Journal of Mathematics and Informatics
Vol. 16, 2019, 41-52
ISSN: 2349-0632 (P), 2349-0640 (online)
Published 19 April 2019
www.researchmathsci.org
DOI: http://dx.doi.org/10.22457/jmi.135av16a4

Journal of
Mathematics and Informatics

Difference Analysis of Online Reviews for Search-Type and Experiential-Type Products: Based on Text Mining Method

Li Wei[1] **and** ***Huang Xia***[2]

School of Economics and Management
Chongqing University of Posts and Telecommunications
Chongqing-400065, China. [1]E-mail: Liweicem@cqupt.edu.cn
[2]Corresponding author. Email:13101257308@163.com

Received 1 March 2019; accepted 10 April 2019

Abstract. Using text mining technology to analyze the differences in the emotional dimension of online reviews of search-type products and experiential products. The results show that the overall satisfaction level and emotional polarity of search-type products are significantly higher than that of experiential products, and there is no significant difference between the emotional strength of search-type products and experience-type products. Finally, this study proposes operational recommendations for the conclusions.

Keywords: experience product; search product; online review; text mining

AMS Mathematics Subject Classification (2010): 62B15

1. Introduction

Online commentary, it means the Internet words of mouth. Refers to the subjective or objective, recommended or derogatory evaluation written by the online consumer based on his or her satisfaction with the goods or services. As a third-party information source, online reviews have the characteristics of fairness, objectivity and richness, which are widely referenced and adopted by consumers. Nelson [4] classifies products into search and experience based on the source and size of the effective information. Most of the search product information comes from officially disclosed product attributes, performance information or professional data information given by industry authorities and institutions [5]. The biggest feature of experiential products is that they are greatly affected by individual differences, therefore, the information of such products is mostly derived from third-party information channels. The existing research is grouped into two levels, depending on the structural role of the product type in the online review differences. The first level uses the product type as a direct cause to study a range of different behaviors that consumers generate for online reviews of different product types. The second level takes the product type as a result and directly studies the online comment feature differences between the two types of products. A large number of studies have pointed out that the different characteristics of the two types of products make consumers have different attention to the information of various types of goods [1]. Zhang Yan-huiand Li Zong-wei [10] pointed out that consumers who purchase search-type products pay more attention to the product characteristics and attribute information disclosed by the official website, while experience-type product consumers focus on the experiential information about the personal use experience. Some researchers explored the negative impact of negative reviews on different types of products on brand impressions [3]. The results showed that consumers of experiential products had stronger feedback on negative comments, which also proved Jiménez's point of view from another angle. Some researches focus on the characteristics of the search product and experience product review text attributes (quantity, number of words, etc.), such as some researchers comprehensive consideration of Online comments specific content constructed analysis indicators for search products and experience products [2]. The results indicate search products Online comment content is more proactive and more effective for consumers, while experiential product reviews are more personalized and diverse [7]. By comparing the initial comments and additional commentary information of the products, Shi Wenhua [6] found that the commentary and comment time intervals of the experiential products were significantly larger than the search products.

The existing research on the difference of online reviews of search-type products and experiential-type products revolves around objective dimensions such as comment level and number (praise, middle, bad, review, etc.), but the emotional attitudes of consumers on these objective indicators are limited. No detailed studies have been conducted to analyze the emotional dimensions of the online review differences between the two types of products [9]. Efficiently and automatically classify and analyze user emotions for online commentary text content, summarizing consumer needs and preferences from the massive comments of information overload. In recent years, the rapid development of text mining technology has made it possible to analyze online sentiment analysis [8]. In order to dig deeper into the differences in emotional dimensions between online search and experiential products, this article classifies products into search and experience types, and uses Python crawler technology to crawl 188,594 online comment data of Taobao. Use text mining techniques and sentiment analysis methods to delve into the differences in the text of different types of product reviews.

2. Research process

2.1. Commodity classification and analysis indicators

According to the difference in the amount of information that consumers receive before purchasing a product, the product is classified into a search type and an experience type. Search-type products (such as brand computers, refrigerators, mobile hard drives, etc.) mean that product information is objective and comprehensive, and consumers can make rational decisions based on the products presented by the sellers, but the cost of search-based products is higher. Experiential products (such as food, makeup, clothing, etc.) often require consumers to personally use the product details, and the seller's product introduction information is not enough for consumers to make rational judgments on the goods. The capacity and speed of the mobile hard disk are more objective, and the chocolate only knows the taste after tasting. Therefore, this article captures the review text data separately from the mobile hard disk as the search product and the chocolate as the experience product.

The Sentiment analysis, also known as the propensity analysis, refers to the analysis, reasoning, and induction of texts with personal emotions to obtain an attitude toward something or something. This study applies text mining techniques to online reviews. Through word and sentence-level sentiment analysis methods, the text data of online commentary is used as a corpus to quantify the user's actual experience in the transaction process and the satisfaction degree of the product. On this basis, the following indicators are used to analyze the differences in online reviews of the two types of products.

(1) Overall Satisfaction. It is measured by the average number of total emotional scores of all reviews of the product. Design the sentiment analysis algorithm to find the total score of all comments on each item, and then find the average value. It is used to measure the overall level of satisfaction of the product in the minds of consumers.

(2) Emotional polarity. It is measured by the number of all emotional word scores/emotional words in the product. The sentiment analysis algorithm is designed to count the total scores of the emotional vocabulary of all the reviews of the commodity, and then calculate the ratio. It is used to measure the emotional polarity of the consumer.

(3) Emotional intensity. It is measured by the number of points/degree words of all reviews of the product. Design the sentiment analysis algorithm to count the total vocabulary scores of all reviews of the product, and then calculate the ratio. It is used to measure the emotional intensity of the consumer.

2.2. Research method-text mining

Online commentary with subjective emotional color guides potential consumers to rational decision-making to a certain extent, analyzes the emotional trend of subjective text in comments, extracts different attribute characteristics of online comment objects, and analyzes the emotional expression of different attribute features of goods. On this basis, the consumer's opinion on the product is unearthed. This paper divides the research framework of online commentary sentiment analysis into three modules according to the processing of structured objects.

(1) Process online comment text. Through the programmatic operation, the text content is unstructured to structured transformation, so that the computer language can more accurately and efficiently identify and process meaningful structural units in the text. The specific operations include: mining the original corpus of the online comment and storing it in the database, loading the word segmentation dictionary for text analysis and cutting work, loading the stop word dictionary to automatically filter the stop words to improve the processing efficiency, and matching the part of speech by the part of speech rules.

(2) Construct a research-specific sentiment dictionary based on the comment object extracted from the original corpus of the online commentary (the object modified by the attribute viewpoint words in the comment) and the commonly used Chinese sentiment word dictionary, including basic emotion dictionary, professional field emotion dictionary, degree emotion dictionary and negative emotion dictionary.

(3) According to the actual situation, the sentiment analysis algorithm is designed. In order to let the sentiment analysis model accurately display the user's emotional attitude, the researcher should set different sentiment analysis standards according to different industry characteristics and original corpus features. In the process of processing, according to the characteristics of the dictionary and the size of the data to make corresponding adjustments.(The specific text mining and analysis ideas are shown in Figure 1 below).

2.3. Data processing

This paper is based on the precise mode of the jieba dictionary for text segmentation. In the information retrieval process, in order to improve the search efficiency and save storage space to some extent, the stop words are automatically filtered out before the natural language data is processed. In this paper, after the comment text segmentation, the dictionary is disabled (the Harbin Institute of Technology stop vocabulary) to remove the stop words to get the corpus. The data processing process is the following three steps.

(1) Construction of emotional dictionary

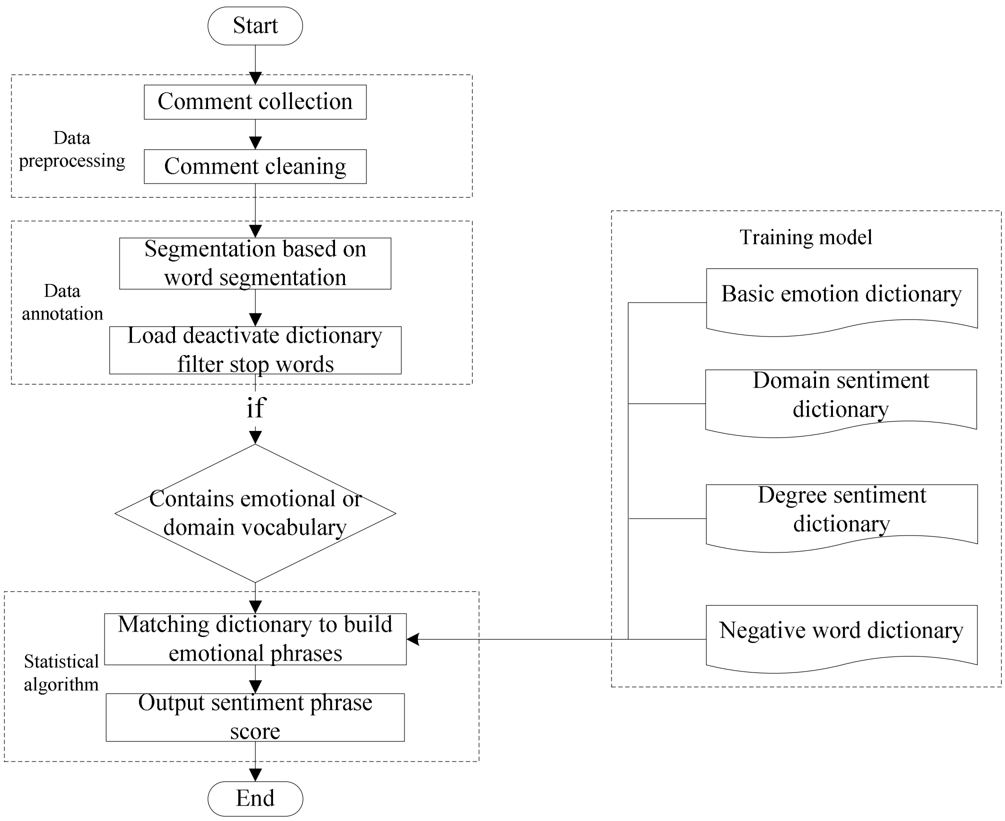

Figure 1: Text mining method and flow chart

Choosing the applicable sentiment dictionary is the key factor for accurate sentiment analysis. The online emotional dictionary divides the emotional characteristic vocabulary into two categories: positive tendency and negative tendency. Considering the simple positive or negative emotional tendency attitude is too weak, it is not enough to present the difference in the emotional dimension between the two types of products. The basic sentiment dictionary of this paper is adjusted according to HowNet's sentiment dictionary: remove uncommon words, and add some words according to the characteristics of jieba participle. The principle of assigning the scores of each word is as follows: divide the temperament intensity of all vocabulary into five grades, 5 points represent very satisfied, 1 point represents extremely dissatisfaction with goods or services, and the principle of vocabulary assignment is as shown in Table 1.

Table 1: Emotional orientation assignment principle (partial)

Vocabulary	Score
Satisfied, will buy back, like, authentic, recommended to buy, trust, beautiful, inexpensive, surprise,	5
Not bad, cheap, suitable, good, reasonable, not expensive, okay, fast,	4
A little expensive, not cheap, not very good, casual, moderate, general,	3
Bad, disappointing, hope for improvement, non-conforming product, greasy, very slow, stunned, cut corners, insufficient memory	2
Garbage, poor, poor quality, disgusting, speechless, bad reviews, scammers, fakes, pits, expired,	1

Then, considering the differences in the characteristics of the two types of products, they have professional emotional vocabulary in their own product fields. Such as "greasy", "too sweet" to express the attitude towards chocolate, and "small", "card" and other expressions on the mobile hard disk preferences. In order to accurately analyze the emotional sentiment of the comments, the respective domain sentiment lexicons were established for the two types of products. The program is designed to filter the words with more frequent frequencies from all the review texts of the two products, and then combine the manual screening to construct the domain sentiment dictionary separately. There are 48 chocolate vocabularies and 50 mobile hard disks, as shown in Table 2.

Table 2: Domain sentiment vocabulary assignment principle (partial)

Chocolate	Score	Mobile hard disk	Score
Sweet but not greasy	5	Thin and light	5
Silky	5	Heat	1
Greasy	1	Shockproof	5

(2) Construction of emotional adjustment dictionary

The emotional sentiment of emotional vocabulary in comments is influenced by its negative and degree words. Therefore, negative words and degree words are called emotional adjustment vocabulary. The appearance of negative words will reverse the embarrassment of emotional words, and the utility is superimposed. Common negative words are as follows: no, zero, very few. The degree word acts on the emotional intensity of the emotional vocabulary: “super card” and “a little card”, and the degree word in front of the emotional word “card” makes the emotional intensity different. Based on the How-Net dictionary, delete the unusual vocabulary, and add the corresponding vocabulary according to the program screening, and finally construct the degree vocabulary of the research. As shown in Table 3 below, according to each degree of word adjustment strength is divided into 5 levels, 5 points represents great adjustment, and 1 point represents small intensity.

Table 3: Degree vocabulary assignment principle (partial)

Vocabulary	Score
Very, extremely, multiply, thoroughly, absolutely, very, completely, all, incomparably	5
deep, disastrously, especially, extraordinarily, extremely, greatly, how, indeed, much	4
still more, increasingly, like that, more, more so, plus, relatively, slightly more, so	3
a bit, a bit too, a little, a little bit, a little more, fairly, more or less, passably, pretty, quite	2
a little less, just, light, merely, not particularly, not too, not very, relative, slight, slightest	1

(3) Sentiment analysis algorithm design

Emotional phrases are defined according to the previously constructed emotion dictionary (basic emotion dictionary and domain emotion dictionary) and emotion adjustment dictionary (negative word dictionary and degree word dictionary).An emotional phrase consists of the emotional vocabulary in each comment and the negative and degree words in front of it. The emotional score of a comment is the sum of all emotional phrase scores. The formula for calculating the emotional value of a comment is:

$$\text{sentiment} = \sum_{i=1}^{n} s(w_i) * (-1)^k * d(adv) \tag{1}$$

where sentiment is the total score of sentiment values of n emotional phrases in a comment, $s(w_i)$ represents the score of the i-th emotional vocabulary, and k is the number of negative words appearing before the emotional word w_i. d(adv) is the adjustment strength of the degree word that appears before the emotional word w_i. Finally, sum the n emotional phrase sentiment values in a comment.

4. Data analysis and conclusion

The experience product is denoted by the value of 1, and the search product is denoted by the value of 2. One-way analysis of variance was performed on three analytical indicators for online review of two types of commodities. The analysis results are shown in Table 4 and Table 5.

(1) The overall satisfaction level of search products is greater than that of experience products. As shown in Table 4, according to the sentiment analysis algorithm design formula, the average value of the overall satisfaction water of the experience type product is 42.384849, and the standard deviation is 12.86620798.The average satisfaction level of search products is 51.206735, and the standard deviation is 15.7759571. And the analysis of variance shows that the overall satisfaction level of search-type products is significantly higher than that of experience-oriented products. Mobile hard drives are typical search products. The merchant provides the consumer with detailed attribute information such as color, model, performance, and memory size on the product home page. And the search-type product industry chain standards, price transparency is high, consumers will have "what you see is what you get" mentality after receiving goods. Therefore, consumers have less perceived perceptions of the product before and after shopping, and thus overall satisfaction is greater. The taste of chocolate is perceived by consumers after being tasted, and in line with the characteristics of experiential products, it is seriously affected by individual taste differences. Therefore, the actual utility of the product and the expected utility are quite different, so in general, the overall satisfaction of the search product is higher than that of the experience product.

Table 4: Descriptive statistics

		N	Mean	Standard deviation	Standard error	95% confidence interval for the mean		Minimum	maximum
						Lower limit	Upper limit		
Overall Satisfaction	1	49	42.384849	12.6820798	1.8117257	38.742130	46.027567	18.8343	73.0305
	2	43	51.206735	15.7759571	2.4058109	46.351612	56.061858	25.5697	98.2667
	all	92	46.508122	14.8084690	1.5438896	43.441375	49.574869	18.8343	98.2667
Emotional polarity	1	49	4.516450	.1272577	.0181797	4.479897	4.553002	4.2909	4.7717
	2	43	4.598831	.0702306	.0107101	4.577218	4.620445	4.3961	4.6958
	all	92	4.554954	.1119226	.0116687	4.531776	4.578133	4.2909	4.7717
Emotional intensity	1	49	3.820922	.0330024	.0047146	3.811443	3.830402	3.7464	3.9042
	2	43	3.831199	.0410903	.0062662	3.818553	3.843845	3.7332	3.9346
	all	92	3.825726	.0371531	.0038735	3.818031	3.833420	3.7332	3.9346

(2) The emotional polarity of online reviews of search-type products is greater than that of experience-type products. The emotional polarity of the online review of experiential products is 4.516450, the standard deviation is 0.1272577, the emotional polarity of the search-type merchandise is 4.598831, and the standard deviation is 0.0702306.From the analysis of variance table, the emotional polarity of search-type products is significantly larger than that of experience-oriented products. This conclusion is consistent with the previous analysis. Because the experience cost of search-type goods is too high, consumers will refer to many factors before buying. At the same time, such commodity information itself is more objective and easy to obtain, and consumer cautious pre-information collection can greatly reduce shopping uncertainty. At the same time, experiential products cost less, which means that their prices are cheaper than search-based products. Different from the standard production process of search products, the quality of experience products is quite different, and many consumers have a contradiction of low price and high quality. Therefore, the emotional polarity of search products is often

significantly higher than the experience products.

Table 5: The ANOVA result

		sum of square	df	Mean square	F	Significant
Overall Satisfaction	Between groups	1782.377	1	1782.377	8.827	.004
	within-group	18173.082	90	201.923		
	All	19955.459	91			
Emotional polarity	Between groups	.155	1	.155	14.209	.000
	within-group	.984	90	.011		
	All	1.140	91			
Emotional intensity	Between groups	.002	1	.002	1.767	.187
	within-group	.123	90	.001		
	All	.126	91			

(3) There is no significant difference between the online strength of search products and the emotional products. The online emotional strength of search products is 3.820922, the standard deviation is 0.0330024, and the emotional strength of the online experience of experience products is 3.831199, with a standard deviation of 0.0410903.Analysis of variance showed that there was no significant difference between the emotional intensity of search-type products and the experience-based products. Degree adverbs act on the emotional intensity of emotional words, and use degree words (very, absolute, etc.) to fully express the emotional psychology of consumers after receiving goods, which can cause consumers to resonate to a certain extent. Consumers use the adverbs of degree and the emotional vocabulary with different polarities to express the mood after receiving the goods. That is to say, degree adverbs are related to the desire of consumers to publish text comments and have nothing to do with the overall level of consumer satisfaction with the goods. The desire of consumers to publish text comments is also affected by personality differences. Therefore, in terms of emotional intensity, there is no significant difference between search products and experience products.

5. Discussion and summary

In view of the differences in the overall satisfaction level, emotional polarity and emotional intensity of search-type products and experiential products, this article presents the following practical recommendations:

Firstly, search-type merchandise sellers should pay more attention to product display details and after-sales service. Search-type goods have the characteristics of higher cost of experience, which makes it necessary for consumers to refer to various aspects of information before decision-making to reduce uncertainty. Therefore, the seller of such products should display the product details in multiple dimensions according to the characteristics of the products sold. This paper proposes two suggestions: First, the study suggests that sellers hire professional copywriters and photo editors to provide consumers with practical and rich product details. Businesses try to ensure "what you see is what you get" as much as possible to ensure that the psychological gap between consumers before and after shopping is reduced. Furthermore, since search-type goods have a standard production process and the price is transparent. Therefore, in order to maintain and improve the competitiveness of goods, sellers of such goods should pay more attention to the quality of goods and after-sales service. Providing consumers with professional and patient customer service staff and safe and secure after-sales service can make sellers stand out among many competitors.

Secondly, experiential merchandise sellers should strive to provide personalized product experience information. In response to the traits of experiential product information that are difficult to obtain and affected by individual differences, this paper proposes two suggestions: First, sellers of such products can select trial personnel for different types of groups, and then create an information form, which contains trial personnel characteristics, preferences, and the degree of preference for the product. And displayed on the product details page for potential consumer reference. In addition, the merchant can sell the demo sample (currently used in the skin care industry), and the experience package is exactly the same as the sales dress, so that the consumer can fully understand the product before the formal purchase. Consumers' maximum understanding of product details can reduce psychological inconsistency before and after shopping, and can greatly improve consumer satisfaction with goods.

Thirdly, to improve the online comment system is of great meaning. The Taobao user comment interface is divided into two categories. The first category only has “descriptive matching” star rating and text comments; the second category includes “praise”, “medium commentary”, “bad review” and text comments. Many Taobao products sell tens of thousands but the comments are only a few thousand, indicating that the vast majority of consumers do not participate in comments after receiving the goods, or just star reviews without entering text content. At this time, the simple good, medium and bad star ratings are too broad to be able to present consumers' preference attitude towards the product, so the reference value is lower. In order to improve such star rating only without the usefulness of text comments, it is recommended to embody a simple star rating, adding a combination of degree words and emotional words as a rating option. For example, super satisfied, very like, very good, will be, not very satisfied, very dissatisfied. The emotional score of “Super Satisfaction” is far greater than “praise”, which is more likely to cause potential consumers to pay attention and resonate, thus greatly improving the usefulness of online comments.

Acknowledgements. The research is supported by the National Social Science Foundation of China (No. 17CGL059).

REFERENCES

1. F.Jia-ming, W.Yu-ying and Z.Zhi-rong, The moderation effect of product types brand reputation on the helpfulness of online product review, *Soft Science*, 3(30) (2016) 108-112 .
2. F.Xin, W.Chen and L.Yuan, The customer churn prediction based on emotional polarity and BPNN, *J. China Acad. Elect. and Infor. Tech.*, 13(3) (2018) 340-345.
3. L.Qi, R.Xiao-jing, Research on how conflictive additional reviews affect perceived helpfulness, *Journal of Management Science*, 30(4) (2017) 139-150.
4. P.Nelson, Informational and consumer behavior, *Journal of Political Economy,* 78(2) (1970) 311-329.
5. S.M.Mudambi and D.Schuff, What makes a helpful online review? A study of customer reviews on Amazon.com, *MIS Quarterly*, 1 (2010) 185-200.
6. S.Wen-hua, W.Lu and S.Na, A comparative study into the impact of initial and follow-on online comments on sales, *Management Review*, 30(1) (2018) 144-153.
7. W.Cui-cui, G.Hui, Research on factors affecting the perceived usefulness of online reviews with additional one-An eye-tracking experiment, *Journal of Modern Information*, 38(12) (2018) 70-77.
8. W.Wei and W.Hong-wei, Comparative network for product competition in feature-levels through sentiment analysis, *Journal of Management Sciences in China,* 9(9) (2016) 109-126.
9. W.Zu-hui, J.Wei and L.Yi-jun, Regular collocation features extraction method in online reviews sentiment analysis, *Journal of Industrial Engineering and Engineering Management*, 28(4) (2014) 180-186.
10. Z.Yan-hui and L.Zong-wei, Analysis of the factors that influence online reviews help fulness: based on the regulating effect of product type, *Management Review,* 28(10) (2016) 123-132.