



Research Paper

## Online Review of Data Mining Applications

Zhi Liping Chen Huanxi

School of Computer and Information Engineering, Anyang Normal University, Anyang, China  
Correspondence: Zhi Liping, School of Computer and Information Engineering, Anyang Normal University,  
Anyang, China. Tel: 1-350-372-8754.

**ABSTRACT:** The rapid development of Internet technology provides us with rich channels for online communication and to express their views, such as Weibo, online news media, online shopping platform, through these channels, a large number of user comments emerged, these user comments include users on certain things, or contains product performance evaluation, people by viewing these content, can understand the information beneficial to themselves, so as to make very high information of valuable decisions. However how to dig valuable information from the amounts of user comments is not an easy thing. Review data mining provides a viable research path for this conundrum. Comment mining is a popular research direction involving natural language processing, machine learning, data mining and other technologies, which has been widely used in public opinion analysis, Internet online advertising, recommendation system and other fields. We give detail in this paper the product review mining that has already shown great commercial value, especially that based on topic models

**KEY WORDS:** online review data mining; product; application

Received 08 Dec, 2021; Revised 21 Dec, 2021; Accepted 23 Dec, 2021 © The author(s) 2021. Published with open access at [www.questjournals.org](http://www.questjournals.org)

### I. INTRODUCTION

The highly developed Internet has provided us with a colorful, fast and comfortable way of life. However, the information explosion brought by the highly interactive network communication has also brought us trouble to a large extent. Due to the characteristics of information diversification, multifarious, massive and fast update speed, how to find hot information and valuable information in a short time has become an increasingly difficult problem for people. Therefore, how to carry out effective data processing and analysis in such information has become an inevitable research hotspot.

#### 1.1 Research background and significance

In the modern society with the high development of information technology, many technological reforms have brought great convenience and fresh and free life experience to people. Online shopping sites that allow people to buy desirable products without leaving home; Network media is a field of considerable competition between the Internet and the media, and the speed of information update is even faster. It provides people with the latest and most timely news and current affairs. After several years of development, microblog has greatly affected people's way of life. Many netizens have become accustomed to following celebrities or friends, knowing their status, and publishing their status through microblog, etc. Social networking sites also provide a free platform for people to share. All of the above ways of information exchange provide users with the function of Posting comments and opinions. Online shopping buyers usually buy a commodity in time in addition to checking the goods from the product information, will check the other has bought the goods from the buyer of goods to make evaluation, so as to decide whether they buy the goods, especially for the high price of goods, usually people are not buying experience and experience with, Browsing the experience and evaluation of buyers has become a very important way to obtain information. In addition, sellers on shopping platforms also attach great importance to buyers' comments, and even encourage buyers to make good comments through many offers or promotions. It is well known that the updating speed of network media is fast. However, the latest information may not be the information that people pay the most attention to and care about. At this time, netizens' comments and opinions as well as search frequency become an important way to discover hot information. As long as a star's micro-blog is a little sensitive, it will immediately become a hot spot for his fans

or others to forward and comment on it, and even affect the trend of the whole society. For example, recently it was very popular to apologize and cherish Ma Yili's article, which made many people imitate each other. The publication of some blog posts on social networking sites sometimes causes many people to express their views and attitudes, which also has a certain impact on social values and social ethos. Technological innovation has made people's lifestyles more diverse, and as hard drives have become cheaper and cheaper, it is no longer difficult to save data, which has contributed to the information explosion in today's society. We may have such experience: when we buy goods on shopping websites, we often lose patience because of the huge amount of information, or we cannot find the information we care about. Whether in the social networking site or on twitter yourself forward, or someone else's comments or views, in a few minutes even after a few seconds to run without a trace, the vast majority of cases users see only a fraction of the content, to understand the whole tendency of comments and content, only by manual browsing even hard to complete.

The collision of advanced technological means and innovative business model has led to the emergence of various e-commerce platforms with their own characteristics in the field of e-commerce. The two more successful models should be online shopping platform and online group buying platform. In both platforms gathered a large number of businesses, with countless goods and services, in order to better serve the consumers, also in order to improve the traffic of their platform, they are retaining customers through a variety of ways, such as online score, online reviews, online recommendation, promotional advertising, etc., the score and comments directly to consumers shopping online provides useful help. Online rating is similar to the "star rating" on Amazon, which enables consumers to know the advantages and disadvantages of a product on the whole, but cannot provide more detailed information. Online reviews to make up for the shortcomings, many buyers are buying the goods or enjoy the services of a merchant, will actively or were invited at the online comments section of the website for the goods or services to make your own comments, consumers can express their feelings from each aspect, such as comments on the characteristics and performance of goods, or comment on service attitude, delivery speed, etc., Other consumers who want to buy the product or experience the service can read the comments and make a decision about whether to make a purchase. As the number of consumers to buy goods will continue to increase, the number of comments will also increase, especially popular goods or services will have hundreds of comments, thousands of comments, consumers generally do not have the patience to browse all the information. What is more troubling is that each consumer's own user experience and the characteristics of the products they pay attention to are not the same, so the form and content of comments will be more or less different.

## **1.2 Research Status**

Comment mining is a through the use of data mining, machine learning, natural language processing technology, on the Internet a lot to the text form, structure, rules of user comments were analyzed, and its purpose is to realize the comments corpus building, product characteristic, the extraction of the user view of extraction and classifying product features and user view area of research. The products in review mining not only refer to physical objects such as mobile phones, TELEVISIONS and books, but also restaurants or some services and entertainment methods, such as KTV. The granularity of product features can be large or small. For example, we can take a restaurant as a product, and its certain dishes, such as barbecue, stir-fry, cake, etc., can be featured. It can also use its cake as a product, so that the bread, bagels, strawberry cake, apple pie and so on in the cake category can be featured. The user's opinion or emotional tendency of the product contained in the comment statement or comment snippet is the opinion expressed by the user, which can be divided into positive (good) and negative (bad). Although review mining is a relatively new research field, it has attracted extensive attention in both commercial and scientific research fields due to its broad application prospect. Comment mining is an interdisciplinary research field of data mining, machine learning, natural language processing and other technologies, so with the development of these technologies, the development of comment mining technology will inevitably have corresponding changes.

## **II. ONLINE REVIEWS**

Online reviews are a special and very important situation of online word of mouth. Word of mouth was first defined by JohanArndt in 1967. It refers to the verbal communication between two or more people for non-commercial purposes regarding goods, services or brands. Then the growth of the Internet introduced the concept of online word of mouth. Defined by Gelb and Johnson in 1995, online word of mouth transforms offline verbal communication into Internet information exchange. Chatterjee came up with the concept of online reviews after discovering that consumers who buy online refer to reviews from people who have previously bought the same product. In the past, many scholars regarded online word of mouth and online reviews as the same concept, but in fact, online word of mouth has a wider scope than online reviews. As Bickart and Schindler found in their study, online reviews are a special form of online word of mouth that has great influence and can be stored for the longest time. Park and Kim proposed in 2009 that online reviews are all positive or negative

expressions of online products by past, current and potential future users, which are obtained by other users from the review community.

### 2.1 Online comments have the following four characteristics

- (1) Large quantity. Due to the cross-time and convenience of the Internet, any user can comment on any business in any review community at any time and place. Some hot businesses or products may even generate a large number of comments in a very short time.
- (2) The subjectivity of reviewers. Online reviews are based on users' experience and situational experience, so their feelings on products and services vary from person to person, with a strong subjective color.
- (3) The text is short and concise. Some review communities set word limits on comments to prevent advertising and comment spam, and because reviewers are rarely motivated to write lengthy comments without being paid, online comments tend to be short and concise.
- (4) Redundant comments. On the one hand, the same user has certain similarities in their feelings for similar products and services; on the other hand, different users have certain similarities in their feelings for the same products and services. Therefore, online comments tend to cause a certain degree of redundancy.

### 2.2 Online review utility study

The utility of online reviews refers to the product reviews that are valuable to users and can help users make correct purchase decisions and avoid wrong consumption decisions. There are some differences between online review utility research and online review credibility research. Online review utility research focuses on the usefulness of reviews, while credibility research focuses on the identification of authenticity of reviews. At present, the research on the utility of online reviews at home and abroad mainly focuses on pre-variable research and post-variable research, and there are also a few studies on the utility of online reviews itself. In this part, this paper mainly lists the research on the ranking of the utility of online reviews.

## III. HYBRID RECOMMENDATION ALGORITHM BASED ON USER PREFERENCE AND PRODUCT CHARACTERISTICS

The hybrid recommendation algorithm based on user preference and product features improves the measurement method of user preference weight and product attribute weight. By mining online review information, users' evaluation of the product as a whole and each attribute is extracted to calculate users' preference degree for different product features. Based on the extraction of feature attributes of the product, the weight of each feature attribute of the product to the whole product is calculated to determine the importance of each feature to the product. Hierarchical user preferences and product features based on online review information mining can better recommend products for users, making the recommendation results more accurate and efficient.

Among them, the user preference model and product feature model are the two key modules of the hybrid recommendation algorithm. The user preference model reflects the user's preference weight of the product attributes from the whole to the local, and the product feature model reflects the contribution of each product attribute to the whole product. Through the user preference model, users can obtain the overall prediction and evaluation of unevaluated products as well as the prediction and evaluation of attribute characteristics. Based on the overall prediction and evaluation of products, users can recommend products, but the next recommendation cannot be realized through the prediction and evaluation of attribute characteristics. This hybrid model through prediction evaluation of the product feature model will attribute of the reduction for the prediction of the whole product evaluation, through product feature model reduction of the overall evaluation and product is obtained by the user preference model prediction of overall prediction evaluation, this paper compares and analyzes selecting two groups of prediction evaluation in the same or similar products to recommend to the user, This can greatly ensure that users are satisfied with the overall and partial attributes of the product and improve the recommendation effect.

### 3.1 Modified cosine similarity algorithm

The modified cosine similarity algorithm corrects the defect of different users' scoring differences due to different scoring scales by subtracting average scores of all items, making the algorithm more accurate. The calculation formula is as follows:

$$S_{uv} = \frac{\sum_{a \in I_{uv}} (r_{ua} - \bar{r}_u)(r_{va} - \bar{r}_v)}{\sqrt{\sum_{a \in I_{uv}} (r_{ua} - \bar{r}_u)^2} \sqrt{\sum_{a \in I_{uv}} (r_{va} - \bar{r}_v)^2}}$$

### 3.2 Similarity algorithm based on user attributes

In addition to the user rating information, the user's own attribute information is also a very important factor in finding the nearest neighbor of the target user. There are thousands of new registered users on the online shopping platform every day, and they have not evaluated any products. At this time, the user cold start problem will occur, and the modified cosine similarity algorithm will lose its effect. In this paper, the user attribute is used to find the nearest neighbor of the user. Suppose there are  $n$  attributes in the user attribute set, then the attribute set of user  $u$  is  $C_u = \{c_{u1}, c_{u2}, \dots, c_{un}\}$ , where  $c_{ui}$  represents the value of user  $u$  on the  $i$ th attribute. The value of each attribute is defined according to the category of the attribute. For category attributes, for example, gender, female = 0, male = 1; For numeric attributes, such as age, set 1-10 years = 0, 11-20 years = 1, and so on, and assign the same values to other user attributes. The similarity calculation formula of user  $u$  and  $v$  based on user attributes is as follows:

$$S'_{uv} = \alpha |c_{u1} \cap c_{v1}| + \beta |c_{u2} \cap c_{v2}| + \dots + \lambda |c_{un} \cap c_{vn}|$$

## IV. CREDIBILITY RESEARCH OF ONLINE COMMENTERS

### 4.1 Research on spam commenter identification

Spam commenter identification research originally belongs to the category of information source credibility research. Since 2007, more and more scholars at home and abroad have paid attention to this topic. There are two types of spammers: positive, overstated reviewers for promotional purposes; Malicious publishers of negative reviews with the intent of discrediting. Domestic and foreign scholars mainly identify whether a garbage reviewer is a garbage reviewer through the behavior model of the reviewer. For example, LimEPetal (2010) proposed four behavior patterns of garbage reviewer: whether there are multiple comments on the same product; Whether there are multiple reviews for different products in the same product group; Whether there is overall scoring bias behavior; Whether there is early bias in rating behavior. On the basis of Lim's research, Qiu Yunfei et al. (2012) added the behavior pattern of spam reviewers "is there any behavior that reviews products far more than purchases?", and conducted linear regression modeling according to each behavior pattern, and calculated the overall score of users by scoring each model. The higher a user's score, the more likely that user is to be a spam reviewer.

### 4.2 Research on the value of reviewers

The reviewer value is the user value of the review community, which emphasizes the user's contribution to the community. The concept of user value was first proposed by Hughes in his study in 1994. Three indicators, Recency, Frequency and Monetary, were adopted to measure the importance of offline customers and their value to merchants. With the deepening of research, more and more scholars have improved the RFM model for different research objects and research environments. Liu wei and ding zhihui (2012) proposed the LAT model based on the RFM model in combination with the characteristics of users in interest-based communities, and selected three indicators (LastTimetoLogin, frequency and totalpoints) to classify user value. Zhao Meng and Qi Jiayin (2014) combined the characteristics of review community users and added "positive comment proportion (P)" on the basis of the traditional RFM model, and compared the new RFMP model with the RFM model. It is found that the new model can better divide review community users into four categories: high-value users, low-value users, traditional users and word-of-mouth users, so that community operators can put forward more targeted user operation strategies.

### 4.3 Research on online comment opinion leader mining

The concept of opinion leader was first proposed by Lazarsfeld in the 1960s and defined as the most influential person in the process of information dissemination. With the development of Web2.0, online comments have produced network opinion leaders. Compared with traditional opinion leaders, network opinion leaders break through the boundaries of communication and exert important influence through online comments.

In recent years, scholars mainly adopt two methods to study opinion leaders: the first is social network analysis, which uses relevant indicators of centrality to dig out the central node in the communication network, namely opinion leader. CAI Shuqin et al. (2013) used the degree centrality in social network analysis to judge whether reviewers in review communities are opinion leaders, and added emotion indicators on the basis of the traditional RFM model to build the RFMS model, which can effectively identify the influence of online reviewers. The other is to use PageRank algorithm to rank users and find opinion leaders by ranking results, but PageRank algorithm is sensitive to noise. In view of various problems of this algorithm, Lv et al. (2011) proposed to use LeaderRank algorithm to quantify the influence of users. Experimental results show that LeaderRank is superior to PageRank algorithm in ranking effectiveness and algorithm robustness. Based on the LeaderRank algorithm, Xu Junming et al. (2015) added reviewers' emotional tendency and reviewers' activity

index, and the improved LeaderRank algorithm had higher accuracy and anti-noise ability compared with the algorithm proposed by Lv.

## V. CONCLUSION

The development and popularization of the Internet make people more and more accustomed to and dependent on the Internet. Especially in the Era of Web2.0, people rely on review communities to know all kinds of information of businesses without leaving home. Users of these review communities are both consumers of information and producers of information. With the advent of the era of big data, the amount of information on the Internet has increased sharply and become uneven. The operators, businesses and users of review communities are all faced with the problem of "information overload". Information overload is a severe challenge to review communities whose core values are comments and information, and a huge trouble to businesses and users who want to acquire valuable content quickly and efficiently. Based on this background, this paper is devoted to the research on the mining model and method of efficient community comments, and verifies and analyzes it, and puts forward effective countermeasures.

## REFERENCES:

- [1]. Zhang Zikun. How do online review platforms affect people's consumption behavior? A study on the influence of information society [D]. University of Science and Technology of China,2010.
- [2]. Lu XIUying. The communication characteristics of Third-party review websites in China from dianping [J]. Publishing & Printing,2011,(2):23-27.]
- [3]. Hong Jiangtao, Chen Liuyin, Huang Pei. The influence of third-party review websites on brand image and consumer behavior of catering enterprises: A case study of Dianping [J]. Finance and Trade Economics,2013,(10):108-117.]
- [4]. Dong DANfeng. Research on influencing Factors of the Usefulness of Community comment [D]. East China Normal University,2013.
- [5]. Zhang Y, Zhang Y, Zhang Y, et al. The influence of online word-of-mouth information on consumer purchasing decision: AN experimental Study [J]. Economic Management,2008(22):36-42.]
- [6]. CAI CHUNjiang. An empirical study on the influence of product brand on the usefulness of online reviews [D]. Chongqing University,2013.
- [7]. Zhang Li. Research on Customer Participation Motivation and Review Effectiveness of Online Reviews [D]. Nankai University,2011.
- [8]. Shi Xu. Research on the Usefulness of Experiential Product Reviews [D]. Huazhong University of Science and Technology,2011.
- [9]. Yan Qiang, Meng Yue. A study on the perceived usefulness of online film reviews [J]. Management Science in China,2013,S1:126-131.