J. Management and Humanity Research
Vol. 3, 2020, 9-15
ISSN: 2582-7766 (online)
Published on 15 July 2020
www.researchmathsci.org
DOI: http://dx.doi.org/10.22457/jmhr.v03a02102

Journal of
Management and Humanity Research

A Review of Research on Identification of False Reviews in E-Commerce

Hui Luo

School of Economics and Management
Chongqing University of Posts and Telecommunications
Chongqing 400065, China. Email: s190701033@stu.cqupt.edu.cn
Address: No.2 Chongwen Road, Nan'an District, Chongqing 400065, China

Received 15 November 2019; accepted 29 December 2020

Abstract. The role of user online reviews in current personalized marketing and e-commerce platforms has attracted widespread attention from academia and enterprises, and has extremely important reference value for both service providers and demanders, while the identification of false reviews is the basis for effective use of user text reviews. This paper takes the problem of identifying users' false evaluation texts as the starting point, reviews the current research status at home and abroad, summarizes and outlines the relevant identification methods based on review texts and reviewers' perspectives in this field of research.

Keywords: e-commerce; false reviews; identification

1. Introduction

E-commerce platforms provide online review systems to serve as a bridge between merchants and consumers. Consumers decide whether to buy a product by browsing review details, and product reviews become an important factor influencing consumers' desire to buy. Since consumers tend to buy products with positive reviews and abandon the purchase of products with negative reviews, many merchants in the competitive e-commerce market manipulate reviews by hiring "water army" to falsify positive reviews with pictures under their own stores, and mislead consumers' purchase decisions by making malicious reviews under competitors' stores. According to scholarly research, a group of fake reviewers [1], which refers to the presence of a certain number of fake reviewers in all reviews of various products on e-commerce platforms, may be a professional group of organizations or friends and relatives of merchants, and then collaborate with each other to post fake reviews of online products. The existence of fake reviews interferes with the authenticity of product descriptions and has a significant negative impact on e-commerce platforms and consumers. Therefore, it is important to identify fake reviews and provide protection for consumers' rights.

The motives of fake reviewers are different from those of real reviewers. Usually, real reviewers express their usage and after-sales experience by rating and evaluating products; whereas the motive of fake reviewers is to promote products and malign competing merchants' products. Existing e-commerce fake review methods identify fake e-commerce reviews from consumer, merchant, customer perception, deceptive language clues, text analysis, etc., which are based on the behavioral characteristics attributes of the review text attributes or reviewers: publisher username, posting interval [2], link address [2-3], and category tags [4]. The study of problems related to fake comments was first proposed by Jindal and Liu [5], professors at U of S Chicago, in 2008, who classified fake comments into three categories: (1) False reviews, also known as deceptive reviews: False reviewers manipulate the content of reviews by hiring "water army" in order to promote goods or malign competitors' goods. (2) Only focus on brand reviews: the reviewer's review content is only related to the brand, seller, manufacturer, etc. of the goods, not the goods themselves and their services. (3) Irrelevant comments: the commenter's comment content is only related to advertising, questions, answers, etc., but not related to the information about the details of the comment. Among them, the detection and identification methods for brand-only comments and irrelevant comments are simple and intuitive; while false comments are difficult to identify and require certain technical support. The difficulty of false comment identification lies in finding effective feature attributes to distinguish false comments from real ones. Recently, scholars have also proposed some new identification methods. For example, Li and Wu [6] used a hierarchical detection mechanism in identifying fake comments; Han [7] proposed a fake comment recognition model based on convolutional neural network; Zhang [8] proposed a detection method of dynamic and static feature fusion and combined with PU-learning classifier to achieve the detection of fake comments. The research results show that the above methods have good results. In order to better understand and apply the recognition methods of false comments, this paper reviews the related literature.

2. The concept and performance of false reviews

Using the literature analysis tool SATI3. 2, Zhu Juan, a domestic scholar, found that although the conceptual distinction between "fake comments" and "spam comments" is vague, the essence of the study is basically the same, and the concept is defined as untrue and deceptive comments posted through the Internet [9].

Kugler, Mukherjee argued that the main reason for the formation of false reviews is due to the false reviewers' attempting to influence the decisions of potential purchasers [10]. Chun-Dong Cheng et al. argued that the motives of fake reviews mainly lied in 3 categories: publicity and momentum, superiority against poor quality, and malicious slander [11].

Chen believes that the publishers of fake reviews can be divided into professional fake reviewers, general fake review publishers and normal reviewers [12], and summarizes the paths of fake reviews into five: （1）Normal reviewers - normal reviews. Consumers make objective and true reviews of their feelings after receiving the product, but there is also another situation where the product may change for better or worse over time, resulting in the previous reviews not matching the current after-sale experience of the product, thus misleading the subsequent consumers' decisions. （2） Normal Reviewers - Over/Under Reviews. Determined by the character traits of normal reviewers, those who are accustomed to tolerance and understanding will give the product a good review, while those with an overly demanding personality will give the product a bad review. (3) Merchants - general false reviewers - excessive reviews. On the one hand, merchants through the "good reviews cash back" or "bad reviews threaten harassment" and other means to tempt consumers to give too high evaluation; or to harass phone calls, messages and other means to force them to delete, modify bad reviews or additional good reviews. (4) Merchant - general false reviewer - too low comments. Can be divided into two categories, the first is the merchant and the consumer has contradiction and discord, which results in consumers intentionally giving too low reviews to retaliate against the merchant; the second is after the purchase of rival goods, the merchant deliberately gives bad reviews of competitors to make profit through malicious competition. (5) Businesses - intermediaries - professional fake reviewers - by hiring "water army" to cause too high or too low evaluation.

3. Fake reviews identification method

The current research on the identification of fake comments is mainly carried out from two perspectives: the text-based perspective and the reviewer-based perspective. Feature selection is a very important step, and most of the reviewed literature is based on multiple features or a combination of features to identify false comments, so I will compare the identification of false comments from two perspectives.

3.1. Recognition based on the perspective of comment text

The feature attributes of comment text content are more conducive to identification studies from the applicability and its text content also has a high value for identification analysis, generally by using relevant methods from linguistics, computer science, and statistics disciplines to establish a model for false comment identification. For example, Deng [12] et al. implemented an online deception recognition system for comment texts based on deceptive language, by extracting and analyzing features such as word frequency, information richness, and content conviction. They applied these feature sets to three previously used classifiers (SVM, NB, and C4.5), and used 5-fold cross-validation. The final experimental results proved that the accuracy of identifying deceptive comments was solved by 80%. A number of other studies have also used the recognition of lexical and n-means grammars in comment texts. Ott [13] et al. obtained a 90% check accuracy rate by using 80 sentiment feature keywords constructed from mono- and binary grammars combined with psychology.

3.1.1. Review usefulness

Previous research shows that the definition of review usefulness refers to whether the product reviews on e-commerce platforms have perceived value to consumers' decision-making process, or in another way, whether consumers' subjective views on product reviews written by other finished buyers are helpful to their own purchase decisions. Some scholars have also analyzed the usefulness of reviews through algorithms, such as information gain algorithms. Review usefulness can be studied from three aspects: the product itself, reviewer characteristics, and review text characteristics.

3.1.2. Reliability model

A logistic regression model is used to quantify the trustworthiness of reviewers and merchants.

(i) The availability of reviewer credibility models:

The difference between review ratings and overall review ratings of merchants can be quantified by calculating the magnitude of the difference between all review ratings of users and the overall ratings of the relevant merchants; the similarity of review ratings for the same merchant is quantified by quantifying two characteristics, namely, first, how many times users have commented on the same merchant, and second, the magnitude of the difference between different reviews.

(ii) Merchant credibility available:

I. Single-example review ratio to analyze, i.e., by using in the evaluation of merchant trustworthiness, this feature is denoted as Sn, that is, the ratio of the number of single case reviewers to the number of all reviews in all product reviews of N merchants.

II. The percentage of burst reviews means that the frequency of posting merchant reviews is sometimes not uniform. At certain special points in time, the cumulative number of merchant reviews will exceed the usual number. For example, peak reviews will occur during holidays, sales or group purchases. However, during peak times, there are also many merchants who hire "watermen" to enhance their reputation or to diminish the reputation of their competitors. Therefore, observing and recording the number of merchant product reviews over time can have a significant impact on building a merchant credibility model. The recorded metric can be labeled as Rn, which is the ratio of the number of reviews posted during peak hours to the number of all reviews for all products from n merchants.

3.1.3. Text analysis (analysis of feature mining and extraction of text)

I. n-gram features

Bag-of-words feature. For a text, only the known occurrences of words are concerned, ignoring their word order and word structure, and the occurrence of each word in the text is represented independently, by counting the number of occurrences of the word or phrase in the document and the frequency of occurrence. The feature is based on the assumption that the occurrence of the Nth word is only related to the previous N-1 words and not to any other word, and that the probability of the whole sentence is the product of the occurrence probabilities of the individual words. These probabilities can be obtained by counting the number of simultaneous occurrences of N words directly from the corpus. Commonly used are the binary Bi-Gram and the ternary Tri-Gram.

II. Lexical features

The feature is the subdivision of sentences, lexical annotation and counting the frequency of occurrence of words with different lexical properties. Scholars such as Li [14] found that real and fake comments showed some differences in lexicality, such as real comments contained more adjectives, nouns, conjunctions, prepositions, and qualifiers, while fake comments contained more verbs, adverbs, pronouns, and prepositional qualifiers. However, the reviews written by professional "water army" do not satisfy this conclusion, as they are more purposeful in imitating real reviews, and will imitate the characteristics of real reviews in details of products and consumption experience, which is more confusing. There is a certain variability in the distribution of lexicality in review texts in different domains, i.e., the distribution of lexicality in texts is related to text types. However, for the analysis of fake review texts, the feature of lexicality is better than the n-gram feature.

III. LDA extraction analysis

LDA model is a document topic generation model, also known as a three-layer Bayesian probabilistic model, which contains a three-layer structure of words, topics and documents. It is a common topic model used in text mining. The so-called generative model means that we consider that each word in a text is obtained by the process of "selecting a topic with a certain probability and selecting a word from the topic with a certain probability". As shown in Figure 1, document-to-topic follows a polynomial distribution, and topic-to-word follows a polynomial distribution. Through topic modeling, the potential topics of the text are obtained, and the distribution of the study text and the real text on different topics is found.

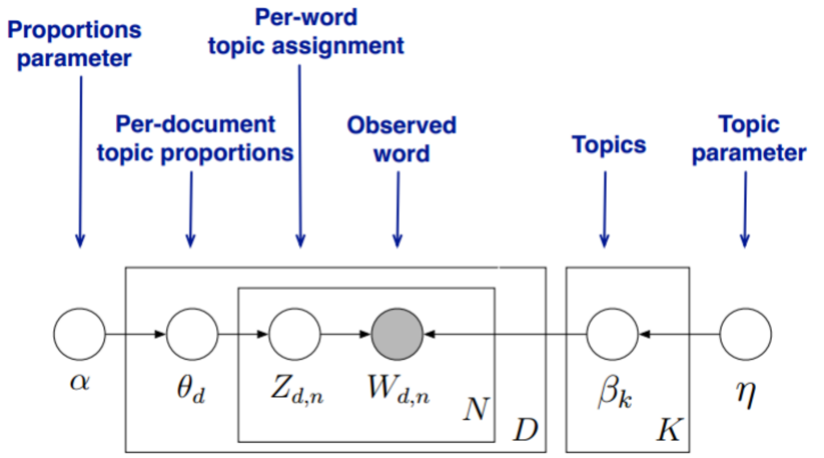

Figure 1: LDA specific model diagram

IV. Deep learning methods - textCNN text classification

Compared with previous less complex machine learning methods such as NB and SVM, convolutional neural networks achieve better results in sentiment analysis, especially in the case of larger datasets, and CNN does not require us to extract features manually. Previous shallow machine learning methods require text feature extraction, text feature representation, normalization, and finally text classification. When using CNN for text classification, the original text is first preprocessed, mainly by-word separation, deactivation, etc., and then the preprocessed text is vectorized. When building the model, the comment text data set is represented as a 600-dimensional word vector; after transforming into a word vector, each sentence can be transformed into the form of a matrix, which is very similar to the use of CNN for image classification. Deep learning is promising for text classification performance optimization.

3.2. Identification based on the reviewer's perspective

The analysis of false reviews from the reviewer's perspective identifies mainly focused on the management domain. This perspective study avoids the interference of fake comments that can imitate real comments. Fake commenters are further identified by identifying the way and characteristics of commenting behavior of fake commenters that are different from ordinary users and by using some methodological techniques to identify fake commenters.

Shao [15] argued that there were differences in emotional polarity between fake reviewers and ordinary reviewers, so the emotional features of review texts were added to the recognition model and combined with the relationship between users and goods to construct a polygraph model for identifying fake reviewers to calculate the unreliability score of users to identify fake reviewers.

Meng and Ding [16] summarized the posting motives of false reviewers in four aspects: sales promotion, denigration, interference, and meaninglessness, and pointed out the possible falsification behaviors as well as hidden behaviors of the publishers of false reviews for the current situation of Chinese Internet trading platforms.

Xiaoting et al. proposed a new identification method for detecting fake reviewers of online products based on review graphs [17], which established a mutually constraining review graph structure through the mutual influence relationship among reviews, reviewers and stores, where the characteristic indicators to measure the degree of cheating include: the credibility of reviews, the fidelity of reviewers and the reliability of stores.

4. Conclusion

This paper summarizes the existing false comment identification methods based on the comment text and commenter perspectives. In particular, the principles and models of identification methods are introduced from the analysis of review usefulness, credibility model and feature mining and extraction of text. Many merchants of e-commerce platforms deliberately manipulate in-money reviews in order to get more benefits in the process of competitive incentives, so that the online shopping platform is filled with a large number of false reviews, leading consumers to make wrong judgments and decisions, therefore, how to use efficient identification methods to identify the deceptive nature of false reviews on e-commerce platforms is the next problem to be solved.

REFERENCES

1. A.Mukherjee, B.Liu and N.Glance, Spotting fake reviewer groups in consumer reviews, *The 21st International Conference on World Wide Web*, NewYork, ACM, 2012, 191–200.
2. N.Jindal, B.Liu, and E.P.Lim, Finding unusual review patterns using unexpected rules, *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, ACM, 2010.
3. A.Mukherjee, B.Liu and J.Wang, Detecting group review spam, *Proceedings of the 20th International Conference Companion on World Wide Web*, ACM, 2011.
4. G.Wu, D.Greene, B.Smyth, and P.Cunningham, Distortion as a validation criterion in tne identification of suspicious reviews, *Proceedings of the First Workshop on Social Media Analytics*, ACM, 2010.
5. N.Jindal and B.Liu, Opinion spam and analysis, *Proceedings of the International Conference on Web Search and Data Mining*. Stanford, USA, 2008, 219-230.
6. J.Li, G.Wu, F.Xie, X.Yao, J.Qi and P.Sun, Research on the identification model of false reviews in life consumption platform, *Journal of Electronics*, 44(12)(2016) 2855-2860.

7. M.T.Han, False comment recognition model based on convolutional neural network, *Science and Technology Innovation*, 6 (2018) 78-79.
8. C.Zhang, *Research on the key technology of false comment detection for e-commerce*, Dissertation of PLA Information Engineering University, 2015.
9. J.Zhu, A review of research on key issues of false reviews of online goods, *Modern Intelligence*, 37(05) (2017)166-171.
10. H.Nie and J.Wang, A review of product review spam identification research, *Modern Library and Intelligence Technology*, 2 (2014)12.
11. L.Kugler, Keeping online reviews honest, *Communications of the ACM*， 57(11) (2014) 20-23.
12. A.Mukherjee and V.Venkataraman, *Opinion Spam Detection: An Unsupervised Approach using Generative Models*, UH－CS －TR－2014， 2014.
13. Y.Chen and C.Lou, Research on the formation path of fake reviews of online goods, *Modern Intelligence*, 1 (2015) 49-53.
14. M.Ott, Y.J.Choi and C.Cardie, Finding deceptive opinion spam by any stretch of the imagination, *49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2011, 309-319.
15. J.Li, M.Ott and C.Cardie, Towards a general rule for identifying deceptive opinion spam, *Proceedings of the Meeting of the Association for Computational Linguistics*, Baltimore, USA, 2014, 1566-1576.
16. Z.Shao and D.Ji, Identification of fake reviewers based on sentiment features and user relationships, *Computer Application and Software*, 5 (2016) 158-161, 172.
17. M.Meng and S.Ding, Analysis of the behavioral motives of publishers of fake product reviews, *Intelligence Science*, 10 (2013)100-104.
18. X.Xu, J.Wei and X.Liao, Research on the identification of spam commenters based on commenter relationship, *Journal of Jimei University: Natural Science Edition*, 21(2) (2016) 146-152.