Association for Information Systems

AIS Electronic Library (AISeL)

WHICEB 2020 Proceedings

Wuhan International Conference on e-Business

Summer 7-5-2020

Analysis of Influencing Factors of Tablet Consumer Satisfaction Based on Online Comment Mining

Zhibin Cui
School of Information,Central University of Finance and Economics,Beijing,102206,China, czb19800355302@163.com

Lu Yuan
School of Information,Central University of Finance and Economics,Beijing,102206,China

Analysis of Influencing Factors of Tablet Consumer Satisfaction Based on Online Comment Mining

Zhibin Cui[1*], *Lu Yuan*[1]

[1]School of Information,Central University of Finance and Economics,Beijing,102206,China

Abstract: How to extract effective information that affects consumer satisfaction from online comments has become a hot issue for customer behavior. This article is based on the data mining of online comments and the research object are the top-selling tablets on the JD platform from October to December 2018. We started by analyzing influential factors such as goods, after-sales service, and logistics, and crawled online review information of nearly 3,000 tablet computers from five major brands. We first use the jieba word segmentation tool to process the user comments, and use TF-IDF to calculate the frequency of different words in the comments to determine the main keywords of the comments. Secondly, we set up a user's sentiment dictionary to determine the sentiment index of the review, and combined the keywords and sentiment index to get the degree of consumer satisfaction on different influencing factors. Finally, we imported the quantified characteristic factors into Clementine 12.0, and established a Bayesian network model of customer satisfaction, thereby obtaining a ranking table of the importance of each factor to product sales. To improve the model robustness, we adopt a multivariate linear model to check the accuracy of the output results. Our research can not only formulate effective product service sales strategies for merchants, but also guarantee customers to experience better products and services.

Keywords: online comments, consumer satisfaction, sentiment dictionary, TF-IDF

1. INTRODUCTION

According to the "China E-Commerce Report 2018" issued by the Ministry of Commerce of 2019, the number of Internet customers who use the Internet for shopping and consumption in China has reached 569 million, an increase of 5.7% compared to 2017 and accounting for 71% of the total number of Internet customers. The number of mobile online shopping customers reached 557 million, an increase of 12% compared to 2017 [1]. Driven by factors such as new retail formats and capital markets, the online retail development index in the first quarter of 2018 maintained a relatively stable development trend.

The above data indicates that more consumers will choose to online shopping. However, due to the time difference between the online and offline B2C business models and the asynchronous nature of the information, customers cannot distinguish the authenticity of the product information described by the seller, and can only rely on the sellers' product descriptions, other customers' post-purchase evaluations, and their own judgments. The online post-purchase comment mechanism of the B2C online shopping platform can provide customers with strong protections and effective solutions of information asymmetry in the development of the digital economy [2]. Online comments are not only an effective feedback method for measuring product sales, but also an important method for customers satisfaction feedback of e-commerce companies and their products or services [3].

In summary, mining product-related quality information or merchant service information from online comments of consumers can effectively measure consumer satisfaction with different aspects of the product and improve product services.

In this research, we use python to crawl the online comment data of five major brands of tablets which have higher sales on JD.com in 2018. The initial data crawled 3087, and the final retained 2912 data. Then we

* Corresponding author. Email: czb19800355302@163.com (Zhibin Cui)

use jieba word segmentation package and TF-IDF method to quantify online comment data. We attribute factors that affect tablet customer satisfaction into ten aspects to build a corpus of attribute features. By regarding the ten quantified factors as input variables, and customer ratings as output variables, we establish Bayesian network models to test the correlation between online comment content and online customers satisfaction reflected in online comments during consumer purchasing decisions.

2. LITERATURE REVIEW

Many foreign scholars have done many studies on consumer comment mining from different aspects. Some scholars have researched from the perspective of comment content. Kim S et al. [4] used the SVM regression analysis method to evaluate the usefulness of comments from the aspects of comment structure, vocabulary, syntax, semantics and metadata. Chevalier et al. [5] analyzed the emotional expression of the text in the comment information, and found that online comments exerted influence on customers' perception of online comments. Moghaddam [6] proposed the problem of predicting the quality of online comments of personalized features, and established a complex probability map model based on matrix factorization and variable factorization. Other scholars have analyzed the impact of online comments on the shopping environment. Adomavicius [7] obtained the customer's true shopping experience by scoring the characteristic attributes of the product and built a customer preference model to analyze the customer's preference characteristics. Dwikesumasari et al. [8] used regression analysis to explore the impact of the functions, innovations, and consumer inertia of travel apps on customer satisfaction.

The domestic academic scholars also make many studies on the issue of online comments. Hao [9] found that more positive emotional expressions and longer comment lengths have a prominent positive impact on the usefulness of comments. Li et al. [10] analyzed the content of online comments, and believed that online comments are the most direct real feedback and needs of customers for products and services. Yan et al. [11] found that the length of comments had a significant impact on the sales of non-hot brands.

In addition, many scholars have analyzed online comments from the perspective of customer satisfaction. Cheng [12] built a TAM model to conduct an empirical analysis of the main factors affecting customer satisfaction with online shopping, and solved the two key factors of perceived usefulness and safety. Li [13] research shows that the number of customer evaluations, the level of product sales, and the evaluation time have a positive correlation with customer satisfaction, but the tendency of positive and negative emotions is difficult to affect customer satisfaction. Lu et al. [14] constructed a gray evaluation model of customer satisfaction, and established customer online comments based on product and object satisfaction evaluation indicators. Wang [15] analyzed the satisfaction degree of each attribute concerned by App usage from the perspective of customer comments through the VIKOR multi-attribute decision-making method.

Many scholars started with sentiment analysis of online comments for customer satisfaction. Shi Wei et al. [16] proposed an emotional computing method from sentence level to document level based on the established fuzzy emotional ontology to calculate the sentiment tendency of Chinese online reviews. Wang et al. [17] based on HowNet semantic similarity polarity calculation method and adverb the magnitude division method analyzes the sentiment polarity and intensity of online comments.

In general, the current research on online reviews mainly focuses on the content mining of online reviews themselves, the impact on product sales, and the impact on other users and consumers' purchasing decisions. Many studies by scholars at domestic and abroad show that online product reviews have a very significant positive impact on product sales and user satisfaction.

3. METHODOLOGY

We crawl customers' online comments through using the jieba word segmentation tool to perform Chinese word segmentation on the customer's online comments, which are text segmented. The stop words are removed according to the stop word list to avoid irrelevant vocabulary interference. Cut online customer comments into vocabularies to form bags of words. Then we establish a corpus that can describe the factors affecting consumer satisfaction in terms of different aspects of the product. The TF-IDF method is used to traverse the bag of words, record the words and their weights in the bag of words, and obtain the characteristics of the factors affecting customer satisfaction of online comments texts vector. Based on the word frequency and weight in the customer's online comments, determine the comment keywords involved in the comment. The keywords are compared with the corpus of consumer satisfaction influencing factors to determine the product influencing factors in this comment. Moreover, we create an emotional vocabulary that can describe the factors affecting consumer satisfaction of product characteristics. Then we form an emotional dictionary that can describe consumer satisfaction by using machine learning method to obtain new customer satisfaction emotional words. The sentiment dictionary is used to perform sentiment calculation on the text content of online comments to judge its sentiment tendency and degree, and obtain the customer satisfaction sentiment index of the influencing factors. According to the customer's sentiment index, determine the customer's satisfaction with this aspect of the product.

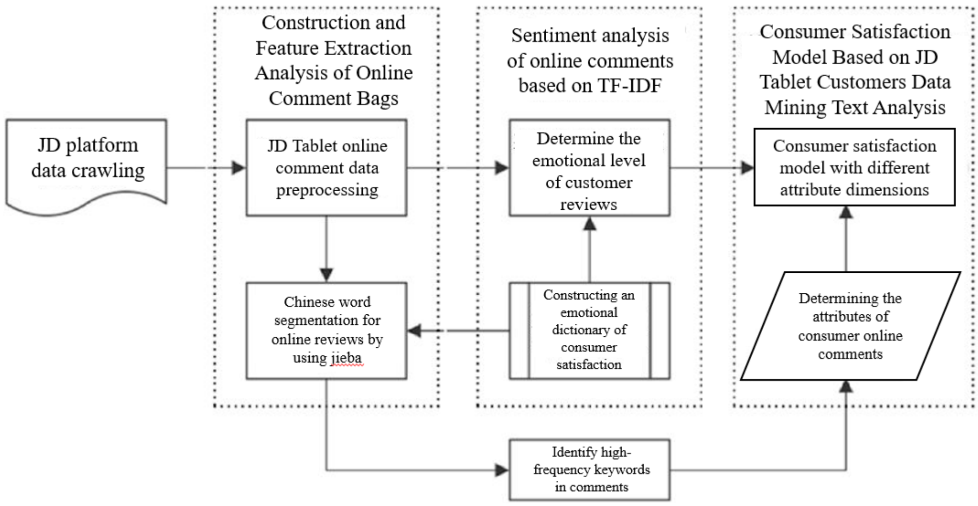

Figure 1. Research framework

We regard the total star rating of the consumer as the dependent variable and ten aspects of the influencing factors as independent variables. We establish the Bayesian network model to analyze the factors affecting the consumer satisfaction of this type of product.

4. RESEARCH CONTEXT

4.1 Data collection

We use python to crawl the customer's online comment data about major brands' tablets on the JD platform and finally chose Online comment data of 5 high-selling tablets including Lenovo Miix 520, ipad 2018, Xiaomi 4, Honor Water play and Microsoft Surface Pro6 from October 2018 to January 2019.

Online comments also contain a lot of redundant information, which does not rule out comments that merchants scramble for sales or comments that are maliciously discredited by competitors. This information will affect the customer's emotions and cause deviations in shopping decisions. For invalid or malicious information, it is eliminated in the comment data acquisition stage to avoid affecting the final result. Finally, we crawl 3087 pieces of initial data and retain 2912 pieces of filtered data. On this basis, we carry out a text analysis and summarize the high-frequency words for each tablet. We extract and retrieve the high-frequency words of each product in the database and count the occurrences.

Table 1. Overall data description

product name	Praise degree	Praise	Average	Bad comment	Crawls number
Lenovo Miix 520	98%	15,000+	100+	100+	492
ipad 2018	99%	780,000+	2700+	4200+	862
Xiaomi 4	99%	91,000+	300+	300+	969
Honor Water play	99%	43,000+	100+	100+	490
surface pro5	99%	42,000+	100+	200+	274

4.2 Data quantization

Although some feature information of the tablet has been summarized in online comments by customers, some of the features of these frequently-occurring data information are redundant and repeated, and the product features expressed by many keywords are semantically close. And there may be some keywords with the emotional characteristics of the commenter. Simply extracting the keywords will weaken their emotional characteristics to some extent. Therefore, we need to further summarize the synonyms, based on the analysis and summary of the factors affecting customer satisfaction, the above high-frequency words are grouped into 10 categories, respectively: logistics, price, appearance, physical attributes, models, Operation, battery, picture quality, sound quality, and merchants. These factors are important feedback that customers will refer to when making a tablet purchase. This information may affect customer satisfaction more or less. we use TF-IDF method to count the word frequency of keywords in each category

Table 2. Variable setting

variable	Meaning	Key words
$Logistics_i$	Customer i 's satisfaction with the product logistics	Logistics, courier, packaging, delivery
$Price_i$	Customer i 's satisfaction with the price of the product	Cost-effective, price, cheap, expensive
$Exterior_i$	Customer i’s satisfaction with the appearance of the product	Appearance, beauty, workmanship, grade, compact
$property_i$	Customer i’s satisfaction with the physical property	Weight, thin, light, size
$Model_i$	Customer i 's satisfaction with the product appearance	Screen, feel, portable, buttons, comfortable
$Operation_i$	Customer i 's satisfaction with the operation of the product	fever, stable, smooth, fast, sensitive, system, simple,
Battery $quality_i$	Customer i 's satisfaction with the product's battery life	Endurance, durability, charging
Picture $quality_i$	Customer i 's satisfaction with the quality of the product	Clear, pixel, picture quality, resolution
Sound $quality_i$	Customer i 's satisfaction with the sound quality	Sound quality, noise
$Services_i$	Customer i 's satisfaction with this product merchant service	After sales, service, customer service

Based on online comment data, customers express their emotions with adjectives when expressing their satisfaction. Words such as "very", "special", "very" all express the customer's inner satisfaction to a certain extent. We establish a sentiment dictionary for customer comments and use TF-IDF method to calculate the word frequency and weight. Different thresholds are set for different scores. This scores ten factors for each comment.

$$\begin{aligned} Y_{rank_i} &= X_{Logistics_i} + X_{Price_i} + X_{Exterior_i} + X_{Physical_i} + X_{Model_i} + X_{Operation_i} \\ &+ X_{Battery_i} + X_{Picture_i} + X_{Sound_i} + X_{Services} \end{aligned} \tag{1}$$

According to the scoring table established above, each variable can be quantified and scored by comparing each comment information. In the scoring process, the scoring is based on the emotional tendency of the modifiers. It should be noted that the Chinese words have corresponding synonyms, such as very beautiful and beautiful, so the corresponding synonyms are converted during the quantification process. For the scores of variable factors that do not appear in the comment information, this article is uniformly identified as 3 points, that is, the emotional color of this variable is average. The figures obtained after the above rules serve as the data for this characteristic factor. At the same time, each comment of JD has an overall score, which has been collected at the time of data collection. The overall score represents the overall customer satisfaction with the purchase, so this score is used as the overall customer satisfaction score for this purchase. Quantify all the data and strictly follow the scoring rules in the processing process to obtain online comment quantification data for analyzing product sales.

5. MODEL RESULTS ANALYSIS

5.1 Results

The processed data was imported into Clementine 12.0 software. After removing invalid content such as member name and comment content, the evaluation star was used as the output variable, and ten factors were used as the input variable. By calculating the importance ranking of each characteristic factor and the conditional probability of each node output by the model.

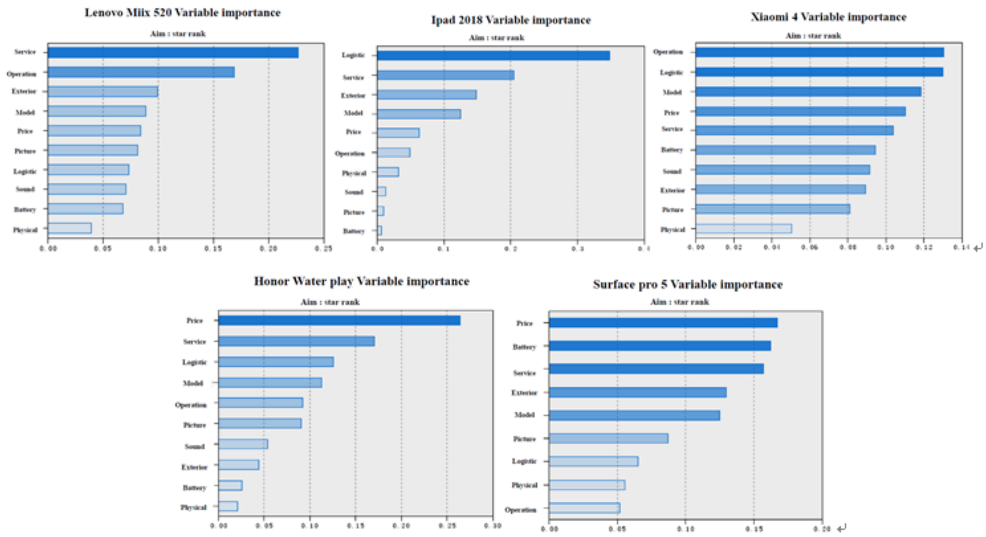

Figure 2. Research output

According to the results of the Bayesian network model, we can see the importance and ranking of ten characteristic factors. The more important the ranking of the characteristic factors, the more it can affect the sales of the product. The importance of the ten characteristics of the five major brands is shown in the following table.

Table 3. Variable importance coefficient output

Independent variable	Lenovo Miix 520	ipad 2018	Xiaomi 4	Honor Water play	surface pro5
Logistics	0.0732	0.348	0.1301	0.08	0.0654
Price	0.0839	0.063	0.1101	0	0.1437
Exterior	0.0992	0.148	0.0894	0.095	0.1204
Physical property	0.0394	0.032	0.0504	0.114	0.0514
Model	0.0885	0.125	0.1184	0.281	0.1203
operation	0.169	0.049	0.1304	0.325	0.0495
Battery quality	0.0681	0.006	0.0946	0.025	0.1565
Picture quality	0.0817	0.01	0.0811	0	0.1184
Sound quality	0.0704	0.013	0.0916	0.043	0.0197
Services	0.2267	0.205	0.1039	0.036	0.1548

5.2 Test of accuracy evaluation

According to the results of the Bayesian network model, we can see the importance and ranking of ten characteristic factors.

Table 4. Multiple regression coefficient output

Independent variable	Lenovo Miix 520	ipad 2018	Xiaomi 4	Honor Water play	surface pro5
Logistics	0.071	0.017	0.044	0.117	0.024
Price	0.064	0.033	0.038	0.170	0.322
Exterior	0.069	0.048	0.025	0.038	0.132
Physical property	-0.013	0.009	-0.056	-0.042	0.029
Model	0.089	0.257	0.130	0.077	0.206
Operation	0.024	0.039	0.064	0.139	0.099
Battery quality	0.134	0.022	0.005	0.031	0.066
Picture quality	-0.028	-0.034	0.009	0.143	0.007
Sound quality	-0.047	-0.008	0.007	0.098	-0.027
Services	0.386	0.362	0.232	0.294	0.398
R^2	0.968	0.920	0.992	0.988	0.957
Adj-R^2	0.964	0.914	0.988	0.985	0.955
F-value	12.903	22.552	8.632	12.951	16.078
p-value	0.000	0.000	0.000	0.000	0.000

According to the results of multiple linear regression, R^2 and adjusted R^2 are an explanation of the model fitting effect. The corrected goodness of the model is above 0.9, indicating that the model has a strong ability to interpret; the probability P values corresponding to F are 0. 000 <0.01, indicating that the 10 independent variables of the five major tablet computers introduced had a significant effect on the dependent variable at a significant level of 0.01.

Secondly, in the coefficients presented, we can see that for Lenovo miix520, the more important factors are merchants, batteries, and models; for iPad, the more important factors are merchants, models, and appearance. For Xiaomi 4, the more important factors are merchants, models, and logistics; for Glory, the more important factors are merchants, prices, and operations; for the surface, the more important factors are merchants, prices, model. In general, although the importance ranking of some factors may be slightly different, the top five factors output by the major models are basically similar, indicating that the results of the two models are basically the same. On this basis, we compare the accuracy of the models. The relevant results are shown in the following table:

Table 5. Model accuracy comparison

	Lenovo Miix 520	ipad 2018	Xiaomi 4	Honor Water play	surface pro5
Accuracy	97.35%	98.21%	95.63%	97.94%	96.52%
Average accuracy	0.984	0.986	0.973	0.975	0.954
	0.881	0.967	0.882	0.913	0.946
Accuracy above 2.0 folds	98.91% of observation	91.89% of observation	94.47% of observation	96.23% of observation	93.42% of observation
Multiple linear regression model accuracy	97.41%	97.63%	96.98%	974.29%	97.65%

As can be seen from the above table, the accuracy rate of Bayesian network models of major brands is above 95%. Among them, the accuracy of the iPad is the highest, which is 98.21%. It will improve the accuracy by more than 2.0. Many changes indicate that it is more feasible to use Bayesian network models to explore user satisfaction factors.

6. DISCUSSION

In this research, we use the text analysis method to mine the online comments of tablet products on the JD platform, and construct a Bayesian network model to analyze the main consumer influence factors affecting tablet of different brands, to provide manufacturers with a reasonable response to increase consumer satisfaction.

In terms of Lenovo miix520, improve the service quality of merchants. When a new product is listed, it is important to focus on the early quality evaluation and improve the quality before promoting it to the shelves. Manufacturers should increase investment in technology research and product maintenance, and further improve the smoothness of CPU processing on the operating system.

In terms of iPad 2018, speed up logistics and distribution, strengthen the efficiency of order and inventory management. They should implement real-time logistics solutions simultaneously, reduce the picking area, speed up transportation, and enhance customer service capabilities.

In terms of Xiaomi 4, enhance customer service capabilities. Combine the types of refunds to determine the core of the problem and improve yourself. For the design of the appearance, it is necessary to consider the experience of most customers more closely and be practical.

In terms of Honor water play, optimize the appearance of the tablet, and further adjust the physical properties of the tablet to make it more convenient and compact, while also maintaining high running smoothness and cost performance.

In terms of surface pro6, improve the battery life of the tablet and choose a battery with a higher storage capacity. Meanwhile, it is also necessary to optimize the power consumption of the tablet application to improve the durability and use time of the tablet.

To achieve differentiated marketing of tablets products, companies should learn from the marketing experience of smart phones. It’s conductive to accelerate the establishment and improvement of unique application software stores, increase the quality management of the plant, and implement detailed project management in the manufacturing process. Companies should improve their after-sales service capabilities and focus on efficient and timely logistics distribution. For some tablet PC series mainly promoted, companies should formulate a clear pricing strategy. Some complementary products can be bundled and presented at the same time. The outstanding advantages of their own products should be emphasized for publicity and word of mouth marketing.

7. FUTURE WORK

This research provides a reference for the future development for tablet manufacturers. At the same time, we point out the weaknesses in this research and the feasible development directions in the future.

The amount of sample crawl data is not representative, and the accuracy of the model needs to be improved. We will crawl larger and more comprehensive online review data to further improve the accuracy and comprehensiveness of the customer satisfaction model.

Characteristic factors are not comprehensive enough. Due to the large amount of data, it is easy to ignore certain feature factors or directly combine two different feature factors into one when summarizing and extracting feature factors. Extracting slightly subjective features may lead to the omission of features that affect customer satisfaction, which will leave the model itself inaccurate.

REFERENCES

[1] Department of Electronic Commerce and Informatization, Ministry of Commerce. China E-Commerce Report 2018, 2019

[2] Xiao Jie. Emotion intensity analysis of online reviews based on text mining [D]. Jinan University,2018.

[3] Zhang Yuanyuan. Empirical research on the influence of online reviews on consumer purchasing decisions [D]. Zhejiang University of Finance and Economics,2015.

[4] Kim S, Pantel P, Chklovski T, et al. Automatically assessing comment helpfulness[J]. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2006, 1:423-430.

[5] Chevalier J,Mayzlin D. The effect of word of mouth online:online book comments[J].Journal of Marketing Research,2006,43(3):348-354

[6] Moghaddam S, Jamali M, Ester M. ETF: extended tensor factorization model for personalizing prediction of comment helpfulness[C]. //Proceedings of the fifth ACM international conference on Web search and data mining. ACM, 2012, 1:163-172.

[7] Adomavicius G,Kwon Y,New recommendation techniques for multi-criteria rating systems[J].IEEE Intelligent Systems,2007,22(3):48-55.

[8] Dwikesumasari P R, Ervianty R M. Customer Loyalty Analysis of Online Travel Agency APP with Customer Satisfaction as a Mediation Variable[C]//Proceedings of the 2017International Conference on Organization Innovation. 2017.

[9] Hao Yuanyuan. Empirical Research on the Impact of Online Reviews on Consumer Perceptions and Purchase Behavior [D]. Harbin Institute of Technology,2010.

[10] Li Jinhai, He Youshi, Ma Yunlei, Li Zhiw en. Building Dynamic User Preference Model Based on Information Mining of Online Reviews [J]. Journal of Intelligence,2016,35(09):192-198.

[11] WANG Jun-jun, YAN Qiang. An Empirical Study on the Impact of Online Reviews of Different Product Popularity on Product Sales [J]. Chinese Journal of Management Science, 2013, S2:406-411

[12] Cheng Hua, Bao Gongmin. Empirical Research on Determinants of Online Shopping Intent [J]. Quantitative and Technical Economic Studies,2003(11):150-153.

[13] Li Jian. Research on the Impact of Online Product Reviews on Product Sales [J]. Journal of Modern lnformation,2012,32(01):164-167.

[14] Lu Pin, Zhong Yan, Tang Hao. Research on Comprehensive Evaluation of User Satisfaction of Online Product Reviews [J]. Acta Electronica Sinica,2014,42(04):740-746.

[15] Wang Zuzhu, Xie Yueyun. User Satisfaction Analysis of App Services from the Perspective of Online Reviews [J]. Journal of Chengdu Technological University,2017,20(04):21-26.

[16] Shi Wei, Wang Hongwei, He Shaoyi. Sentiment Analysis of Chinese Online Comments Based on Semantics[J]. Journal of the China Society for Scientific and Technical Information, 2013, 32(8):860-867.

[17] Wang Hongwei, Song Yuan, Du Zhanqi, et al.Evaluation of Service Quality for Express Industry Through Sentiment Analysis of Online Comments [J]. Journal of Beijing University of Technology, 2017, 43(3): 402-412.