

Research on Identification Method of Anonymous Fake Reviews in E-commerce

Lizhen Liu^{*1}, Xinlei Zhao², Hanshi Wang³, Wei Song⁴, Chao Du⁵

Information and Engineering College, Capital Normal University, Beijing 100048, P. R. China

*Corresponding author, e-mail: liz_liu@126.com¹, xinlei8912@sina.com², necrostone@sina.com³,
wsong@cnu.edu⁴, cnuie_dc@126.com⁵

Abstract

In this paper, a new method has been proposed for identifying anonymous fake reviews generated by click farmers in E-commerce and improves the identification rates. Anonymous fake reviews are different from the genuine reviews. They could be distinguished based on the credibility of users, the average daily number of evaluations, the content similarity, and the degree of word overlapping. The proposed method takes into account these 5 features to calculate the fake reviews content by constructing multivariate linear regression model. Experiments show that this preliminary work performed well in identifying fake reviews in Chinese E-commerce website. The extracted features are also useful to identifying the fake reviews when the reviewer's identification is not accessible.

Keywords: volume of fake reviews, feature extraction, multi-linear regression, click farming

Copyright © 2016 Universitas Ahmad Dahlan. All rights reserved.

1. Introduction

It is observed that the E-commerce and opinion-sharing website are gaining their popularity following the emergence of Web 2.0 [1]. Online commenting and reviewing are widely adopted in many E-commerce platform such as Amazon, Taobao, and Tmall, which allow users to exchange their purchasing or consuming experiences. Positive reviews can increase shop's reputations and attract more customers, while negative ones may bring out potential sales loss [2]. Consumers use these reviews not only to receive word-of-mouth (WOM) information on products, such as quality, suitability and utility, but also to input their own reviews to advice other consumers [3]. As the influence of online review becoming a critical factor of the E-commerce market, unfortunately, it is noticed hundreds of click farming groups have sprung up in recent years which delivering bundles of fake positive reviews to merchants requiring a quick and dirty way to boost their popularity. The consumers who make purchase decision rely on those fake reviews posted by 'click farmer' could be disappointed as the product will not meet their expectation [4]. Moreover, many E-commerce platforms protect the privacy of the users through providing anonymous evaluating services, potentially facilitates the generation of fake reviews.

Now, the data security is the biggest issue for the consumers in E-commerce [5]. So anonymous fake reviews analysis has drawn many researchers' attention. The reason of investigating the fake reviews in E-commerce is to identify the fake reviews and help the consumers to get genuine information of the product or service they want consume. Therefore, how to identify the anonymous fake reviews in E-commerce is a necessary work. But, very few researchers explore anonymous fake review identification topic till now. Y. Fu and B. H. Dong [6] provide a model to extract the fake reviews from Taobao and Tmall website, but the method they employed is rely on some internal business data, thus may not applicable to other E-commerce platforms. This research tries to set up a model to identify on anonymous fake review purely rely on public reviews. Instead of using special customer database, our model achieves the goal by extracting review feature and model training upon public reviews.

For the problems mentioned above, the paper firstly proposes five assumptions of the feature extraction based on analysis of evaluation processing. Simultaneously, a concept named VoFR is defined to compute fake reviews. Then a supervised model for finding click farming product and detecting fake reviews is developed with VoFR. Finally, the experiment

shows the supervised VoFR model accuracy is 92.5%, precision is 94.7%, the recall is 90%. In addition, the unique contribution of the paper fills the gaps in identification of anonymous fake reviews.

The rest of the paper is organized as follows. Related work is reviewed in Section 2. Based on analysis of evaluation processing, we propose five assumptions in Section 3, followed by extraction of the feature functions and supervised VoFR model in Section 4. Section 5 presents our detailed experimental results. We summarize this work in Section 6.

2. Related Work

Generally, the research related to analyzing the fake reviews can be roughly fallen into two categories: identifying the text of fake reviews and identifying the click farmers.

For fake text detection, Jindal, et al., [7-9] proposed the concept of fake reviews in 2007 at the first time. He collected the reviews from Amazon and manually labelled the fake reviews. Logistic regression is employed in his research to identify the fake reviews. Lai, et al., [10] proposed a recognition method named unigram model. The grammars and formats of the text are used as features for classification. Ott, et al., [11] formulated the problem of identifying fake reviews as a binary classification problem. Besides that, the fake reviews are solely identified through the texts in [12-14]. Even through huge progress has been made in text based fake review detection in recent years, the methods require high level understandings of the words in the reviews, which is hard to archive since the fake reviews are intended to mislead the consumers.

For detecting the click farmers, researchers assume the click farmers could generate more fake reviews than the genuine reviews. References [4, 15] track the behavior of consumers to detect the click farmers. Liu, et al., [16] exploited an 'undesired rule' to identify the fake reviewers. However, the aforementioned algorithms could only identify a specific category of click farmers. Mukherjee, et al., [17, 18] considered the generating of fake reviews as a group behavior, he proposed a triple-cross model to identify the click farmers through integrating the group features and individual features. References [19, 20] considered the relations between the reviewers, reviews and shops to identify the click farmers.

In this paper, both the userID and historic activity of the consumers are not available under anonymous condition, so that we propose a new way to solve the problem. The fake reviews are identified through feature extraction and model training. The details of the methods are as following.

3. Click Farming Evaluation Processing

To better understand the fake reviews and the feature of click farmers, we investigated the pipeline of generating fake reviews in Taobao & Tmall, two largest E-commerce platform in China. A detailed model of click farming is studied comprehensively. As shown in Figure 1, the survey found that fake reviews are generated exactly the same as the genuine reviews. As the review is made anonymously, there is no way to access the real identification of the consumers, which brings difficulties to our investigation. Therefore, we proposed five assumptions to describe the anonymous reviews: the users' credibility, the average daily number of evaluations, the similarity of reviews and products description, the overlap between reviews and ratio between sales volume and shop running time. Features are extracted based on these assumptions and then used to train a model to identify the fake reviews.

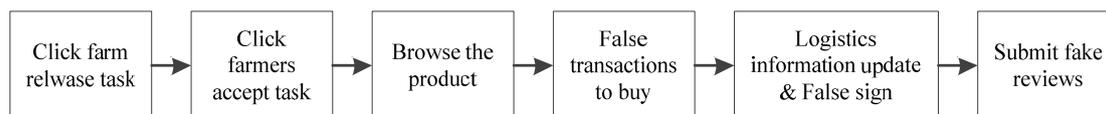


Figure 1. Click farming process

3.1. Users Credibility

Once consumers finished their shopping and evaluating, merchants in Taobao&Tmall will give them a positive or negative rating. The 'positive' rating will add one point, 'negative' rating reduce one point, thus caused consumers to produce the different user credibility. So, the more consumers use their ID for shopping the more credibility they have.

Users credibility reflect both the number of goods that the consumer has purchased and the credit of the consumer has. The credibility cannot be hidden even if the consumer choose anonymity. According to our investigation, the click farmer's salary is partially depending on the credibility. Higher credibility is corresponding to higher salary. However, what we found is E-commerce platform always restrict the maximum number of reviews that can be posted by one account in a certain period. So the click farmers, in usual, has multiple userID whose identity has not been verified by the E-commerce platform. These auxiliary accounts always have lower buying records and lower user credibility, which at the end, declines the average credibility of the product whose reviews are generated by the click farmers. Also, since the cost of hiring a click farmer with higher credibility is higher than hiring a click farmer with lower credibility, the merchant always choose to hiring more click farmers to generate more reviews instead of hiring someone has more experience. Hence, we assume that negative correlation exists between the users credibility and number of fake reviews that a product received:

Assumption I: The average user's credibility of a product with more fake reviews is lower than the average user's credibility of a product with more genuine reviews.

3.2. Average Daily Number of Evaluations

The evaluation time can be obtain from the website, almost all the merchants choose click farming is for improving sales and rating. The high sales mean better reputation, better products. Similarly, these higher reputation and sales shops will attract more consumers to buy, and then the daily number of reviews will increase. These merchants have a common characteristic that the average daily number of evaluations is lower.

Assumption II: The scalping products average daily number of evaluations is lower than normal products.

3.3. Similarity of Reviews and Products Description

Since fake reviews are not generated through any customer experience, the contents of the fake reviews are always monotonous and tiresome. The only source for click farmers to obtain the information about the product is through reading its online description. Since the fake reviews are usually generated and organized based on the description, result in a high similarity between the fake reviews and the description of the product.

Assumption III: The similarity between the fake reviews and the description of the product is higher than the similarity between the true reviews and the description of the product.

3.4. Overlap between Reviews

According to expectancy theory in reference [21], the motivation force experienced by an individual to select one behavior from a larger set is some function of the perceived likelihood that that behavior will result in the attainment of various outcomes weighted by the desirability of these outcomes to the person. Since it is not reward able to publish a carefully written review, which is generally between 0.3 to 1.3 dollars, the click farmers, who are lack of real customer experience, always choose to edit and reorganize previous reviews. Therefore, the fake reviews always plagiarize each other. Reference [22] referred to a method of detecting text plagiarism, COPS which detected document overlap by relying on string matching and sentences. Learning the method, the overlap between reviews is calculated.

Assumption IV: The overlap between fake reviews is higher than the similarity between the genuine reviews.

3.5. Ratio between Sales Volume and Shop Running Time

Based on statistics, we found that there is a relative constant ratio between the selling volume and the time of the online shop established. A click farming shop usually has higher selling volume but short running time.

Assumption V: The click farming product has a lower ratio between selling volume and selling age than a normal product.

4. Supervised VoFR Model

We propose a new method to identify click farming product by computing the volume of face reviews (VoFR). The VoFR is computed based on the features proposed in Section 3, a multi-linear regression model is adopted, and the result of the multi-linear regression is the VoFR model. Here, the VoFR of clickfarming product is 1, the VoFR of a normal product is 0. The model is defined as following:

$$VoFR = \beta_0 + \beta_1\phi_c(p_i) + \beta_2\phi_t(p_i) + \beta_3\phi_s(p_i) + \beta_4\phi_o(p_i) + \beta_5\phi_r(p_i) \quad (1)$$

$\phi_c(p_i)$, $\phi_t(p_i)$, $\phi_s(p_i)$, $\phi_o(p_i)$, $\phi_r(p_i)$ respectively represent the feature function of users credibility, the average daily number of evaluations, similarity of reviews and product description, the overlap between reviews, and the ratio between sales volume and registration time. β_{0-5} are the weights, they are learned through the training dataset.

4.1. Credibility Feature Function

Although the userID is anonymous, neither could track any historic information of any click farmers, user's credibility can be obtained. The weight is assigned based on Table 1. The footnote represents the weight of products' review.

Table 1. The Weight ω_{ij}

Users credibility	ω_{ij}
0 gold crown	1
1 red heart	0.9
2 red heart	0.8
3 red heart	0.7
4 red heart	0.6
5 red heart	0.5
1 diamond	0.4
2 diamond	0.3
3 diamond	0.2
4 diamond	0.1
5 diamond and above	0

In Taobao, 0 Golden Crown represents user with the lowest credibility. User with the greatest credibility is assigned the lowest weight to ensure the credibility of product fall into the interval of [0, 1].

Calculation of average credibility following formula:

$$\phi_c'(p_i) = \bar{C}_{ij} = \frac{\sum_j (c_{ij} \cdot \omega_{ij})}{|R_i|} \quad (2)$$

$$\phi_c(p_i) = \frac{\phi_c'(p_i) - \min(\phi_c'(p_i))}{\max(\phi_c'(p_i)) - \min(\phi_c'(p_i))} \quad (3)$$

\bar{C}_{ij} represents the average degree of credibility of the product p_i 's review r_j , $|R_i|$ represents the number of product p_i of all reviews, $\sum_j (c_{ij} \cdot \omega_{ij})$ represents the sum of the number of products p_i 's reviews r_j with the corresponding weights ω_{ij} . The average credibility is computed which contains all the fake and genuine reviews of one product. The final value is normalized through max-min normalization.

4.2. Time Feature Function

Time feature function is used to name the average number of evaluations during one day, is represented by $\phi'_t(p_i)$:

$$\phi'_t(p_i) = \frac{R_i}{T} \quad (4)$$

$$\phi_t(p_i) = 1 - \frac{\phi'_t(p_i) - \min(\phi'_t(p_i))}{\max(\phi'_t(p_i)) - \min(\phi'_t(p_i))} \quad (5)$$

Wherein, R_i represents all the reviews of p_i 's, T represents the total days of collection to obtain R_i .

4.3. Similarity Feature Function

Since the click farmers do not have customer experience, their reviews are always based on the description of products, so that the similarity between the reviews and the product can be used to character the genuineness of the review.

$$\text{sim}(\mathbf{r}_{ij}, \mathbf{d}_i) = \frac{\mathbf{r}_{ij}^T \cdot \mathbf{d}_i}{|\mathbf{r}_{ij}| |\mathbf{d}_i|} \quad (6)$$

$$\phi'_s(p_i) = \text{Mean}(\text{sim}(\mathbf{r}_{ij}, \mathbf{d}_i)) \quad (7)$$

$$\phi_s(p_i) = \frac{\phi'_s(p_i) - \min(\phi'_s(p_i))}{\max(\phi'_s(p_i)) - \min(\phi'_s(p_i))} \quad (8)$$

\mathbf{r}_{ij} represents the vector of review r_j for product p_i , \mathbf{d}_i represents the vector of the description of product p_i , $\text{sim}(\mathbf{r}_{ij}, \mathbf{d}_i)$ is the cosine similarity between product p_i 's review and its description, $\phi_s(p_i)$ is the normalized similarity score.

Reviews datasets and description datasets are built to calculate the aforementioned similarity, the process is shown in Figure 2.

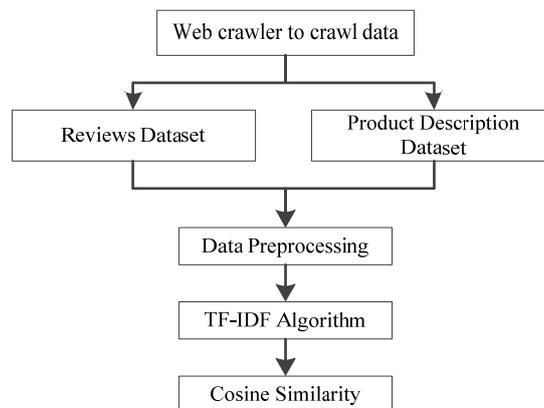


Figure 2. Calculation the similarity

The similarity between each review with the product description is calculated to measure the objectivity of reviews. The higher the degree of similarity between products with its description, the less emotional words it used, the more objective of the review. Based on the AssumptionIII, the more objective of the review, the higher possibility of the review belongs to a fake review.

4.4. Overlap Feature Function

A word set for each review of product p_i is built, excluding any duplicated words. The average overlap value between two word sets is computed as following:

$$R_i = \{r_{ij}\}_{j=1}^{N(p_i)} \quad (9)$$

$$\phi'_o(p_i) = \text{Mean}\left(\frac{|R_i \cap R_j|}{\max(|R_i|, |R_j|)}\right) \quad (10)$$

$$\phi_o(p_i) = \frac{\phi'_o(p_i) - \min(\phi'_o(p_i))}{\max(\phi'_o(p_i)) - \min(\phi'_o(p_i))} \quad (11)$$

R_i represents word collection contains all the word sets of p_i , $\phi'_o(p_i)$ represents the average overlap value between any pair of word sets, $\phi_o(p_i)$ represents the normalized value of overlap score.

4.5. Ratio Feature Function

Analysis the selling volume find that the selling volume of click farming shops are less than the normal shops, however, their registration time only have slight difference. Hence we add this feature into our model.

$$\phi'_r(p_i) = \frac{G_i}{L_i} \quad (12)$$

$$\phi_r(p_i) = 1 - \frac{\phi'_r(p_i) - \min(\phi'_r(p_i))}{\max(\phi'_r(p_i)) - \min(\phi'_r(p_i))} \quad (13)$$

G_i represents the total selling volume of product p_i , L_i represents the age of shop p_i 's, we employ the same data normalization method to normalize our data.

5. Experiment Analysis

The overview of the proposed method is shown in Figure 3.

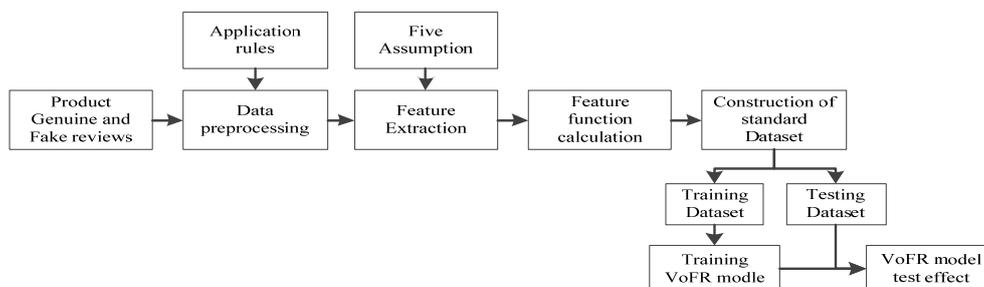


Figure 3. Experimental flow chart

80 pairs of product are randomly selected as the training dataset, 20 pairs as a testing dataset. A multi-linear regression is learned on training dataset, and tested on the testing dataset.

5.1. Training Dataset Collection

To investigate the features of fake reviews, understand the pipeline of click farming, 116363 reviews from Taobao and Tmall platform are collected. The genuine reviews are selected and downloaded from the shops where the authors had purchased something from. The shops have truly good consumer experience and existed for a long period of time. To collect the fake reviews, the author pretended to be a click farmer. We found that fake reviews are also overwhelmed in Tmall platform. After data preprocessing, features are extracted on the basis of assumptions mentioned in section 3.

5.1.1. Fake Reviews Dataset

Through investigation, we found that click farming products are usually come with kick-started online shops whose reviews are few and selling volumes are low. The quality of the products that they sell is not necessarily worse than their competitors. The fake reviews are mainly generated for improving their selling volume, revenue, and rankings. Hence, the following rules are employed in preprocessing.

Rule1: exclude any reviews that contain images or additional comment.

It is not usual to insert images or additional comments in the fake reviews.

Rule2: exclude any reviews that human cannot identify.

A label cannot be assigned if human could not distinguish the review.

Rule3: exclude any reviews that contain advertisements.

In practice, these reviews can be excluded through simple heuristic rules.

Rule4: exclude any reviews that contain real shopping experience.

Rule5: exclude any negative reviews.

Based on five rules, reviews of neck and shoulder massage have been preprocessed, as shown in Table 2.

Table 2. Amazon, Taobao and Tmall data comparison

Reviews text	Rules
Product s is also product, gave a product choice elders birthday gift! Like a product mother, product customer service attitude, logistics fast! Excessive additional period of time to continue. http://img.alicdn.com/bao/uploaded/i1/197630138383715927/TB2luNRhXXXXXaQXpXXXXXXXXXX XX_!0-rate.jpg	Rule1
http://img.alicdn.com/bao/uploaded/i2/197630138383737204/TB2OGdYhXXXXX4XpXXXXXXXXXX XX_!0-rate.jpg	
Product very product, praise.	Rule 2
Professional tattoo, please add QQ *****	Rule 3
Product packaging, housing also has a corresponding manufacturers. But the product gap is large, does not meet the proper quality brand. The use of fever phenomenon, the general attitude of customer service Mike.	Rule 4
It's completely fake, I'm not black you, what is to spread the products, on the value of a 20 yuan a massage head is crooked, the intensity is very small, yet I bought 40 yuan a product, who do not believe, who regret, asked people to click farming all the praise!	Rule 5

After pretreatment the fake reviews dataset has a total of 80 products, 51 merchants, 10708 reviews, related to the eight categories of clothing, footwear, electronic appliances, and furniture etc.

5.1.2. Genuine Reviews Dataset

Contrast to fake reviews, genuine reviews are also from Ali's Taobao&Tmall website. Facing the same product and price, the higher the ranking the more likely to be selected. So click farmers prevalent in newly established shop or a new product. For merchant who sales rankings already high, there is no need for click farming, because these shops in pre-sales have accumulated a great deal of popularity. So we choose higher sales merchant, and choose the recent reviews to set up genuine reviews dataset. High sales of products 'reviews generally have tens of thousands, according to the time sequence, we choose the reviews nearly 2

months, and then pretreatment, artificial remove advertising reviews, get genuine reviews dataset, used for VoFR model training. The genuine reviews dataset has a total of 80 products, 50 merchants, and 10565reviews, related to eight categories.

5.2. Supervised VoFR Model Analysis

Feature function’s calculations show us, there is different between the fake reviews and genuine reviews. Expressed as $\phi_{c.t.s.o.r}(p_{shill}) > \phi_{c.t.s.o.r}(p_{truth})$ in Fig.4-8. We choose 40 pairs of the data as sample.

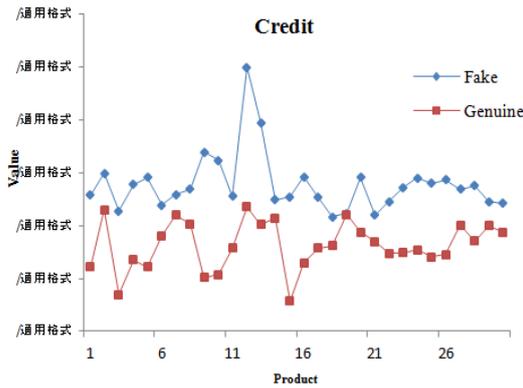


Figure 4. Credit’s comparison

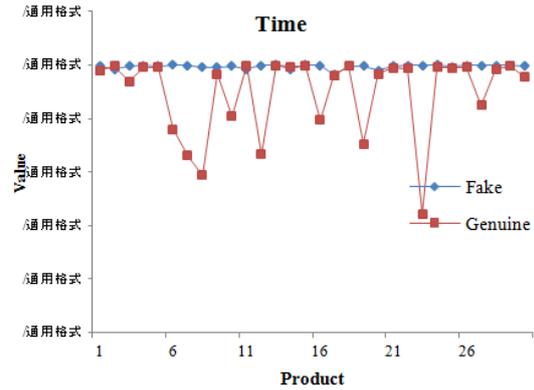


Figure 5. Time’s comparison

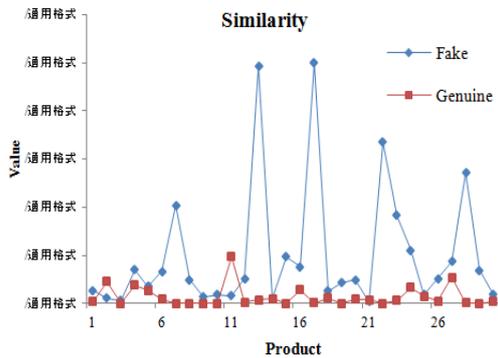


Figure 6. Similarity’s comparison

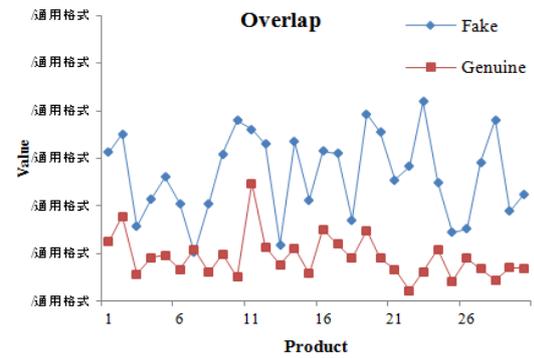


Figure 7. Overlap’s comparison

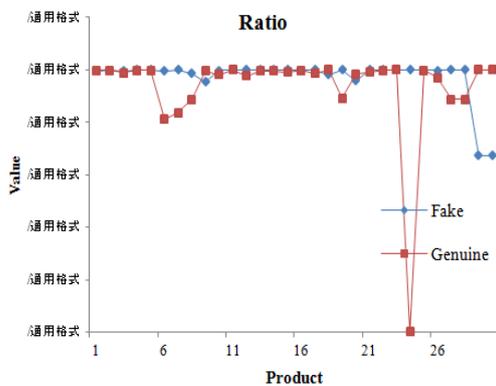


Figure 8. Ratio’s comparison

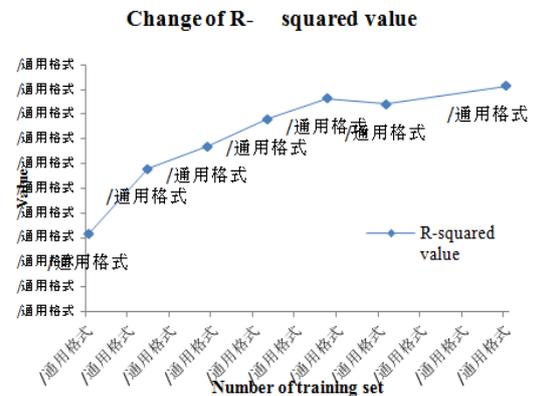


Figure 9. Change of R^2

For the training dataset 80 pairs of products' features function results, using the amount of data accumulated gradually fitting manner VoFR model, and the model is fitted with 10, 20 and 40 pairs datasets gradually accumulated, and the R^2 's changes are found as shown in Figure 9.

According to the change of R^2 , when choosing 60 pairs, R^2 has a downward trend. When choosing of 80 pairs, R^2 has improved. It is speculated that as the dataset increase, model fitting a gradual increase in the degree, the final choice of 80 pairs as the training set, get the model parameters as shown in the following Table 3. Final determination VoFR model:

$$VoFR = -0.937 + 1.41 \times \phi_c(p_i) + 0.297 \times \phi_l(p_i) + 0.157 \times \phi_s(p_i) + 0.92 \times \phi_o(p_i) + 0.236 \times \phi_r(p_i)$$

Table 3. Model parameter

β_{0-5}	β_0	β_1	β_2	β_3	β_4	β_5
Value	-0.937	1.410	0.297	0.157	0.920	0.236

Test datasets is composed by 20 click farming products (32252, 13 shops) and 20 normal products (2453, 13 shops). The critical value is 0.5. If VoFR's output is higher than 0.5, the product is identified click farming, otherwise is normal product. The classification results are shown in Figure 10:

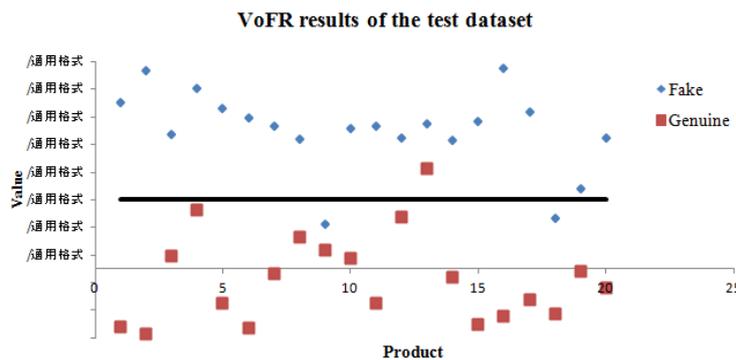


Figure 10. VoFR results of the test dataset

5.3. Evaluation Indicators

The evaluation indicators: accuracy, precision, and recall. Here defined as follows:

$|TP|$ is the number of click farming products correctly to be predicted click farming.

$|FP|$ is the number of the normal products wrong to be predicted click farming.

$|FN|$ is the number of click farming products wrong to be predicted the normal.

$|TN|$ is the number of normal products correctly to be predicted the normal.

Formula as follows:

$$Accuracy = \frac{|TP| + |TN|}{|TP| + |FP| + |TN| + |FN|} \tag{14}$$

$$Precision = \frac{|TP|}{|TP| + |FP|} \tag{15}$$

$$Recall = \frac{|TP|}{|TP| + |FN|} \tag{16}$$

Accuracy represents an accuracy of click farming products and normal products can be correctly classified. Precision represents the click farming products can be successfully detected the accuracy rate. It reflects the accuracy of the classification results of VoFR model. Recall that said the probability of a correct classification of the click farming, said the proportion of click farming in total products.

Now, almost all research on fake reviews are relying on Amazon. We compared the different from the reviews data on Amazon, Taobao and Tmall. As shown in Table 4. Different data makes it difficult on the study.

Table 4. Amazon, Taobao and Tmall data comparison

E-commerce platform	Rating	ReviewersID	Reviewsusefulness	Time	Users credibility	Users other purchase	Users ranking
Amazon	√	√	√	√	√	√	√
Taobao	x	x	x	√	√	x	x
Tmall	x	x	x	√	√	x	x

Table 4 tells that anonymous users cannot be track the purchase history and the reviews cannot be evaluated by rating, so we propose VoFR modle to classify clickfarming products and the fake reviews.

Reference [6], the dataset is also from Taobao and Tmall. The click farmers' userID is obtained, tracking purchasing informationfor analysis and getting 14 features. SVM algorithm and KNN algorithmare respectively applied for identification click farmers in Taobao. Its evaluation indicators and ours are shown in Table 5.

Table 5. Evaluation IndicatorsComparison

Method	Accuracy	Precision	Recall	TP	TN	FP	FN
SVM	93%	88%	100%	30	26	4	0
KNN	78%	81%	73%	22	25	5	8
VoFR	92.5%	94.7%	90%	18	19	1	2

First, Reference [6] was analyzed for click farmers'userID and purchasing information, which is difficult to obtain. In the era of data security, such an approach also could infringe on user privacy, it is unsafe. Our dataset is from the public information in Taobao and Tmall. Second, since the data is too specific in Reference[6], thereby causing poor can migrate. Third, comparing Reference [6]'s 14 features, our research has 5 features. But the result of evaluation index is not much difference. Precision is better than SVM and KNN. So that, some features in Reference[6] are useless. Forth,Reference[6] conducted research for the consumers, while this paperresearch for products which can be applied for anycategory, more comprehensive data is available.

6. Conclusions and Future Work

Click farming in E-commerce is an unignorably and challenging issues, it misleading the consumer's purchase decisions. This paper compares the click farm pattern and normal shopping pattern, then proposes the application of the feature function and multi-linear regression method of constructing VoFR model. By calculating the volume of fake reviews to identify click farming products, real consumer's decision is provided to users. Experimental datasets obtain from China's largest E-commerce platform's (Taobao and Tmall) reviews, and manual tagging the anonymous user reviews of standard data sets to ensure the accuracy of the data. Experiments show that by calculation and analysis, the five features functions as input data for VoFR model are effective identification method.

The future work will be further study for VoFR model, extracting feature functions and features of importance rating further improve the accuracy of VoFR model. And the method is extended to other fake reviews to identify missing information, and to expand the applicability of

the research results. In addition, for the text of reviews still can make a further research on emotional information. Emotional lexicon will be built to improve the efficiency of VoFR model.

Acknowledgement

This work was supported in part by the Humanity & Social Science general project of Ministry of Education under Grants No.14YJAZH046, National Science Foundation of China under Grants No. 61303105 and 61402304, the Beijing Natural Science Foundation under Grants No. 4154065, the Beijing Educational Committee Science and Technology Development Planned under Grants No.KM201610028015 and KM201410028017, Beijing Key Disciplines of Computer Application Technology, Science and technology innovation platform, Teachers' Construction & Teacher' s Teaching Promotion.

References

- [1] A Heydari, MA Tavakoli, N Salim, Z Heydari. Detection of review spam: A survey. *Expert Systems with Applications*. 2015; 42: 3634-3642.
- [2] H Sun, A Morales, X Yan. *Synthetic review spamming and defense*. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. 2013: 1088-1096.
- [3] CG Harris. Detecting Deceptive Opinion Spam Using Human Computation. 2012.
- [4] EP Lim, VA Nguyen, N Jindal, B Liu, HW Lauw. *Detecting product review spammers using rating behaviors*. In ACM International Conference on Information & Knowledge Management. 2012: 939-948.
- [5] A Sarwar, MN Khan. A Review of Trust Aspects in Cloud Computing Security. *International Journal of Cloud Computing and Services Science (IJ-CLOSER)*. 2013; 2: 116-122.
- [6] Y Fu, BH Dong. Research on identifying spammer in E-commerce. *Scientific Decision-Making*. 2015: 79-94.
- [7] N Jindal, B Liu. *Review spam detection*. In International Conference on World Wide Web. 2007: 1189-1190.
- [8] N Jindal, B Liu. *Opinion spam and analysis*. In International Conference on Web Search and Data Mining. 2008: 219-230.
- [9] N Jindal, B Liu. *Analyzing and Detecting Review Spam*. In ICDM. 2007: 547-552.
- [10] CL Lai, KQ Xu, RYK Lau, Y Li, L Jing. *Toward a Language Modeling Approach for Consumer Review Spam Detection*. In e-Business Engineering (ICEBE), 2010 IEEE 7th International Conference on. 2010: 1-8.
- [11] M Ott, Y Choi, C Cardie, JT Hancock. *Finding deceptive opinion spam by any stretch of the imagination*. In Meeting of the Association for Computational Linguistics: Human Language Technologies. 2011: 309-319.
- [12] G Wu, D Greene, B Smyth, P Cunningham. *Distortion as a Validation Criterion in the Identification of Suspicious Reviews*. In The Workshop on Social Media Analytics. 2010: 10-13.
- [13] G Wu, D Greene, P Cunningham, Draig, *Merging multiple criteria to identify suspicious reviews*. In ACM Conference on Recommender Systems. 2010: 241-244.
- [14] M Ott, C Cardie, J Hancock. Estimating the Prevalence of Deception in Online Review Communities. *Eprint Arxiv*. 2012: 201-210.
- [15] YF Qiu, JK Wang, LS Shao, DY Liu. Research on Product Review Spammer Detection Based on Users' Behavior. *Computer Engineering*. 2012; 38: 254-254.
- [16] N Jindal, B Liu, EP Lim. *Finding unusual review patterns using unexpected rules*. In ACM Conference on Information and Knowledge Management, CIKM 2010. Toronto, Ontario, Canada. 2010: 1549-1552.
- [17] A Mukherjee, B Liu, J Wang, N Glance, N Jindal. *Detecting group review spam*. In International Conference Companion on World Wide Web. 2011: 93-94.
- [18] A Mukherjee, B Liu, N Glance. *Spotting fake reviewer groups in consumer reviews*. In International Conference on World Wide Web. 2012: 191-200.
- [19] G Wang, S Xie, B Liu, PS Yu. *Review Graph Based Online Store Review Spammer Detection*. In Proceedings of the 2011 IEEE 11th International Conference on Data Mining. 2011: 1242-1247.
- [20] XJ Lin, JJ Wei, Y Liu, XW Miao. Research on review spammer detection based on review graph. *Journal of Fuzhou University: Natural Science Edition*. 2015; 43: 170-175.
- [21] RL Oliver. Expectancy Theory Predictions of Salesmen's Performance. *Journal of Marketing Research*. 1974; 11: 243-253.
- [22] AH Osman, N Salim, A Abuobieda. Survey of Text Plagiarism Detection. *Computer Engineering & Applications Journal*. 2012; 1.