Exploring Review Spammers by Review Similarity: A Case of Fake Review in Taiwan

Min-Yuh Day[1], Chih-Chien Wang[2], Chien-Chang Chen[1], Shao-Chieh Yang[2]
[1] Tamkang University, Taipei, Taiwan
[2] National Taipei University, Taipei, Taiwan
Email: myday@mail.tku.edu.tw, wangson@mail.ntpu.edu.tw, ccchen34@mail.tku.edu.tw, steven821015@gmail.com

Abstract

Understanding the phenomenon of spam reviews in social media is now an emerging and important issue since some enterprises may hire spammers to post fake reviews to promote their product or demote product of their competitors. The hired spammers are paid based on the fake reviews. Thus, these spammers may rewrite previous reviews as new review to earn the money. Thus, review similarity maybe a cue to detect fake reviews. Although literature had investigated the spam reviews, the review similarity of real review spammers is relatively unexplored. The objective of this paper is to explore the review spammers with a real case of fake reviews in Taiwan by investigating the cosine similarity and content length of reviews. We have proposed a text mining approach for a better understanding the phenomenon of fake reviews. The empirical results suggested that when comparing with normal reviews, the spam reviews were longer and with higher content similarity.

Keywords: Spam, Fake reviews, Text mining, Cosine similarity

1. Introduction

With the popularity of the Internet, Internet has penetrated into our daily life. People now express their comments and share their ideas about product, service or others on the Internet. These shared comments also serve as an important reference for other people to make decisions.

To encourage people to share their idea, most social media or product review website allow people to share/publish their comments or experience anonymously. The anonymous feature makes online space a place for "free-of-speech", in which people can say almost anything. Almost no one will screen the contents if the contents do not break the law and regulation.

Since many consumers will consult opinions from internet and people can provide their opinions without limitation, some unethic companies perceive the opportunistic opportunity of manipulating the majority of online opinion by providing online opinions. Thus, these unethical companies begin to hire spammers to publish fake reviews to promote reputation of themselves or to attack their competitors.

Because of the lack of appropriate filtering mechanisms on the Internet, fake reviews are flooding the Internet. In 2015, Amazon filed legal action against 1,114 spammers because of the fake reviews, which had seriously affected Amazon's goodwill, misled the consumer, and influenced the seller's trust to Amazon.

Spammers usually post in a short time with the intention to lead the opinions. They tend tomislead users to make inappropriate decision[1]. The spammers are usually part time or full time workers hired by companies to distribute fake reviews. They create fake reviews in exchange of payment. To save time and efforts, spammers may duplicate and modify previous reviews as new reviews. Thus, spam reviews may be similar with each others.

This study used a real case of fake reviews in Taiwan to discuss the similarity of fake reviews. We proposed the idea that the similarity could be a cue to detect fake reviews since that similarity among fake reviews are usually higher than that among ordinary reviews and similarity between fake reviews and ordinary reviews.

2. Related Works

2.1. Spam reviews

Spam refers to send bulk messages that the audiences do not want. In the age of internet, people get messages from e-mail, instant messaging, blog, news media, social networking, web search... and so on. They receive ordinary messages, advertising as well as spams from these media.

The history of spam can be traced back to the 1970s[2]. The initial idea of spams limits to spamming e-mail. However, due to the development of internet applications, there are new issue of spamming, such as spam in web search engine and social media.

Due to the rapid evolution of social media, social spam is now a great challenge. Chakraborty, et al. [3] argued that there were four kind of social spams. First, Malicious Links, which usually contains damage or fraud

ISBN: 978-1-941968-42-0 ©2017 SDIWC

link or other means to harm users or computers. Second, Fake Profiles, which usually provide fake personal information to avoid being found and tempted to keep in touch with the normal users. Third, Bulk Submissions, which contains a group of comments published multiple times with the same or similar text. Fourthly, Fraudulent Reviews, which claim that the product is good, or give a negative comment to attack products of competitors, even if the commenters did not have consumption experience on using the product.

Both bulk submissions and fraudulent reviews mentioned by Chakraborty, et al. [3] are relative with spam review. Since online reviews are a valuable message source for consumers to make purchase decision, companies are highly concerned with the online opinions to the product. However, not all online reviews are truthful and trustworthy since some of them are fake reviews by spammers. Previous studies had conduct review spam detection using various machine learning techniques [4].

Supervised learning were usually used to anti-fake review detection. Previous literature usually use review text itself and reviewer information as cue to detect fake review [4].

2.2. Text Mining and Similarity Analysis

Dang and Ahmad [5] mentioned that about 90% of real world data is unstructured. It is impractical to manually analyze the large number of unstructured textual information. Thus, as a result, text mining techniques are being developed to mechanize the process of analyzing this information.

Losiewicz, et al. [6] revealed that text mining architecture composed of three functions: Data collection, data warehousing, and data Exploitation. Each of these three functions included two sub-functions: Data collection contains data source selection and file selection; Data warehousing contains data conversion and data storage; Data exploitation contains data mining and data presentation. Similarity analysis in text mining are used to explore the degree of association between two documents or two sentences. We assume that the same person will have similar terms and characteristics in writing. Thus, we convert words in a posted review into tokens to calculate similarity of different posts. If the similarity of the two documents is higher, it means that there are more words in common in these two documents.

There are some approaches to calculate the similarity. For example, cosine similarity calculates the angle of the two vectors in the high dimension space; Jaccard similarity calculates the degree of similarity between the two sets; Euclidean distance calculates the actual distance between two vectors; Manhattan distance calculates the sum of the absolute wheelbase on the street map.

Lau, et al. [7] calculates Amazon's review similarity analysis using cosine similarity. In their study, if similarity above some threshold, they manually reviewed them to determine if they were spam or not.

Table 1. Fake and Normal Reviews in the Current Study

	Fake Reviews	Normal Reviews	Total
Post	434	7457	7891

Jindal and Liu [8] divides the word of mouth into three categories: Fake opinion, ordinary review, and non-comment. They collected 5.8 million reviews of products on Amazon generated by 2.14 million users and counted similarity by Jaccard similarity to judgment real or fake reviews. They got accuracy rate of 78% in their study.

Lin, et al. [9] collected 2000 fake reviews among 155080 normal reviews by Jaccard similarity. They regarded reviews as fake reviews if similarity was higher than or equal 0.7. They using those data to training the model by Logistic regression and SVM for fake review detection and try to use this model detecting other database. They got precision rate of 85% in their study.

Algur, et al. [10] used cosine similarity to detect fake from normal reviews. They considered duplicated reviews and near duplicated reviews as spam reviews, and regarded unique reviews as non-spam reviews.

As mention above, previous studies had used content similarity as feature for detecting fake reviews. However, previous studies only assume that duplicated and near duplicated reviews were fake reviews. Few previous studies, if any, had used real fake review case to compare the similarity among fake reviews and among ordinary reviews. Thus, this study uses a real case of fake reviews to reveal the correctiveness of previous assumption that some fake reviews are duplicated or near duplicated from previous reviews.

3.Methodology

3.1. Data Corpus

The data we used in this study were the same as that used in prior research [11, 12]. In this study, we focus on the original posts for understanding the fake review posts and normal posts. As shown in Table 1, we collected 7891 posts, including 434 fake review posts and 7457 normal posts

3.2. Analysis Methods

Figure 1 reveals a two phases analysis framework for understanding the phenomenon of spam review. Firstly, we calculate the number of Chinese characters in phase one. We calculate the average length of all posts, spam posts and normal posts. Then, content similarity scores were calculated to understand if the fake reviews are duplicated or near duplicated from other fake reviews.

Jieba (https://github.com/fxsjy/jieba), a Chinese segmentation tool, was used to segment Chinese contents into word tokens. We calculate similarity of words token by Cosine similarity.

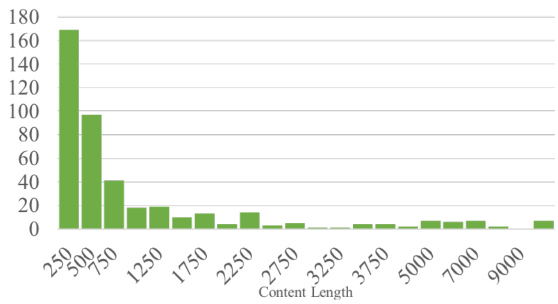
4. Data Analysis and Results

4.1. Content Length

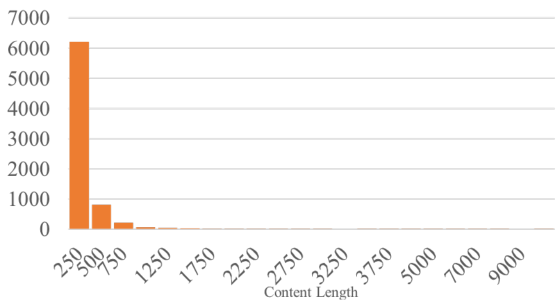
Table 2 shows the statistical summary of the content length of reviews. The mean length for all reviews posts is 236.93 Chinese characters. The mean length of spam reviews posts is 1058.91 Chinese characters, while the mean length (189.09) of normal review posts is 189.09 Chinese characters. In average, fake reviews were longer than the normal reviews.

Figure 2(a) shows the length distribution of spam reviews. In the collected 434 spam reviews, we found that 69 spam reviews are with less than 250 Chinese characters (38.94%), 265 spam reviews are with more than 250 Chinese characters (61.06%).

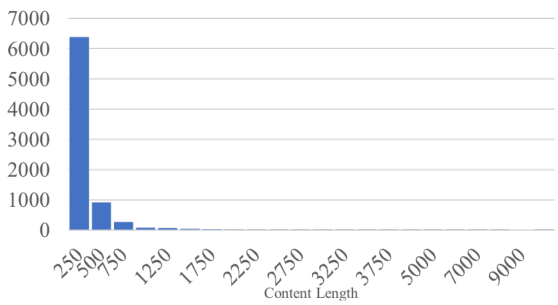
Figure 2(b) shows the length distribution of normal reviews. In the collected 7,457 normal reviews, we found that 6,207(83.24%) normal reviews were shorter than 250 Chinese characters Only 1250 normal reviews were longer than 250 Chinese characters (16.76%).



(a) Spam Reviews



(b) Normal Reviews



(c) All Reviews

Figure 2 Length Distribution of the Reviews

Table 2 Statistical Summary of the Content Length of Reviews

	All posts (n=7891)	Spam posts (n=434)	Normal posts (n= 7457)
Mean	236.93	1058.91	189.09
Standard Deviation	639.85	1906.06	425.01
First Quartile	65	168	63
Median	110	346.5	105
Third Quartile	202	1006.25	188

In Figure 2(c), a total of 7891 reviews were collected we found that among the 7891 reviews, most reviews (80.80%) content length was shorter than 250 Chinese characters. The results suggest that the length of review content is short.

The empirical analysis results show that the length of spam review is longer than normal reviews' review content in average. We argued that spammers might use long detailed reviews to persuade normal users.

4.2. Review Similarity

We use Cosine similarity measurement to analyze the similarity among reviews. We divided reviews into three groups by length: Short reviews (review length shorter than 250 Chinese Characters), middle reviews (review length between 251 and 750 Chinese Characters), and long reviews (review length longer than 750 Chinese Characters).

There were 169 fake reviews that were shorter than 250 Chinese characters (table 3). We randomly selected same amount (169) of normal posts for comparison purpose. In table 3, we found that the similarity among

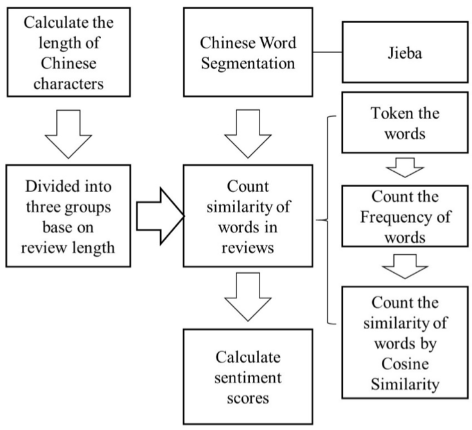


Figure 1 Content Similarity Analysis Procedure

Table 3 Similarity Analysis for Short Reviews (review length shorter than 250 Chinese Characters)

	(1) Similarity among Fake Reviews	(2) Similarity between Fake and Normal Reviews	(3) Similarity among Normal Reviews
Average Similarity	0.21	0.17	0.16
Standard Deviation	0.12	0.11	0.10
F value (P value)	F=664.09 (p<0.001)		
Post Hoc Test	(1)>(2)>(3)		

Notes: There were 169 short length fake reviews in our corpus. We randomly selected another 169 short length normal reviews for comparison purpose.

Table 4 Similarity Analysis for Middle Length Reviews (review length between 251 and 750 Chinese Characters)

	(1) Similarity among Fake Reviews	(2) Similarity between Fake and Normal Reviews	(3) Similarity among Normal Reviews
Average Similarity	0.37	0.32	0.30
Standard Deviation	0.15	0.15	0.15
F value (P value)	F=683.27 (p<0.001)		
Post Hoc Test	(1)>(2)>(3)		

Notes: There were 138 middle length fake reviews in our corpus. We randomly selected another 138 middle length normal reviews for comparison purpose.

Table 5 Similarity Analysis for Long Length Reviews (review length longer than 750 Chinese Characters)

	(1) Similarity among Fake Reviews	(2) Similarity between Fake and Normal Reviews	(3) Similarity among Normal Reviews
Average Similarity	0.50	0.39	0.34
Standard Deviation	0.23	0.23	0.21
F value (P value)	F=1060.00 (p<0.001)		
Post Hoc Test	(1)>(2)>(3)		

Notes: There were 127 long length fake reviews in our corpus. We randomly selected another 127 long length normal reviews for comparison purpose.

fake reviews were the highest, and the similarity among normal reviews are the lowest. The average similarity between fake review and normal reviews are in the middle. There were 138 fake reviews with content length in the range of 251 to 750 Chinese characters (Table 4). We randomly selected the same amount (138) of normal reviews for comparison purpose. In table 4, we also found that the similarity among fake reviews are the highest, and the similarity among normal reviews are the lowest. The average similarity between fake review and normal reviews are in the middle.

There were 127 fake reviews with content length of more than 751 Chinese characters (Table 5). We randomly selected the same amount (127) of normal reviews for comparison purpose. In Table 5, we found that the similarity among fake reviews are the highest, and the similarity of the group among normal reviews are the lowest. The average similarity between fake review and normal reviews are in the middle.

Table 5 reveals that the average similarity coefficient among long fake review is 0.50 while the average similarity coefficient is 0.34 among long normal review. The difference of similarity coefficient between long fake reviews and long normal review is 0.16. Table 6 reveals the similarity analysis results for all reviews (do not divide reviews into three groups). The average similarity coefficient among fake review is 0.33 (do not divide reviews into three groups), while the average similarity coefficient among long normal review is 0.25(do not divide reviews into three groups). Thus, review length is a potential moderator when using content similarity as a cue to detect fake reviews.

5. Discussion

Based on the content similarity analysis for fake review, we found that the similarity among fake reviews were higher than that among normal reviews or between normal reviews and fake reviews, no matter the review content is with short or long. However, if we did not divide the reviews based on length of the reviews, we can not observe this phenomenon of similarity since the similarity are low between long and short length reviews.

Secondly, we found that normal reviews with 250 or less Chinese characters accounted for 83.24% of normal reviews. However, spam posts with 250 or less Chinese characters accounted for only 38.94%. The research results suggest that the spam posts are generally more longer than normal posts.

The contributions of this paper are three folds. First, we have proposed a texting mining approach and explored the review spammers with a real case of fake review in Taiwan by investigating the cosine similarity and content length of reviews. We discovered the spam reviews tend to have higher content similarity and longer reviews than normal reviews

Second, we used text mining techniques with Cosine similarity for analyzing the similarity of spam reviews post. We found the content similarity among fake

Table 6 Similarity Analysis for All Reviews

	(1) Similarity among Fake Reviews	(2) Similarity between Fake and Normal Reviews	(3) Similarity among Normal Reviews
Average Similarity	0.33	0.27	0.25
Standard Deviation	0.20	0.18	0.17
F value (P value)	F=1685.86 (p<0.001)		
Post Hoc Test	(1)>(2)>(3)		

Notes: There were 434 fake reviews in our corpus. We randomly selected another 434 long length normal reviews for comparison purpose.

reviews are higher than the similarity between fake and normal reviews and similarity among normal reviews.

Third, based on the analysis of content length, spam reviews are longer than normal reviews. This empirical analysis results suggest the fact that spammers would likely to use longer and detailed contents to persuade the consumers to believe the review content they post on Internet. Based on this observation, we should keep more attention on long length reviews when we want to detect fake reviews.

6. Reference

[1] S. Xie, G. Wang, S. Lin, and P. S. Yu, "Review spam detection via temporal pattern discovery," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2012, pp. 823-831: ACM.

[2] F. Brunton, *Spam: A shadow history of the Internet*. MIT Press, 2013.

[3] M. Chakraborty, S. Pal, R. Pramanik, and C. R. Chowdary, "Recent developments in social spam detection and combating techniques: A survey," *Information Processing & Management,* vol. 52, no. 6, pp. 1053-1073, 2016.

[4] M. Crawford, T. M. Khoshgoftaar, J. D. Prusa, A. N. Richter, and H. Al Najada, "Survey of review spam detection using machine learning techniques," *Journal of Big Data,* vol. 2, no. 1, p. 23, 2015.

[5] D. S. Dang and P. H. Ahmad, "A Review of Text Mining Techniques Associated with Various Application Areas," *International Journal of Science and Research (IJSR),* vol. 4, no. 2, pp. 2461-2466, 2015.

[6] P. Losiewicz, D. W. Oard, and R. N. Kostoff, "Textual data mining to support science and technology management," *Journal of Intelligent Information Systems,* vol. 15, no. 2, pp. 99-119, 2000.

[7] R. Y. Lau, S. Liao, R. C.-W. Kwok, K. Xu, Y. Xia, and Y. Li, "Text mining and probabilistic language modeling for online review spam detection," *ACM Transactions on Management Information Systems (TMIS),* vol. 2, no. 4, p. 25, 2011.

[8] N. Jindal and B. Liu, "Opinion spam and analysis," in *Proceedings of the 2008 International Conference on Web Search and Data Mining*, 2008, pp. 219-230: ACM.

[9] Y. Lin, T. Zhu, H. Wu, J. Zhang, X. Wang, and A. Zhou, "Towards online anti-opinion spam: Spotting fake reviews from the review sequence," in *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*, 2014, pp. 261-264: IEEE.

[10] S. P. Algur, A. P. Patil, P. Hiremath, and S. Shivashankar, "Conceptual level similarity measure based review spam detection," in *Signal and Image Processing (ICSIP), 2010 International Conference on*, 2010, pp. 416-423: IEEE.

[11] C.-C. Wang, M.-Y. Day, and Y.-R. Lin, "Toward understanding the cliques of opinion spammers with social network analysis," in *Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on*, 2016, pp. 1163-1169: IEEE.

[12] C.-C. Wang, M.-Y. Day, and Y.-R. Lin, "A Real Case Analytics on Social Network of Opinion Spammers," in *Information Reuse and Integration (IRI), 2016 IEEE 17th International Conference on*, 2016, pp. 623-630: IEEE.