Applied Mathematics and Nonlinear Sciences

Users' Sentiment Analysis of Shopping Websites Based on Online Reviews

Xiaohong Wang, Shuang Dong [†]

Management College, Beijing Union University, Beijing 100101
China

Abstract With the rapid development of online shopping, how to explore the value of online reviews, so as to give full play to their role in potential users' purchasing decisions. Based on text mining and quantitative analysis, this paper studies the sentiment analysis of online reviews on B2C shopping website. The main attributes of commodity or service are extracted based on the order of word frequency in the online reviews. Text analysis method is used to judge the relationship between attributes of commodity or service and its emotional words. The fine-grained sentimental polarity and intensity of attributes are identified to analyze users' concerns and preferences. The research shows that users pay more attention to the configuration and after-sales service of mobile, and have a positive sentimental orientation to most of attributes, especially unlocking function, hand feeling attribute and logistics service; and have a neutral sentimental orientation towards the attributes of battery and memory, and a negative sentimental orientation towards the membrane of mobile phone. The results can provide a reference for consumers to make purchasing decisions, for enterprises to improve product quality, and for shopping platform to optimize service.

Keywords: online review; shopping website; sentiment analysis; text mining; sentimental polarity
AMS 2010 codes: 03Bxx.

1 Introduction

The rapid development of e-commerce has deeply influenced the behavior pattern of consumers, and online shopping has become the normal way of shopping for web users currently. Because of the inability to feel the physical goods, consumers know the quality and performance characteristics of goods by IWOM (Internet Word-of-Mouth) before shopping. The shopping experience of other consumers will have an important impact on their purchasing decisions. Online reviews, as an important form of IWOM, evaluate goods in the form of text and are believed to be more trustworthy than advertising and expert recommendations [1,2].

Users' sentimental orientation of online reviews can have a greater impact on the consumption psychology and purchasing decisions of potential users [3]. Online reviews include a large amount of information about the description and use feeling of commodity attributes, as well as consumers' feedback on commodity or service.

[†]Corresponding author.
Email address: gltdongshuang@buu.edu.cn

How to analyze the sentimental orientation of these rich and complex online reviews and mine the hidden rule, in order to better provide decision-making basis for consumers, enterprises and shopping platform. It has become a hot research topic in marketing and the popular research direction in the field of text mining [4].

Based on the sentiment analysis technology, this paper directly explores the consumers' sentimental orientation to the attributes of commodity or service from comments text, analyzes the user satisfaction and preference, and then discusses how the consumers, commodity producers and online shopping platforms to manage and utilize online reviews effectively.

2 Literature Review

Sentiment analysis, also known as opinion mining, refers to the process of semantic analysis of online reviews using text mining technique to identify the consumers' attitudes and opinions of goods and services, to excavate the distribution of consumer's sentimental orientation [5, 6]. Sentiment analysis on online reviews involves some technologies, such as natural language processing, machine learning and text mining. The research mainly focuses on the analysis of sentimental polarity and strength, value mining and so on.

Analysis of sentimental polarity and intensity is mainly through discussing the text content of online reviews to identify consumers' sentimental orientation or polarity and determine the polarity intensity. According to the different granularity of text processing, the analysis of sentiment polarity is mainly divided into the fine-grained analysis and the coarse-grained analysis. The former is the analysis on sentimental polarity and intensity of feature-viewpoint based on corpus and dictionary. And the latter is the judgment of overall sentimental orientation or polarity of text based on model and machine learning. The existing researches are mostly about the analysis of coarse-grained sentimental polarity [7].

Valuable online reviews can help consumers improve their decision-making efficiency and provide additional benefits for businesses. The research on value mining of online reviews mainly concentrates upon its usefulness, impact on product sales and purchasing intention. The usefulness of online reviews is a subjective perception of whether online reviews are helpful to consumers' purchasing decisions. There may be differences in the proportion of positive sentiment and negative sentiment of different reviews and its degree of concern which leads to differences in the usefulness of reviews, and the reviews with strengths and weaknesses of product will be more useful [8–10]. The factors that can significantly affect the sales volume of online products are mainly the number of online reviews, product attention, timeliness of reviews, usefulness of reviews and attribute inconsistency, and the polarized reviews are more powerful than neutral ones [11, 12]. The influence research of online reviews on consumer's purchasing intention is mainly related to the influence factors, influence mechanism and influence power [13–15].

Online reviews often involve the commodity attributes of quality, performance, function, price, logistics and service attitude, as well as their experience and feelings of using commodity or service [16]. The emotional expression of users is also complex with affirming certain aspects and criticizing other aspects of commodity or service. To better exploit the user's attitude towards different attributes of commodity or service and deeply study its sentimental orientation, on the basis of previous research achievements, this paper attempts to use natural language processing and ontology construction technology to extract the main features of online reviews, and determine the relationship between attributes of commodity or service and its emotional words. By designing the quantitative calculation rules of sentimental polarity, the user's fine-grained sentimental orientation about the attributes of commodity or service is analyzed deeply.

This paper explains how to perform the users' sentiment analysis of shopping websites based on online reviews. First of all, it describes the research background and literature review on sentimental polarity analysis of online reviews and its value mining; next it explains the research process of sentimental polarity analysis and how to collect and process data, preprocess text content, extract the main attributes of commodity or service, and identify the relationship between attributes and emotional words; and then it provides the reference for consumers, commodity producers and online shopping platforms according to analysis results; at the end of the

full text, it comes up with the foresight and future work.

3 Research Design

The design process of study is shown in Figure 1, including major steps such as data collection, text content preprocessing, attribute extraction and identification of corresponding relation, classification and assignment of sentimental orientation, calculation of sentimental polarity intensity.
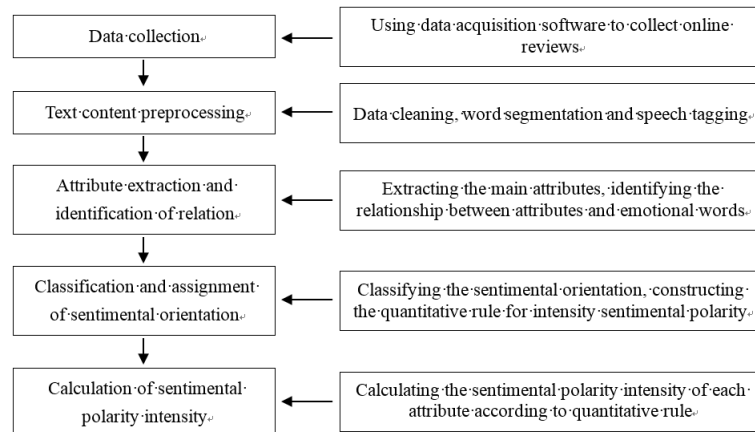
Fig. 1 The research process of sentiment analysis

3.1 Data Collection and Text Content Preprocessing

Data acquisition software is applied to collect the online information of relevant commodity and clean obtained data, and text mining system is used for text processing and data mining. The words playing a key role in Chinese sentiment analysis, such as nouns, adjectives, adverbs and verbs, should be highlighted.

Online reviews usually involve multiple attributes or services of a goods and comment on these attributes separately. The sentence with an emotional word or an evaluative word usually expresses the user's opinion about a particular attribute of a commodity or service. It is necessary to carry out word segmentation, speech tagging and sentence processing for each online review. Then the content is preprocessed, the synonyms are processed, and the low-frequency words without meaning and the high-frequency words without obvious meaning are removed.

3.2 Attribute Extraction of Commodity or Service

Attributes extraction of goods and services refers to the extraction of users' concerns on a product in online reviews, aiming to understand the users' attention to functions, performance and services of product. Attributes extraction is the basis of sentiment analysis of online reviews. Extraction method, mainly relying on the statistical theory, is to identify the product attributes based on predefined rules to count the frequency of nouns appearing in a specific domain, some attributes of commodity or service are extracted as the reviews objects according to the order of occurrence frequency and then users' sentiment analysis is carried out.

3.3 Identification of Relationship between Attributes and Emotional Words of Commodity or Service

Emotional words refer to those words in online reviews that can more clearly express users' position, attitude and feelings towards the attributes of commodity or services. By means of text analysis, such as speech tagging, the corresponding relationship between commodity attribute words and emotional words is judged according to

Table 1 The quantitative rule for the intensity of sentimental polarity

Sentimental polarity	Negative words and degree adverbs	Sentimental polarity intensity
Positive emotions	extremely, very, most	2
	relatively, slightly	1
Neutral emotions	still, right, just right	0
Negative emotions	not good, insufficient, no	-1
	very, too	-2

its co-occurrence. Firstly, attribute words are used as the reference to find its emotional word in order backward, and then emotional words are used as the reference to find its degree adverbs in order forward.

3.4 Classification and Assignment of Sentimental Orientation

Use the positive words or negative words in comments to determine the user's sentimental orientation or sentimental polarity to an attribute of commodity or service. When making the judgment of sentimental orientation, it is necessary to combine the same attributes of commodity or service with the same meaning, and classify the sentimental orientation of these attributes into the same category.

The sentimental intensity expressed by words with the same sentimental polarity may be different. For example, "good-looking" and "amazing" are the positive sentimental polarity, but the polarity intensity of the latter is obviously greater than that of the former. The degree adverb can also change the sentimental intensity of its modifier, such as the sentimental polarity intensity of "very satisfied" is significantly greater than that of "satisfied". In order to understand the differences in users' sentimental orientation towards the attributes of commodity or service, it is necessary to quantify the polarity intensity of emotional words and emotional words with degree adverbs. It is divided into five different grades according to the intensity of positive emotional words and negative emotional words, as shown in Table 1.

4 Empirical Analysis

Tmall.com, JD.com and Suning.com, accounting for more than 80% of market scale of B2C online shopping in China, are selected as the samples. These B2C shopping websites have great influence and its own characteristics, such as the marketing of Tmall.com, the supply chain of JD.com, and the offline physical stores of Suning.com. Online reviews information of popular goods is taken as the samples data to study the users' sentiment analysis of B2C shopping websites.

Huawei, leading manufacturer of mobile phones in China, has a large fan base which is good for data collection. Online reviews about Honor 8 mobile phones are abundant in shopping websites and it can well support the research of this paper.

4.1 Data Collection and Processing

Octopus information acquisition software is used to capture the comments of mobile phone in Tmall.com, JD.com and Suning.com. The time of data collection is from the release date (listed on July 11, 2016) to January 10, 2017. After deleting repeated comments from the same user, a total of 14502 online reviews about Honor 8 are obtained.

Text mining system developed by our project team is adopted to segment words, calculate the word frequency and sort by ascending order. It can be fully prepared for extracting the attributes words through text content preprocessing and updating the dictionary of disused words and synonym.

Some words (or strings) that have nothing to do with this study are added to the dictionary of disused words, such as kkkkk, x6, 58, 12, hglll and other meaningless low-frequency words, and mobile phone, Tmall, this,

Table 2 The attributes classification of commodity or services

Theme classification	High-frequency words for the attributes of commodity and services
appearance	entirety, appearance, color, cute face
function	function, photograph, playing games, video, fingerprint, unlocking
performance	performance, operation, signal, charge, reaction
configuration	memory, system, screen, battery, membrane
price	price, cost performance
quality	quality, sound quality, hand feeling
service	logistics, customer service, express, deliver goods, delivery

huawei, today and other high-frequency words without obvious meaning. Some words with the same meaning are added to the dictionary of synonym, such words as photographs, cameras, self-portraits, photographing and other words are defined as "photograph" in the dictionary.

4.2 Attribute Extraction of Commodity or Service

The high-frequency words in online reviews of shopping websites can reflect the users' concerns about goods and services. In general, the higher the word frequency of a commodity attribute or service attribute, the more the user is concerned about the property or service of the commodity. In this paper, the high-frequency words that line in the top 100 in shopping website (Tmall.com, JD.com and Suning.com) are selected. After removing the words with less obvious context, there are 30 high-frequency words about the attributes of commodity or services. Based on reviewing the related literature and combining with the characteristics of mobile product, 30 high-frequency words are divided into 7 categories in this paper, namely appearance, function, performance, configuration, price, quality and service, as shown in Table 2.

The proportion of online reviews on the attributes of Honor 8 is shown in Table 3, from which the attributes of commodity or services that users are concerned about can be learned. Users are most interested in the hand feeling attribute with the largest number of online reviews (9.69%). There are more online reviews on the attributes of commodity or services including screen, appearance, logistics, photograph, customer service, operation, express, charge, system, battery, delivery goods, and their proportion of reviews are more than 5%. Among them, the attributes such as memory, system, screen and battery belong to the configuration theme, and the attributes such as logistics, customer service, express and deliver goods are classified as service theme. It shows that users usually pay more attention to the configuration and after-sales service of mobile phones. While users pay less attention to the commodity attributes of price, sound quality, color, entirety, delivery, and the number of reviews accounts for less than 2%.

4.3 Attribute Analysis of Commodity or Service

According to quantitative rules of sentimental polarity intensity designed in Table 1, the value of sentimental polarity intensity and its mean value (N), the same attribute of commodity or services corresponding to emotional words, is calculated one by one. The mean value of sentimental polarity intensity of online reviews of Honor 8 is shown in Table 3. Most of the attributes of commodity or services show a positive sentimental orientation, i.e.,$N \geq 0.3$. There are two attributes show a neutral sentimental orientation, which is $0 \leq N < 0.3$. Only one attribute indicates a negative sentimental orientation with $N < 0$. The value distribution of sentimental polarity intensity of the main commodity attributes is shown in Table 4.

As shown in Table 4, users are most satisfied with unlocking function, hand feeling attribute and logistics service of Honor 8, and the mean value of their sentimental polarity intensity is close to 1.5 (as shown in Table 3). As for the unlocking function of mobile phone, its proportion of positive sentimental polarity is 92.8%, among which 63.79% of users report that the unlocking function is very fast, convenient and user-friendly. The positive sentimental polarity of hand feeling attribute is 96.58%, the number of users who feel very good, excellent and

Table 3 The proportion and mean value of sentimental polarity intensity of attributes

Commodity attribute	Proportion(%)	Mean of intensity
hand feeling	9.69	1.49
screen	8.59	0.87
appearance	8.38	1.38
logistics	7.90	1.46
photograph	7.23	1.08
customer service	6.87	1.24
operation	6.80	1.16
express	6.26	1.38
charge	5.86	0.82
system	5.39	0.80
battery	5.27	0.19
deliver goods	5.12	1.13
fingerprint	4.99	1.41
cute face	4.82	1.44
function	4.61	0.98
reaction	3.99	1.12
unlocking	3.35	1.51
cost performance	3.35	1.25
signal	3.12	0.83
membrane	3.05	-0.46
quality	2.94	1.12
memory	2.83	0.11
performance	2.44	1.16
playing games	2.36	0.68
video	2.27	0.95
price	1.88	0.61
sound quality	1.85	0.95
color	1.68	1.13
entirety	1.39	0.98
delivery	1.12	1.16

Table 4 The distribution of sentimental polarity intensity of the main commodity attributes

Attribute of commodity or service	Distribution of sentimental polarity intensity of the main commodity attributes (%)				
	-2	-1	0	1	2
unlocking	0.62	4.12	2.47	29.01	63.79
hand feeling	0	0.57	2.85	43.77	52.81
logistics	3.25	4.21	0.44	27.30	64.79
cute face	0.29	0.14	1.72	50.79	47.06
fingerprint	1.36	5.90	4.08	27.38	61.27
appearance	0	0.33	6.94	46.99	45.75
express	2.11	5.90	3.00	30.14	58.84
playing games	5.46	15.02	11.95	41.30	26.28
price	1.85	13.33	18.15	54.81	11.85
battery	5.43	26.36	22.78	34.57	10.86
memory	7.71	24.38	25.62	33.33	8.96
membrane	2.35	50.82	39.76	4.24	2.82

first-class account for 52.81%. The proportion of positive sentimental polarity of logistics service is 92.09%, among which 64.79% of users report that logistics is very fast, very awesome and so on.

Users are also satisfied with the attributes of cute face, fingerprint, appearance, express, cost performance, customer service, performance, operation, deliver goods, delivery, reaction, quality, photograph, function, entirety, sound quality, screen, signal, charge and system, and their mean value of sentimental polarity intensity is between 0.7 and 1.45, and their proportion of positive sentimental polarity is relatively high.

Users are satisfied with the attributes of playing games and price, and their mean value of sentimental polarity intensity is between 0.3 and 0.7. While the polarity intensity of battery and memory are relatively small, between 0 and 0.3, and their negative sentimental polarity is relatively high with 31.79% and 32.09% respectively. This shows that users believe the battery and memory configuration of mobile phone is general. The user's negative feedback on battery mainly focuses on less strong battery endurance, less use time, battery breakthrough and other aspects. The negative emotional comments on memory configuration are mainly concentrated in the system with more memory, less available memory and the inability to install memory card.

The mean value of sentimental polarity intensity of membrane attribute is negative, accounting for more than half of negative sentimental polarity of 53.17%. This indicates that most users are dissatisfied with the membrane of mobile phone. The problem of membrane feedback is mainly manifested in the lack of membrane delivery, poor appearance after film coating, and poor membrane matching of mobile phone.

5 Discussion and Conclusion

Based on users' sentiment analysis of shopping websites, the following conclusions are drawn. (1) Web users usually pay more attention to the configuration of HUAWEI mobile, such as the attributes of memory, system, screen, battery, and the after-sales service of mobile, such as the attributes of logistics, customer service, express and deliver goods. (2) Users have a positive sentimental orientation to most of attributes, among which the most satisfying ones are unlocking function, hand feeling attribute and logistics service. (3) Users show a neutral sentimental orientation towards the attributes of battery and memory, and a negative sentimental orientation towards the membrane of mobile phone. The results can provide a reference for consumers to make purchasing decisions, enterprises to improve product quality, and shopping platform to optimize service direction.

For consumers, the rapidly increasing number of online reviews provides consumers with more comprehensive information, and also brings information overload which affects consumers' judgment of online reviews.

The consumers' attitude can be automatically recognized in the huge amount of online reviews through users' sentiment analysis on online reviews. It can help consumers to understand the distribution of commentators' position, attitude or sentimental orientation towards the attributes of commodity or service, so as to make the right purchasing decisions. Those users of shopping websites are satisfied with most of attributes, such as hand feeling, appearance, express, customer service, performance, operation, delivery, reaction, quality, photograph, screen, charge and system. Most users think the phone's battery and memory configuration are general, and are dissatisfied with membrane of phone. Consumers have different expectations for quality, function and performance of products due to their different concerns and preferences. Consumers can pay attention to the attributes of commodity or service that they are interested in, so as to provide the decision reference for purchasing goods.

For commodity producers, users' sentiment analysis on online reviews can reflect consumers' concerns or preferences about products, for example, the battery endurance is not too strong and the use time is not too long; the system takes up more memory and has less available memory, the actual memory configuration does not accord with propaganda; sometimes playing games or playing video will not flow smoothly; The sound is harsh when the volume is high. The feedback information comes directly from consumers, reflecting their real idea and representing their actual needs, and should be highly valued by enterprises. Enterprises should conform to the objective reality in product promotion and solve the consumer's problem and dissatisfaction in time according to consumer feedback, in order to continuously improve product quality, service and performance. Commodity producers should formulate product marketing and development strategies in line with consumer needs to improve the image of enterprises and brands so as to win the advantages of market competitive.

For shopping platforms, users' sentiment analysis on online reviews can report the consumers' concern or feeling preference about the services provided by shopping platform. For example, the logistics speed is slow, the attitude of express service is not good, the customer service language is too stiff or impersonal, and the delivery is not timely. The feedback information represents the true feelings of consumers, and shopping websites can analyze its own problems and self-examination to continuously improve its service level. They can optimize delivery process, improve logistics speed and service quality, and provides users with better service experience through the integration of online and offline resources.

Although several valuable conclusions have been drawn from the research, there are some limitations: the research data of samples is limited and the characteristics of users' sentimental orientation of different shopping websites can also be further analyzed. In the future research, more data and samples will be used to conduct a comparative study on users' sentiment analysis of different shopping websites based on online reviews.

References

[1] Liu, Y. Word-of-Mouth for Movies: Its Dynamics and Impact on Box Office Revenue, Journal of Marketing, 2006,70(3), 74-89.

[2] Mudambi S M and Schuff D. What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com, MIS Quarterly, 2010, 34(1), 185-200.

[3] Archak N, Ghose A, Ipeirotis P G. Deriving the Pricing Power of Product Features by Mining Consumer Reviews, Management Science, 2011, 57(8), 1845-1509.

[4] Kozinets R V, de Valck K, Wojnicki A C. Networked Narratives: Understanding Word-of-Mouth Marketing in Online Communities, Journal of Marketing, 2010, 74(2), 71-89.

[5] Wang W and Wang H W. The Influence of Aspect-Based Opinions on User's Purchase Intention Using Sentiment Analysis of Online Reviews, Systems Engineering Theory and Practice, 2016, 36(1), 63-76.

[6] Zhang Z Q, Ye Q, Li Y J. Literature Review on Sentiment Analysis of Online Product Reviews, Journal of Management Sciences in China, 2010, 13(6), 84-96.

[7] Turney P D and Littman M L. Measuring Praise and Criticism: Inference of Semantic Orientation from Association, ACM Transactions on Information Systems, 2003, 21(4), 315-346.

[8] Duan G P. What Do Consumers Think Online Reviews Are More Useful? Influence Effect of Social Factors, Management World, 2012, 28(12), 115-124.

[9] Shi W H, Gao Y, Hu Y Y. Influencing Factors of Online Reviews Usefulness Based on Sentimental Orientation and Observational Learning, Journal of Beijing University of Posts and Telecommunications (Social Sciences Edition),

2015, 17(5), 32-39.

[10] Schlosser A E. Can Including Pros and Cons Increase the Helpfulness and Persuasiveness of Online Reviews? The Interactive Effects of Ratings and Arguments, Journal of Consumer Psychology, 2011, 21(3), 226-239.

[11] Chris F, Anindya G, Batia W. Examining the Relationship between Reviews and Sales: The Role of Reviewer Identity Discloser in Electronic Markets, Information Systems Research, 2008, 19 (3), 291-313.

[12] Chevalier J A and Mayzlin D. The Effect of Word of Mouth on Sales: Online Book Reviews, Journal of Marketing Research, 2006, 43(8), 345-354.

[13] Du X M, Ding J Y, Xie Z H, et al. An Empirical Study on the Impact of Online Reviews on Consumers' Purchasing Intention, Management Review, 2016, 28(3), 173-183.

[14] Zhou M H, Li P M, Mou Y P. Effects of Online Reviews on Purchase Intention of Consumers: The Mediation of Psychological Distance, Soft Science, 2015, 29(1), 101-104,109.

[15] Niu G F, Li G Q, Geng X X. The Impact of Online Reviews' Quality and Quantity on Online Purchasing Intention: The Moderating Effect of Need for Cognition, Journal of Psychological Science, 2016, 39(6), 1454-1459.

[16] Ma Y L. The Influence of the Conflict of Online Reviews on Consumer Attitudes, On Economic Problems, 2014, 36 (3), 37-40.

This page is intentionally left blank