

Vlad and Martin [18] conducted detection on Singletons reviewers in fake reviews, who registered multiple names deliberately, posting each review under a different name. They made an important assumption: singleton reviewers is lack of complete imagination and are not able to rewrite each review completely. They prefer to rephrasing, switching some synonyms, and keeping the sentiment consistency to all reviews.So, semantic similarity has been measured from the perspective of terms and topic distribution, and meanwhile singleton reviewers have been identified indirectly.

In Atefeh Heydari et al.'s work [5], suspicious time intervals are captured from the sudden rapid increase in the number of the reviews. The reviews in suspicious time are analyzed combining the rating deviation, content similarity and activeness of reviewers, which is beneficial to the time efficiency due to narrowing the detection range.

David S.et al. [**?**] used binomial regression to identify those reviewers with an abnormal rating for products. The proposed method stemmed from overarching assumption that the majority of reviews are posted by truthful reviewers. Therefore, the method does not pay attention to the text-based features, but focused merely on the rating differences between spammers and the majority of truthful reviewers.

With an in-depth study on features of fake reviews, Lijing et al. [10] introduced the concept of user-credibility and shop-credibility, and established a fake review identification model, which integrated the reviewer's behaviour characteristics, businesses characteristics and review texts.

In the above method, the supervised learning technique focuses on analyzing the review text, and uses the natural language processing technology to extract the part of speech, grammar, sentiment and other features of the review text to distinguish the authenticity of the content. However, this method heavily depends on a large number of labeled examples of fake and truthful opinions by domain experts, which is a time-consuming and costly endeavour. The unsupervised recognition method does not need to mark a large amount of training corpus, but the problem of low recognition rate is generally found. Furthermore, most of the heuristic features proposed in the existing methods are based on individual opinion spammer,ignoring the characteristics of group review spamming. Therefore, the current identification of opinion spam has not yet formed an effective solution.

3 Feature representation

For detecting the fake reviews, the first thing is defining the features. In this paper, we mainly considered two scopes of features to indicate spamming activities, one is user's review content features and one is user's behavioral features.

Suppose there is a review sequence $R_p = r[1], r[2], ..., r[n]$ for a product p, where $r[i]$ is a review sorted by time arrival and contains a variety of information, such as user name $r[i].u$, release time $r[i].t$, review rating $r[i].r$, review title $r[i].title$ and review content $r[i].content$ etc.

3.1 Review content features

Spammers usually consider the following factors while writing fake review: first, to generate a review as quickly as possible and second, to express their emotions as strongly as possible to promote or demote a product. Therefore, we employ following criteria to score reviews to detect users with low reputation:

(1) Duplicate of target review

In order to generate more reviews and get more economic benefits as soon as possible, spammers often copy the text of existing reviews or make minor changes for the same or different target products. For example, they posted the same review to express different products, or just changed the product name . We can judge whether the target review is a duplicate review by measuring their similarity with other reviews from dataset. Due to the large number of reviews in dataset, the similarity is based on the following two cases:

①Duplicated reviews on the same user id in different products.

②Duplicated reviews for different user id of the same product.

Corresponding to above two cases, we consider two copy behaviors separately. One is to copy his/her own previous reviews and the other is to copy the reviews written by other users on the same product.

①Review similarity between the same user id($User\_Similarity$)

Review similarity between the same users is used to determine whether the user of the target review copied his/her own previous reviews. Therefore, the similarity is calculated between the target review and each of his/her own previous reviews, and choose the largest similarity as User_Similarity.

$$User\_Sim(r[i].u) = Max(Similarity(r[i], r[j])) \tag{1}$$

Where $r[i]$ indicates the target review, $r[j]$ indicates the jth review of his/her own previous reviews. The similarity is calculated by using the cosine formula of vector space model.

②Review similarity between the same product($Production\_Similarity$)

Review similarity between the same product is used to judge whether a target review is a copy of another user's reviews on the same product. Also, we can calculate the similarity of a target review and the rest of reviews for the same product by other users, and choose the largest similarity as Production_Similarity.

$$Production\_Sim(r[i].u) = Max(Similarity(r[i], r[k])) \tag{2}$$

Where $r[i]$ indicates the target review, $r[k]$ indicates the kth review by other users on the same product. Similarity is calculated by the same cosine formula of vector space model.

Duplicated score of review is determined by the above both user_similarity and production_similarity.

$$Duplicated\_Score(r[i].u) = 1 - (0.5 * User\_Sim(r[i].u + 0.5 * Production\_Sim(r[i].u) \tag{3}$$

(2) Review opinion

Generally, spammers tend to express strong emotional tendencies to promote or demote some target product while writing reviews, and their sentiment polarity tends to be closer to the poles, i.e, extremely strong or very weak. Conversely, genuine reviews are more likely to be closer to the middle of variance distributions.

$$O\_Score(r[i].u) = \begin{cases} 0, & if\ Sen\_Value(r[i] > \tau)\ or\ Sen\_Value(r[i]) < \tau \\ |Sen\_Value(r[i])|, & otherwise \end{cases} \quad (4)$$

Where $O\_Score(r[i])$ indicates sentiment tendency of the target review $r[i]$, τ is the threshold value which denotes the emotion polarity beyond which reviews expressed are thought to be suspicious.

The sentiment value of the target review is a challenge work. We firstly extracted the opinion word by using association rule mining.After that, original polarity of the opinion word was obtained based on PageRank model combined with mixture relevance relation. Finally, both the location of the opinion word and the context of modified information were taken into account to complete the task of the sentiment analysis of the target review. Furthermore, a greedy hill climbing search to learn τ is performed and the final estimated value is 0.758.

(3) Personal expression

Through the careful observation of the reviews, we find that the personal expression also has an implicit role for judging whether a user is a spammer. While producing a genuine review, user often express their experiences in the form of the first person to all aspects of the product. For instance, often we see such reviews, "After reading a lot of very negative review on this album, I thought I should hate this album, but I love it and it's hard to say why.", "I love her lyrics and I love the songs and what can I say she has a unique voice which I think is hot and very cool!!!", etc,. However, there are some sentences that use the "you", "you should....." and others to express, in order to recommendations or guide other consumers how to do. By this, we think such reviews are very suspicious. We can use the proportion of the first-person pronoun appeared in reviews, "I", "my", "me", "we", "us", "our", and the second-person pronouns "you", "your" to examine target reviews.

$$PersonExpre(r[i]) = \frac{FirstPerson\_Num(r[i])}{SecondPerson\_Num(r[i])} \quad (5)$$

Where $FirstPerson\_Num(r[i])$ indicates the number of the first-person pronouns in the target review $r[i]$, $SecondPerson\_Num(r[i])$ indicates the number of the second-person pronouns in the target review $r[i]$. To facilitate the comparison, we take a normalized value as the final Personal Expression Feature value $(NormalPersonalExpressionScore, NPE_Score)$.

$$NPE\_Score(r[i]) = \frac{PersonExpre(r[i])}{Max(PersonExpre(r[i])} \quad (6)$$

(4) Review length

Users tend to write review of appropriate length to express his/her true feelings about the use of product. Too long or too short are likely to be suspicious. Spammers sometimes deliberately write longer reviews to get more consumers' attention, and thus get more helpful feedback from consumers. Meanwhile, too short reviews often cannot clearly express the user's true feelings. Therefore, spammers sometimes simply use a few simple words to express, such as "good" or "bad" and so on.

$$RL\_Score(r[i].u) = \begin{cases} 0, & if\ NRLength(r[i] > \xi_1)\ or\ NRLength(r[i] < \xi_2) \\ NRLength(r[i], & otherwise \end{cases} \quad (7)$$

Where NRLength(r[i]) indicates normalized length of the target reviews (Normal Review Length, NRL), which is calculated as follows:

$$NRLength(r[i]) = \frac{|r[i]|}{Max(|r[k]|)} \tag{8}$$

$|r[i]|$ indicates the number of words in the target review. ξ_1 and ξ_2 are the parameters which denote the length beyond/below which reviews posted are considered to be suspicious. Similarly, a greedy hill climbing search to learn ξ_1 and ξ_2 was performed and in our subsequently experimental environment, the final estimated values were ξ_1=0.82,ξ_2=0.18.

(5) Helpful Feedback

Before making a purchase, users tend to look over the reviews from the other customers, and feedback it by marking that whether a review is helpful to their purchase. Naturally, we believe that the more positive feedback on a review is, the higher quality of it, and the lower probability of becoming a fake review, and meanwhile the lower probability of the reviewer becoming spammer. Consequently, helpful feedback could be:

$$HRF\_Score(r[i]) = \frac{Helpful\_FeedNum(r[i])}{FeedNum(r[i])} \tag{9}$$

Where *Helpful_FeedNum*$(r[i])$ indicates the number of users who believe the target review $r[i]$ is helpful to their purchase and feedback them after reading. *FeedNum represents*$(r[i])$ the number of all feedback review for the target $r[i]$.

3.2 Reviewer behavior features

Some of specific behavior of reviewer also has certain relevance with spammers. So, we examine fake reviewers from the following three behavioral characteristics.

(1) User Activity

Spammers usually tend to register multiple different user accounts on multiple forums or websites. When they accept a new task, they may use the newly registered user id to write and publish fake reviews. Until it is completed, they give them up. Accordingly, user's activity will be a critical aspect for detecting spammers. The activity score can be measured by examining whether the user id has release review on other products at other time.

The value could be computed as:

$$UA\_Score(r[i].u) = \frac{tf(r[i].u)}{\sum_{j=1}^{tf(r[i].u)} T(r[j].t, r[j+1].t)} \tag{10}$$

Where $tf(r[i].u)$ indicates the number of reviews written by the user $r[i].u$, $r[j].t$ refers to the release time of $r[j]$, $T(x, y)$ is the time interval between the two reviews x and y. It is important to note that when the user id of target review does not submit their review on other products at other time, the interval is considered to be infinite.

(2) Review Deviation

Many e-commerce sites usually provide review ratings of product. Assuming 5-star rating system, 1 star is the lowest level, indicating that the sentiment of customer for the product is the most negative; whereas 5 star represents the highest level, indicating that the product is most satisfactory and the review is belong to the positive reviews. Therefore, the spammers tend to raise or suppress the target product or brand by rating. For example, a user post a negative reviews about a product, while others wrote positive reviews on the same product. In this case, the reviewer is more likely to be spammer.

Or the user's reviews on a series of brand products are positive, representing that the reviewer is highly suspected as spammer. Given a set of *ratings{r[j].r,r[j+1].r, ...,r[m].r}*allocated to a product p by reviews, the deviation of a review is measured between the rating value and the normal value:

$$RD\_Score(r[i].u, R_p) = \frac{r[i].r - Avg\limits_{(j=1,2,...m)\wedge j\neq i}(r[j].r, r[j+1].r, ..., r[m].r)}{Avg\limits_{(j=1,2,...m)\wedge j\neq i}([r[j].r, r[j+1].r, ..., r[m].r]))} \tag{11}$$

Where $r[i].r$ indicates posted rating by the user $r[i].u$, $Avg([])$ denote the average of review sequence R_p(not containing the target review $r[i]$ for product p.

(3) Other rating score

According to the 5-star rating systems, ratings can be divided into three levels: ①Good (Rating≥ 4); ②bad (Rating ≤ 2.5); ③average (2.5 ≤ Rating≤ 4). Therefore, a customer's review ratings for products mainly belong to the following four conditions: ① all good or bad; ②some good and other average; ③some bad and other average; ④some good, some bad and some average. If the review ratings for other products are all fallen in the first category, it is the most feasible possibility that the reviewer is a spammer. Using the following formula to calculate the proportion of three levels.

$$\begin{aligned} GoodRatio(r[i].u) &= \frac{ReviewNum_{good}}{TotalNum} \\ BadRatio(r[i].u) &= \frac{ReviewNum_{bad}}{TotalNum} \\ AveRatio(r[i].u) &= \frac{ReviewNum_{average}}{TotalNum} \end{aligned} \tag{12}$$

According to the scale of three levels, the score of the reviewer's behavior is judged by formula (13).

$$OR\_Score(r[i].u) = \begin{cases} 0, & \text{if } GoodRatio(r[i].u)=1 \text{ or } BadRatio(r[i].u)=1 \\ Max(GoodRatio(r[i].u), BadRatio(r[i].u), AveRatio(r[i].u)), & otherwise \end{cases} \tag{13}$$

4 Proposed methodology

4.1 Reviewer reputation model for suspected spammers detection

Reputation refers to the extent of one user access to public trust, favor, and popularity, mainly focusing on qualitative evaluation. In the product reviews, each customer is allowed to present and share their own opinions for product quality, performance and price. We believe that the reviewer's reputation has a direct relation with the quality of reviews.

Consequently, the proposed method for detecting opinion spammers stems from two overarching assumptions regarding reviewer reputation:

(1) Review, written by user with lower reputation, shows lower quality and higher possibility to be fake. Consequently, with higher probability, the reviewer is spammer;

(2) if one customer has high reputation, his/her reviews are high quality, and are more helpful to others. Therefore, the user's reputation could be measured to detect suspicious spammers.

By modeling both the content of reviews and behavior of the reviewer, we propose eight features to describe the reputation of the reviewer and show the likelihood of the reviewer being

spammer from different aspects. Let reviewer's reputation was graded over [0,1]. Values close to 0 signify low reputation for reviewer and greater extent to which user are marked spammers. Similarly, Values close to 1 signify high reputation for reviewer and greater extent to which users are genuine. Therefore, if the eight normalized values of a targeted reviews are lower, the lower reputation value of the reviewer, and the greater the probability of becoming the fake reviewer is. Conversely, the higher the eight normalized values is, the higher the value of reputation of the reviewer, and the greater the likelihood of becoming a real user is. Hence, detection indicators are then used to score each reviewer and the value of reviewer's reputation will be:

$$Reputation\_Value(r[i].u) = \alpha_1 * Content\_Features(r[i].u) + \alpha_2 * Behavior\_Features(r[i].u) \tag{14}$$

Where $Content\_Features(r[i])$ indicates content characteristics of the target review $r[i]$ and include various indicators described in section 3.1.Similarly, $Behavior\_Features(r[i].u)$ refers to behavioral characteristics of the reviewer who wrote the target review $r[i]$ and is also presented in section 3.2. α_1 and α_2 are both parameters that represent the proportion of content characteristics and the behavioral characteristics in the model.

Ultimately, reviewers with reputation scores lower than the defined threshold are marked as suspected spammers.

$$L_{r[i].u} = \begin{cases} L_{normal} & \text{Reputation\_Value(r[i].u)}>\tau \\ L_{suspected\_spammer} & \text{Reputation\_Value(r[i].u)}<\tau \\ random & Reputation\_Value(r[i].u)=\tau \end{cases} \tag{15}$$

This section construct the reviewer's reputation quantification model by analyzing the target reviews' two sets of such indicators on content and the behavior characteristics of the reviewer which may indicate spamming activities, to find out reviewer with low reputation.

4.2 Spammers identification based on the K-center clustering

After the suspected spammers were detected, we believe that many of them may not all spammers. So, further screening should be need on this basis. In fact, spammer group is a group organization in which they work together to promote or damage their reputations of the target product.To achieving this, they will be close link together. This behavior is represented as a cluster structure on the network structure diagram. Therefore, based on the low-reputation user set, we adopt clustering algorithm to further distinguish between genuine users and spammer group.

The crucial problem in clustering algorithm is how to measure similarity. Different similarity measurement will result in different clustering results. In spamming activities of product reviews, each member of spammer group will publish multiple fake reviews in a short period of time once they received the designed task by the manufactures or the store owners. The purpose of it is to achieve the effect of taking control of sentiment on the target product. Therefore, users' posting time interval is proposed to measure similarity. The shorter the time interval between the two users, the more similar to each other, and the greater probability designed in the same cluster.

The users' posting time interval is calculated as the following formula:

$$TimeInterval\_Score(r[i].u, R_p) = \frac{T(r[i].t, r[j].t)}{Max(T_p)} \tag{16}$$

Where, $T(x, y)$ is similar to that defined in equation (10), $Max(Tp)$ indicates the max posting time interval of all reviews pairs for the product $r[i].p$.

Because our ultimate goal is to identify spammers, the clustering results could be reduced two clusters, one for the fake reviewers and the other for genuine users group. Based on the clustering number, the k-center algorithm is used to cluster the user set with low-reputation, and then spammers is identified finally.

5 Experimental analysis

5.1 Dataset construction and labeling

The dataset used in this research is Music product reviews from Amazon.com provided by Hu and Liu. Generally, spammers post fake reviews on a particular product or brand. Therefore, the reputation calculation of reviewer is based on their reviews. This requires us to firstly select a certain type of product reviews from the 7705 products of the Music data set. We counted the number of reviews for each different product and selected the product with the most reviews as the data set of our experiment (product number=B0002GMSC0), which contains 1424 comments.

Moreover, the dataset was marked by seven volunteers for evaluation purpose. They first refer to the methods in the report "30 Ways You Can Spot Fake Online Reviews" which distinguish the reviews as deceptive and truthful opinions, and then make use of the relation between fake reviews and reviewers to achieve a standard annotated corpus. Seven volunteers were from four college students, two postgraduates, and one PHD candidate. In order to be able to complete the labeling task rapidly and accurately, seven volunteers are also divided into two groups and each group contains two undergraduate students and a postgraduate respectively. Undergraduate students are responsible for annotating, while postgraduate is checking. When the two undergraduate students had made different judgments, the postgraduate is responsible for decision. Finally, the PHD candidate checks and confirms the conflict labels from the two groups again and determines the label by referring to the results of two postgraduates. The entire process judges were made to work in isolation to prevent any bias.

5.2 Experiments results and analysis

Experiment1: Impact of parameters for the reputation model

In our method, reputation score of reviewer plays a crucial role in the detection task. However, there are two parameters influencing the accuracy of the detection in the reputation model. In formula(14), $\alpha 1$ indicates the proportion of the review content, and $\alpha 2$ corresponds to the proportion of the reviewer behavior. Obviously, the greater the proportion, the greater the role of the corresponding factors in the overall evaluation system, and therefore the effective will be different. Therefore, the purpose of this experiment is to find best configuration for our method.

Based on our experience and actual situation, we firstly tested different parameter values. Then, the stepwise addition method was used to evaluate the effect of different parameters, with step of 0.1, ranging from 0 to 1. The experimental results under precision, recall and the F value criteria are shown in the table below.

Table 1: The impact of parameters

Parameter Value	Precision	Recall	F	Parameter Value	Precision	Recall	F
$\alpha 1 = 0.0, \alpha 2 = 1.0$	0.69	0.87	0.77	$\alpha 1 = 0.6, \alpha 2 = 0.4$	0.81	0.08	0.16
$\alpha 1 = 0.3, \alpha 2 = 0.7$	0.74	0.81	0.77	$\alpha 1 = 0.7, \alpha 2 = 0.3$	0.75	0.05	0.09
$\alpha 1 = 0.4, \alpha 2 = 0.6$	0.85	0.73	0.79	$\alpha 1 = 1.0, \alpha 2 = 0.0$	0.67	0.03	0.06
$\alpha 1 = 0.5, \alpha 2 = 0.5$	0.82	0.14	0.24				

Results from the table 1 clearly indicate that the performance of the model with the two parameters $\alpha 1 = 0.4, \alpha 2 = 0.6$ is superior to other configuration, indicating that comparing to the content-based indicators the behavioral characteristics has a more significant effect on accurately detecting spammer group. Thus, the greatest values are used in the subsequent experiments. Another interesting observation in table 1 is that the recall decreases sharply with the increase proportion of the content-based feature in the entire reputation model. Its poor performance could be attributed to the number of candidate low-reputation users. In the users' reputation model, with the proportion of content-based feature increasing, the number of candidate low-reputation users is greatly reduced. On this basis, the consequently performing clustering results in high precision of detection but low recall. On the contrary, while reducing the proportion of the content-based feature, it results in a substantial growth in the number of candidate low-reputation users, which add to the noise in the clustering, and thereby reduce the accuracy of detection but increase the recall.

Experiment2: Impact of detection metrics

In the second experiment, to further analyze the role of the content-based features and behaviour of reviewers in the spammer detection, eight influence metrics are tested separately. We implemented the experiment with different combination of the metrics, in which feature options refer to the sequence indicators, their values "1" or "0" indicates the corresponding feature work or not work.

Table 2: The impact of the detection metrics

Features Options	Precision	Recall	Features Options	Precision	Recall
01111111	0.52	0.82	11110111	0.82	0.19
10111111	0.74	0.21	11111011	0.12	0.10
11011111	0.62	0.73	11111101	0.75	0.39
11101111	0.73	0.57	11111110	0.75	0.39

From table 2, we can see that the reviewer behavior features could more influence the performance for spammer groups detection compared with the review content, especially the indicator of the user activity. When user activity feature does not work, very few candidate low-reputation users are returned, and most of them are not spammers, so their precision and recall are both very low.

Moreover, according to the results of the experiments, in the reviewer behavior indicators, the review deviation and other rating score of the reviewer are the same impact on spammer group identification.The reason is that both features return the same number of candidate low-reputation users, which lead to consistent identification recall and precision.

Another interesting observation is that among the five review content indicators, the target review's features of duplicate and sentiment have more impact on the spammer group identification than the other three. When the duplicate review indicator does not work, the results show that the excessive candidate low-reputation users are returned, indicating that reputation model does not perform the filtering of low-reputation user well, resulting in the noise increase and lower detection accuracy.Simultaneously, while the review opinion indicator does not work, too few low-reputation users are returned, indicating that a large number of real low-reputation users are filtered out, resulting in a rapid decrease of the recall.

Experiment3: Comparison of the models

In order to further verify the effectiveness of the proposed method, we conduct a comparison experiment with the Lin's work [12]. They proposed an unsupervised learning method for fake review detection. Actually, there exists a close correlation between fake reviews and spammers. A review written by a spammer is a fake review with high probability, and a fake review is almost certainly written by a spammer. Therefore, the experimental results are comparable. The specific experimental data is shown in Table 3.

Table 3: Performance comparison results

	Precision	Recall	F-Score
Our method	0.85	0.73	0.79
Lin's work	0.81	0.70	0.75

As can be seen from the data in Table 3, the model proposed in this paper clearly obtained better performance results. For the reason of analysis, in Lin's work [12], the proposed six characteristics of the review text and the reviewer's behavior are all individual spammers, ignoring the characteristics of group organizations among spammers.Therefore, some normal reviewers are mistaken for spammers. At the same time, the model proposed in this paper has gone through two stages.The first stage is to detect the suspected opinion spammers. The users with low reputation value will be regarded as suspected spammers.They are filtered out through the reputation value combining the review text and the reviewer's behavior.On this basis, the characteristics of the group behavior between the spammers are fully explored, and the clustering analysis is used to identify the spammers. The two-stage joint implementation ensures the accuracy of opinion spammers identification.

6 Conclusion and future work

In this paper, an effective opinion spammers detection approach is put forward.Based on the idea that the reviewer's reputation has a direct relation with the quality of reviews, in this paper, we first propose a reviewer's reputation model which employs the target review's content-based characteristics(context similarity, opinion sentiment, review length and helpful feedback) as well as behavior-based features(authors' activeness, review deviation and rating) to assign reputation scores to each reviewer and distinguish suspected spammers with low-reputation reviewers from ordinary reviewers.On this basis, the k=center clustering algorithm is used to perform for suspected spammers to ultimate identify spammers due to the observation that the spammers' posting time intervals burst.

On Amazon's Music real dataset, we constructed a set of annotated review, and verify the effectiveness of spammers identification method. Besides, we also analyze the effects of the eight kinds of features on identification model. The experimental results are encouraging and indicate that the spammers identification method based on reputation and clustering algorithm poses high accuracy and recall, and good performance is achieved. Furthermore, performance comparison results between the proposed method and Yuming' work show better detection accuracy.

To date, the spammer detection is still open issue.Two suggestions should be worthy attention in the future research direction. The first direction could be a deep exploring and analyzing more features of linguistic, relations, and psycholinguistic of opinion spammers to distinguish from ordinary reviewers. These valuable features could be benefit to improve the performance of detection.

Another suggestion for future work is transformation from detection to analysis social and psychological impact of spamming activities on customers. At present, a lot of work focuses on how to identify the existing spammers or spammer groups. However, harmful spammers activity can affect many potential customers psychology and decision-making. Therefore, analysis the impact of spamming activities on customer participation is our next work.

7 Acknowledgments

This work was supported by the National Science Foundation of China under Grant Numbers 71861014, 71762017 and 61662027, and the Project of Hunan Provincial Education Department(17A113,18A441,17K015) and Hunan Provincial Philosophy and Social Science Fund(16YBA228).

Bibliography

[1] Banerjee, S.; Chua, A.; Kim, J.(2015). Using Supervised Learning to Classify Authentic and Fake Online Reviews, *Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication*, 938–942, 2015.

[2] Crawford, M.; Khoshgoftaar, T.M.; Prusa, J.D. et al.(2015). Survey of Review Spam Detection using Machine Learning Technique, *Journal of Big Data*, 2(1), 1–24, 2015.

[3] Dewang, R.K.; Singh, A. K.(2015). Identification of Fake Reviews using New Set of Lexical and Syntactic Features, *Proceedings of the sixth International Conference on Computer and Communication Technology*, 115-119, 2015.

[4] Dong, M.; Yao, L.; Wang, X.(2018). Opinion Fraud Detection via Neural Autoencoder Decision Forest, *Pattern Recognition Letters*, 1-9, 2018.

[5] Heydari, A.; Tavakoli, M.; Salim, N.(2016). Detection of Fake Opinions using Time Series, *Expert Systems with Application*, 58, 83-92, 2016.

[6] Heydari, A.; Tavakoli, M.; Salim, N. et al. (2015). Detection of Review Spam: A Survey, *Expert Systems with Applications*, 42 (7), 3634–3642, 2015.

[7] Hua, N.; Boseb, I.; Koh, N. et al.(2012). Manipulation of Online Reviews: An Analysis of Ratings, Readability, and Sentitnents, *Decision Support System*, 52(3), 674-684, 2012.

[8] Jindal, N.; Liu, B. (2008). Opinion Spam and Analysis, *Proceedings of the First ACM International Conference on Web Search and Data Mining (WSDM)*, 219-229, 2008.

[9] Lau, R.Y.K.; Liao, S.Y.; Chi-Wai Kwok, R.; Xu, C. et al.(2014). Text Mining and Probabilistic Language Modeling for Online Review Spam Detection, *ACM Transactions on Management Information Systems*, 2(4), 1–30, 2011.

[10] Li, J.; Wu, G.S.; Xie, F. et al.(2016). Research of Fraud Review Detection Model on O2O Platform, *Journal of ACTA Electronica Sinica*, 44(12), 2855-2860, 2016.

[11] Lim, E.; Nguyen, V.; Jindal, N. et al.(2010). Detecting Product Review Spammers using Rating Behaviors, *Proceedings of the 19th ACM International Conference on Information and Knowledge Management(CIKM)*, 939-948, 2010.

[12] Lin, Y.; Zhu, T.; Wang, X. et al.(2014). Towards Online Review Spam Detection, *Proceedings of the companion publication of the 23rd International Conference on World Wide Web Companion*, 341–342, 2014.

[13] Liu, Y.; Pang, B.(2018). A Unified Framework for Detecting Author Spamicity by Modeling Review Deviation, *Expert Systems With Applications*, 112, 148-155, 2018.

[14] Luca, M.; Zervas, G. (2016). Fake it Till You Make It: Reputation, Competition, and Yelp Review Fraud, *Harvard Business School Working Paper*, 62, 3412-3427, 2016.

[15] Mukherjee, A.; Liu, B.; Wang, J. et al.(2011). Detecting Group Review Spam, *Proceedings of the 20th International World Wide Web Conference (WWW)*, 93-94, 2011.

[16] Ren, Y.; Ji, D.(2017). Neural Networks for Deceptive Opinion Spam Detection: An Empirical Study, *Information Sciences*, 385-386, 213-224, 2017.

[17] Savage, D.; Zhang, X.; Yu, X. et al.(2015). Detection of Opinion Spam based on Anomalous Rating Deviation, *Expert Systems with Applications*, 42(22), 8650-8657, 2015.

[18] Vlad, S.; Martin, E.(2015). Detecting Singleton Review Spammers using Semantic Similarity, *Proceedings of 24th International Conference on World Wide Web Companion*, 971-976, 2015.

[19] Zhang, W.; Bu, C.; Taketoshi, Y. et al.(2016). Cospa: A Co-training Approach for Spam Review Identification with Support Vector Machine, *Information*, 7(12), 1-15, 2016.

[20] Zhang, D.(2017). High Speed Train Control System Big Data Analysis based on Fuzzy RDF Model and Uncertain Reasoning, *International Journal of Computers Communications & Control*, 12(4), 577-591, 2017.

[21] Zhang, D.; Sui, J.; Gong, Y. (2017). Large Scales Software Test Data Generation based on Collective Constraint and Weighted Combination Method, *Tehnicki Vjesnik*, 24(4), 1041-1050, 2017.