ISSN: 2688-9323

Analysis on Negative Impact of Logistics Service Quality based on LDA Model

Xue Zhou*, Qing Kai

School of Economics and Management, Xidian University, Xian 710126, China

*168192985@qq.com

Abstract

[Objective]By analyzing the negative review data concentrated in the logistics industry, the negative index which has a great influence on the logistics service can be found.[Methods]We take the complaint cases of the website of the State Post Office as a source of data, and the data was cleaned and processed to make an available feature set.Then by using a weakly supervised LDA model, we add some priori knowledge of logistics to generate seed words manually to extract data information.[Results]It turns out that reliability and interactivity have the greatest impact on customer satisfaction, we'll give some targeted suggestions based on this.[Limitations]The weakly supervised LDA model used in this paper is less researched at present, and its accuracy and accuracy still need to be improved.[Conclusions]The logistics enterprises can improve their own service quality for specific aspects, so as to improve consumer satisfaction and maintain their own long-term development.

Keywords

customer satisfaction; complaint cases; text mining; topic models.

1. Introduction

With the rapid development of the Internet and the continuous expansion of the scale of online shopping, China's express delivery industry has also expanded rapidly. The China Express Development Index shows that the index of China's express delivery industry in 2018 was 1765.3, an increase of 25.5 points over 2017. Service Quality Index It is 133.4, an increase of 30.2% year-on-year [1]. This is the first time that the growth rate of the logistics service index exceeds the development scale index, which means that after the large-scale and high-speed growth of the logistics industry in China, it is gradually transforming from a large logistics country to a strong logistics country and gradually improving the service quality. Become a key part of service management in the logistics industry.

How to rationally optimize the quality of logistics services has also attracted widespread attention from scholars at home and abroad. Part of the existing related research is based on macro theory and countermeasures [2], and it has also derived the impact of individual logistics factors on consumer satisfaction. Research [3] and research on the impact of consumer purchasing behavior in e-commerce platforms [4]. Among them, most of the literature is on the impact of logistics service evaluations on consumer purchasing behavior in e-commerce platforms [5-7]. The two major e-commerce platforms represented by Ali and Jingdong have adopted completely different logistics models, so some scholars have also conducted comparative studies on the differences in logistics models [8]. In addition, some scholars have an influence on online reviews and their effectiveness. Related research on sex [9]. Existing research divides online reviews into positive reviews, neutral reviews, and negative reviews by content. Negative reviews often mean loss, and people hate loss, so the study considers negative reviews to be more effective And has a greater influence [10]. If you can make full use of the negative evaluation of the quality of logistics services, in-depth analysis of these data information will help logisticsIndustry to make decisions and gradually improve the quality of their own logistics services. However, the current research on logistics big data mainly focuses on online reviews of shopping websites, among which there are a small number of reviews related to the quality of logistics services and insufficient negative evaluation information. Literature appears to focus on the negative evaluation of logistics services.

Based on the above background, the text selects an important information platform for consumers to focus on publishing negative evaluations of logistics services. The logistics complaint website of the State Post Office of China uses the consumer complaint cases on this website to analyze and study the service quality of the logistics industry with a view to Help improve the overall service quality of the logistics industry. By collecting appeal cases on the complaint website of the State Post Office as data support, the main research contents of this article are as follows: First, use a more authoritative and widely used industry-based logistics service quality based on SERVQUAL and LSQ The evaluation model determines the evaluation dimension of the service quality of the logistics industry, and divides the service quality evaluation of the logistics industry into six dimensions and indicators of each dimension (that is, the six topics of the theme model). Second, through the case of the complaint website of the State Post Office Perform data collection, preprocessing, word segmentation and other processing processes to form a usable data set. Features are extracted based on syntactic analysis, and Word2vec clustering performs feature extraction and cluster analysis on data. Based on prior knowledge in the field of logistics, features generated based on clustering Set artificially generated seed words. Based on the generated seed words, makeUse the weakly supervised topic model to extract the potential topic structure of logistics service quality from the data set, that is, the six dimensions of the service quality evaluation model. Finally, use statistical methods to classify and analyze the topics to obtain each service quality dimension for logistics. The impact of the negative evaluation of services. Finally, based on the analysis results, a path to improve the quality of express delivery services of logistics companies is proposed, in order to improve the level of service quality in the logistics industry.

2. Establishment of Logistics Service Quality Evaluation Scale

2.1. Theoretical Basis of Logistics Service Quality Evaluation

Research on service quality began in the late 1970s, and since then, service quality issues have aroused research interest among many scholars. Representative Gronross [11] proposed customer perceived services based on the basic theory of cognitive psychology The concept of quality believes that service quality is a subjective category, and its strength depends on the comparison between customer expectations and actual perceptions. This theory has a groundbreaking effect on the study of service quality. In the decades that followed In China, scholars have conducted a lot of valuable research on service quality and related issues, and gradually subdivided into various fields.Among them, the service quality evaluation models such as SERVQUAL and LSQ have a greater impact on the logistics industry.

In 1988, the PZB Group [12] proposed a "difference model" for service quality, began to study the evaluation of perceived service quality, and proposed a SERVQUAL evaluation system. The evaluation system includes 5 dimensions, namely: tangible, responsive, Reliability, empathy, and assurance.There are 22 secondary indicators in five dimensions. In 2001, MENTZER et al. [13] proposed the LSQ model for measuring the quality of logistics services. Research on service quality in the field.Gross, Suther and others believe that service quality includes both the result of the service and the way to provide the service, which is commonly referred to as the quality of technology (quality of results) and function (quality of processes, quality of

ISSN: 2688-9323

interactions) [14]. Therefore, to evaluate the service quality of a logistics enterprise, we must not only evaluate its technical quality, but also its functional quality. Technical quality and functional quality together build the evaluation dimension of the logistics enterprise service quality. The dimension of the SERVQUAL model focuses on the " Evaluation of "functional quality", Brady Cronon research shows that SERVQUAL model indicators describe the attributes of one or more service quality processes and also support this view [15];Technical and functional quality are the two most important factors in the evaluation of logistics service quality, which have also been recognized by scholars such as (Mentzer, Flint, and Hult; Mentzer, Flint, and Kent).

To sum up, the evaluation of the service quality of logistics companies must start from two main aspects: "technical quality and functional quality", and combine with the actual situation of China's logistics enterprises, drawing on the "functional quality" evaluation connotation in the servqual model and the establishment of the The "technical quality" evaluation content of the lsq model of logistics service quality evaluation to determine the evaluation dimension of service quality.

2.2. Establishment of Logistics Service Quality Evaluation Scale

The logistics service quality evaluation scale used in this article is an evaluation system based on the servqual scale and the lsq model, which are representative of service quality evaluation.The evaluation system is a combination of the two. The recognition of experts in the industry.

Existing research shows that the evaluation of the service quality of logistics enterprises needs to start from two aspects: "technical quality" and "functional quality", while the servqual model focuses more on "functional quality", and the evaluation connotation of the lsq model is more biased towards " "Technical quality". Therefore, based on the combination of the two and the actual situation of China's logistics enterprises, the specific indicators of this article's logistics service quality evaluation scale are as follows [16]:

3. Data Preprocessing and Feature Extraction

The overall research process of this article is shown in Figure 1:

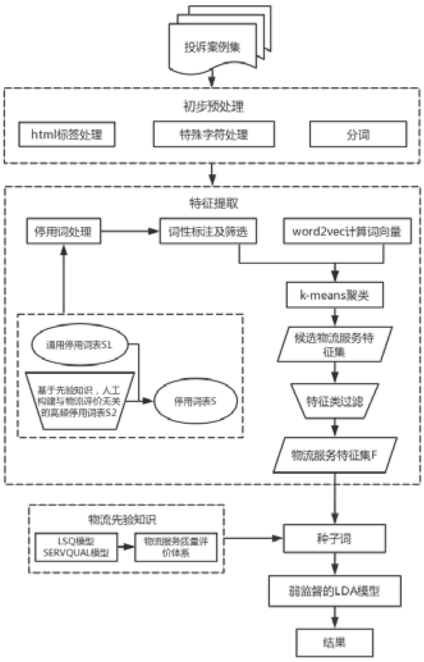

Figure 1. Research process

Table 1. Evaluation scale of logistics service quality

Dimension	Secondary indicators
Convenience	c1 Wide coverage and reasonable location
	c2 Multiple delivery methods
	c3 Mailing process
reliability	r1 The outer packaging and goods are intact
	r2 Express companies have high credibility
	r3 Parcel delivery is accurate
	r4 accurate logistics information displayed on the Internet
Responsiveness	t1 order response time
	t2 Update speed of logistics information
	t3 parcel waiting time
Interactivity	e1 Staff attitude
	e2 can provide personalized services
	e3 Is the courier company proactive and swift in responding to errors?
	e4 Whether the express company handles customer complaints and makes customers satisfied?
Economical	p1 Provision of services in line with fees paid
	p2 elasticity of shipping and pricing
	p3 reasonable price
safety	s1 Express shipments that will not pose a safety hazard to the country, organization, or citizen
	s2 The security of service personnel and shipments is protected by the security measures of express delivery companies
	s3 Express companies and related personnel will not disclose all information about express

First, use the Python crawler to collect the appeal cases on the National Post Office's logistics appeal website, and perform data cleaning and preliminary preprocessing. Second, based on the processed data, add prior knowledge in the logistics field, filter it, and manually construct stop words. Based on this, the part-of-speech tagging and screening work is performed to generate a set of characteristic words of logistics service quality. Under the guidance of the logistics service quality evaluation system established earlier, the final seed words are artificially generated and a weakly supervised LDA model , Enter the seed word result, and get the final topic extraction result. The specific implementation process is as follows:

3.1. Data Acquisition

This article collects data from more than 10,000 cases disclosed on the complaint website of the State Post Office. The sample includes all case data accumulated from the initial selection of cases to the beginning of the research. All analyses in this article are performed using the Python programming language. The specific implementation method is as follows:

First, use the web crawler [17] method to collect the required case data. The web crawler can automatically obtain the required web page content through a program that sets rules, and its purpose is to crawl the URL of a website or application Get the content of the webpage, and then parse out the valuable information we need [18]. This article uses Python to crawl all 9377 complaint cases from the complaint site of the logistics website from 2009 to 2018, and the data is stored in json format by default.

Since the data source of this article is unstructured text data, each case is in the form of a long text, so to obtain effective information in it, you need to use parsing and feature extraction.

3.2. Preliminary Preprocessing

Preprocessing is the preparatory work before subsequent feature and topic extraction.It mainly includes segmentation and part-of-speech tagging of the case text, and segmentation of the sentence so that each clause reflects only one corresponding topic to improve the subsequent analysis results. accuracy.

(1) Remove special characters

Due to the lack of a unified format and the randomness of comment expression, there is a serious problem of colloquialization in consumer complaint cases on the complaint website. Consumers will not care whether there are typos in the comments, or whether the language description is appropriate, and due to the collected The data contains many meaningless special characters, such as spaces, incorrect punctuation, and traditional characters.Therefore, you need to remove the special characters from the collected data first to make the specifications unified and convenient for subsequent research.

(2) Word segmentation

Word segmentation [19] is to cut a piece of text into a sequence of word units according to certain specifications. Before data mining, you need to perform word segmentation on the Chinese character sequences. There are currently many open source word segmentation tools available. Use jieba word segmentation to segment the pre-processed dataset. Jieba segmentation is a Python Chinese word segmentation component that is mainly used to implement Chinese word segmentation, part-of-speech tagging, and keyword extraction, and has several different modes for users to choose, so in It is widely used in Chinese word segmentation processing.

(3) Stop word processing

After word segmentation, there are still many meaningless mood words like "ba", "?", "?", "", Or turning words like "even", "but", and Or some symbols and words that are meaningless to the analysis result, such words are called stop words. The next step is to remove these stop words, that is, to stop the word operation.

In this paper, due to subsequent actual analysis, different stopword strategies will be used: the first strategy is based on the topic model, and the Chinese stopword list containing commonly used stopwords is obtained through the network.The jieba word segmentation algorithm also has its own stopword lexicon. The stop word after the integration of the two is used as a new stop word lexicon. The second strategy is based on the analysis of sentiment tendency, and the affective words, degree adverbs, and negative words in the stop word lexicon are eliminated on the basis of the first. Reduce the impact of information loss in the word segmentation process on the results of sentiment analysis. The third strategy manually analyzes the high-frequency words appearing in the result of the previous step to manually select words that are meaningless to subsequent research. Add them to the stop list, In this way, the segmentation effect is maximized, and a new stopword list S is finally obtained.

The segmentation result after removing stop words is shown in Figure 2:

ISSN: 2688-9323

"n", "傲慢": "a", "办事": "n", "没有": "v", "赔偿": "v", "损失": "n", "消协": "v", "投诉": "v", "承担": "v", "责任": "n", "城意": "n", "来": "v", "恳请": "v", "主持公道": "n", "维护": "v", "合法权益": "n"}, {"到": "v", "官方网": "n", "投诉": "v", "回复": "v", "说": "v", "陆运": "n", "时限": "n", "慢": "a", "是": "v", "原因": "n", "到户": "v", "有": "v", "解决": "v", "需求": "v", "帮助": "v", "希望": "v", "等到": "v", "工作人员": "n"}, {"寄": "v", "到": "v", "安": "v", "市区": "n", "对方": "n", "收到": "v", "是": "v", "讲": "v", "说法": "v", "是否": "v"}, {"买": "v", "明信片": "n", "纪念奖": "n", "知道": "v", "去": "v", "兑奖": "v", "时": "n", "邮局": "n", "没有": "v", "奖品": "n", "跑遍": "v", "说": "v", "没了": "v", "到": "v", "总局": "n", "人员": "n", "留下": "v", "手机号": "n", "开始": "v", "电": "n", "询问": "v", "问": "v", "得到": "v", "回答": "v", "奇怪": "v", "中奖率": "n", "高": "a", "没": "v", "领奖": "n", "期限": "n", "可": "v", "应该": "v", "没错": "v", "省局": "n", "发": "v", "可能": "v", "要": "v", "拖": "v", "没底": "v"}, {"韵": "n", "

Figure 2. Part-of-speech tagging results

3.3. Feature Extraction

After the text preprocessing is completed, we begin to extract the logistics service features. The feature extraction process in this article is as follows:

(1) Extraction of display features based on part of speech

Usually, a clause is a sentence after the text is divided by punctuation such as a period, semicolon, and question mark. For example, the text "邮件到达没有投送, 6月28日到达无锡." It consists of two clauses. For this article to be resolved Questions, clauses should contain at least one of the six themes. According to observations, the characteristic feature words in complaint cases are usually nouns, noun phrases, or verbs, and opinion emotions are often adjectives, verbs, or nouns. So based on the labeling results, this article filters out some irrelevant clauses by only retaining clauses containing nouns, verbs, or adjectives.

The significance of clauses in this article is to improve the stability of clustering in the subsequent feature extraction process. In principle, a clause only reflects one topic type. By clustering the clauses, the clustering results of other features are reduced. Impact. Of course, a single clause belongs to only one topic type is ideal. Many times, a clause contains multiple topic types. This article only focuses on multi-feature clauses that express a parallel relationship and only include nouns, noun phrases, or verbs. Sentence, other forms of multi-feature clauses are not further processed. At the same time, the tag library is normalized, and the tags are clustered using the word2vec method.Some semantically similar tags are clustered together through clustering to achieve The role of semantic weight loss [20].

(2) Feature word clustering

Through the previous step, high-frequency feature words on different topics can be extracted. Feature word clustering is a method of clustering feature words that express the same or similar attributes into one category to eliminate semantic repetition, such as "价格", Both "价钱" and "价" are classified as "价格" [21]. Similarly, feature word clustering will be very sparse if expressed using a traditional vector space model. To avoid this sparseness, this article uses custom similarity The degree measurement method measures the distance between feature words by combining semantic similarity and viewpoint similarity based on feature and viewpoint co-occurrence information.The calculation of viewpoint similarity refers to the method in the literature [22]. Feature words are clustered. The method of extracting feature words from part-of-speech path templates in the literature [23]. The feature words used by the evaluation object generally appear in the form of nouns, noun phrases, or verbs, and the emotional words used to express opinions are often adjectives and nouns. Or verbs are the main, and the collocation between feature words and emotional words usually has a pattern similar to "noun + adjective", such as "物流速度慢" and "配送服务差". This article is based on a similar method.The threshold is set to a length of 4, frequency threshold value of 5%, statistical template was filtered and extracted high-frequency part of speech paths 8 shown in Table 2 as templates, the final extract all candidate feature words.

ISSN: 2688-9323

Table 2. High-frequency feature words (top 50)

Feature words	Word frequency	Feature words	Word frequency
公司	3254	态度	928
电话	3187	**派件**	899
投诉	3064	**希望**	899
说	2912	客户	891
快件	2642	**送到**	886
收到	2590	查询	834
联系	2587	**信息**	790
收件人	2230	**无人**	788
派送	1977	时间	765
处理	1616	**表示**	756
客服	1521	**寄**	747
申诉	1467	回复	740
到达	1428	**收件**	720
显示	1327	**解决**	669
签收	1314	**地址**	642
送	1292	货物	642
消费者	1176	货	631
赔偿	1141	答复	608
人	1042	发出	603
问题	941	丢失	573

4. Topic Model Construction based on Prior Knowledge of Logistics

By extracting high-frequency feature words, we understand the words that appear more frequently in user complaint cases, and have a preliminary understanding of the main concerns of users when conducting negative evaluations of logistics services. In order to further explore different topics in user complaint cases This paper will use the method of weakly supervised lda topic model for analysis.

4.1. Weakly Supervised Lda Model Construction

(1) Introduction to the lda model

The weakly supervised lda topic model used in this paper is different from the previous lda model.It analyzes the text under the given topic and automatically judges the probability that the text belongs to a certain topic based on the artificially generated seed words. Compared with traditional ones, The word frequency greatly improves the accuracy and recall of the subject identification [24].This article uses a latent semantic analysis-based text mining method for topic mining, mainly using a latent Dirichlet distribution-based topic generation model [25]. In this topic model, a series of topics are generated in the form of a polynomial distribution For each text, each word is also sampled from these topics in a manner that obeys polynomial distribution, thus constituting the process of the model generating text around the topic.

(2) Principle of weakly supervised lda model

The main idea of the weakly supervised LDA model used in this paper is to treat each text as a mixed probability distribution of all topics, and treat each topic as a probability distribution on words. Therefore, when there are D documents , T topics, and W words, the probability of the ith word in a document can be expressed as:

$$P(w_i) = \sum_{j=1}^{T} P(w_i \mid z_i = j) P(z_i = j) \tag{1}$$

In the LDA model, the parameter z represents the topic, and the parameter w represents the word.The expression P (zi = j) in Equation (1) represents the probability that a word from the document belongs to the topic j, and P (wizi = j) Represents the probability that the word is i when the word belongs to topic j. P (zi = j) can be expressed as a polynomial distribution of the document on the topic, recorded as $\theta_j^d = P(z = j)$ Let P (wizi = j) be a polynomial distribution of the topic on the word, recorded as $\varphi_w^j = P(w \mid z = j)$.

After adding the Dirichlet prior to the above generative document ideas, we get the well-known LDA model, where θ Indicates the distribution of documents on topics, φ Represent the distribution of topics on words, then add θ with φ The prior distribution (dirichlet distribution of parameters α and β respectively), so that we can get the mathematical expression of the dependency relationship between the parameters of each layer of the LDA model:

$$\begin{aligned}
&w_i \mid z_i, \varphi^{(z_i)} \sim Discrete(\varphi^{(z_i)}) \\
&\varphi \sim Dirichlet(\beta) \\
&z_i \mid \theta^{(d_i)} \sim Discrete(\theta^{(d_i)}) \\
&\theta \sim Dirichlet(\alpha)
\end{aligned}$$

The LDA model is a probabilistic generative model.The process of generating a text is as follows: ① For document d, from $Dirichlet(\alpha)$ Sampled θ (d); ② For topic z, from $Dirichlet(\beta)$ Sampled φ (z); ③ for each word wi and the topic zi, from the polynomial distribution θ Zi = P (zi | θ), From polynomial distribution φ In the sampling wi = P (wi | zi, φ).

4.2. Seed Word Generation

(1) Feature filtering

Feature class filtering is mainly to filter the feature word clustering results to obtain more accurate results. In this paper, all words in the same feature class are treated as the same feature. In the same way, all words in the same opinion class are treated as the same. Viewpoints, and then build a feature-viewpoint co-occurrence matrix to filter features through mutual information pmi [26].First, stop word elimination is performed, and by counting the frequency of words, the words that appear most frequently in each feature class are used as the topic words of this class, so that the appearance of other words in each class contributes to the frequency of the topic words, and the opinion words are similar. Among them, the correlation matrix of features and viewpoints.

(2) Corpus construction

The corpus construction is based on each type of feature and is represented by a set of words related to the topic of the category. For example, the feature "寄件" can be represented by a vector composed of keywords such as "寄", "发出", and "寄出" Each word has a different degree of contribution to the feature, such as "贵" can only describe "价格", which can be directly mapped to the feature "价格"; while words like "快" can describe multiple features For example, "物流速度", "客服响应速度", etc., need to be determined by combining multiple words in the context; words with broad meanings such as "不错" and "不行" can describe almost all features without distinguishing features. Consider filtering it out.

Based on the results of syntactic analysis, this article only retains nouns, adjectives, and verbs. For features that appear explicitly in the comment, the words that appear in the clause in which the feature is included are added to this type of feature prediction. The corpus format is expressed as {feature, {feature Words}, {<terms, word frequency>}}, where {feature words} represents all the expressions of feature words appearing in the feature category, {<terms, word frequency>} represents the terms and appearances that appear under the feature category The number of times used to describe feature related words.

(3) Seed word generation

Seed words are a key part of mapping text data to different topics.The comprehensiveness and accuracy of seed words affect the final topic extraction results.Therefore, it is not enough to automatically generate seed words based on the high-frequency feature words we filter. It also needs manual screening and elimination [27].

In the process of artificially generating seed words, we fuse semantically similar and similar feature words so that multiple synonyms can be expressed in the same vocabulary. Manually remove unwanted high-frequency feature words such as 快递; **快件**; **包裹**; 消费者. For example, these feature words cannot represent any topic, so they are meaningless. On the basis of the above results, the feature corpus is filtered by adding prior knowledge in the field of logistics, and the partial seed word results are generated as follows:

Table 3. Some seed words

theme	Seed word
Convenience	寄送，取货，距离，方便，区域，自取
reliability	保证，发错，失职，丢失，包装，准确
Responsiveness	速度，效率，延误，滞留，期限，催单
Interactivity	素质，耐心，踢皮球，沟通，妥善处理
Economical	收费，贵，服务费，成本，损失，金额
safety	液体，药品，违禁品，重要文件

4.3. Topic Extraction

Substitute the seed words into the lda model, and the running result is shown in Figure 3, from which we can get the probability that each text contains a certain topic (Note: each document can contain multiple topics at the same time). 100 cases have been manually labeled. For the comparison of the topics, the analysis found that it is appropriate to set a threshold of 0.15, that is, when the probability of a topic to which this article belongs is> 0.15, it is determined that it belongs to the topic.

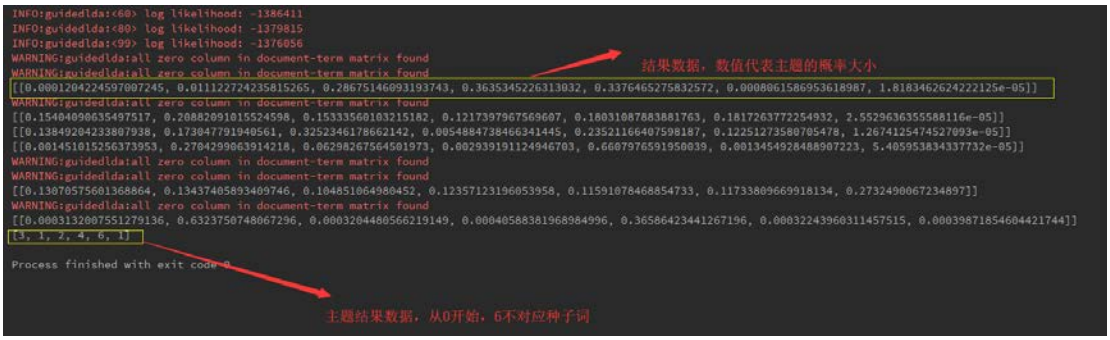

Figure 3. Subject extraction results

By using weakly supervised lda topic model analysis to complete a total of 9377 data, we obtained the distribution of topics involved in the negative evaluation case of logistics services, as shown in the following table:

Table 4. Analysis of influence of various topics

theme	Convenience	reliability	Responsiveness	Interactivity	Economical	safety
Influence	976	2724	1304	7188	956	384
Impact ratio	10.41%	29.04%	13.91%	76.66%	10.20%	4.10%

(Note: Total number of documents 9377)

4.4. Theme Impact Analysis

From the above results, we can see that in the negative evaluation of the service quality of logistics enterprises, the six dimensions of influence from high to low are interaction, reliability, responsiveness, convenience, economy, and security. Among them, interaction According to the analysis results of this article, we propose the following specific countermeasures and suggestions to logistics companies to improve their service quality.

(1) Continuously improve awareness of logistics services

In the negative evaluation of logistics companies, the interaction has the largest impact, which fully reflects the high requirements of customers for the service awareness of logistics companies. Modern logistics companies should not only have qualified transportation capabilities, but also continuously improve their service awareness. Image, service personnel skill level, business quality, ability to communicate with customers, handling logistics customer complaints, and responding satisfactorily, we must improve customer satisfaction by establishing relevant systems and improve the overall level of logistics services.

(2) Strengthening logistics service management

The reliability dimension is the most important requirement for customer perception, and it includes the punctuality, accuracy, and stability of logistics services provided by logistics companies. From the analysis results, it can be seen that the proportion of reliability in the negative evaluation of logistics services is only lower than The interactivity indicates that reliability is still an important aspect of customer perception of logistics services, so logistics enterprises should continue to strengthen their logistics service management.

(3) Vigorously develop logistics information technology

In the process of modern logistics services, the proportion of responsiveness is getting higher and higher, which reminds various logistics companies to realize the full informationization of the logistics process as an inevitable requirement. Some logistics companies do not focus on collecting customer demand information and have not established it. Customer information management system, these factors will affect the security, accuracy and timeliness of logistics services. Therefore, logistics enterprises should strengthen the investment and construction of information technology, establish a logistics information management system, and use the information platform to realize the logistics process Sharing information with customers, improving the customer's response capabilities, and the application level of information technology will be the key factors that determine the future development space of logistics companies.

(4) Pay attention to the cost of logistics services

The economic dimension includes relatively reasonable fees paid for logistics services, the logistics methods provided are adapted to customer needs, and are reasonable. The value-added services provided are reasonable and reasonable. They are welcomed by customers. The analysis results reflect the current content of this part. Satisfaction among consumer groups is far from enough, and logistics companies need to pay enough attention to the relationship between service costs and customer satisfaction. Logistics costs and the quality of service provided by logistics are a pair of counter-benefit relationships, and logistics companies should establish science The logistics service pricing mechanism allows customers to accept reasonable logistics service fees.

5. Summary and Discussion

This paper aims at the negative evaluation of the service quality of logistics companies, uses emerging text processing technology to perform data mining and preprocessing, and then performs word segmentation and feature word clustering on the processed data.Finally, a weakly supervised lda topic model pair is used. The collected data were analyzed. Based on the objective and accurate basis, several suggestions were made to the logistics enterprises based on the research results, in order to improve the service quality of the logistics enterprises, further improve consumer satisfaction, and maintain the logistics enterprise's Long-term development has brought greater benefits to logistics companies.

The difference between this paper and the existing research is that the data set is the centralized negative evaluation of logistics services, rather than the data of previous positive and negative reviews. Compared with the mixed data research, the centralized negative evaluation used in this paper is conducive to accurate analysis. Judging the inadequacy of the logistics service level of the logistics enterprises, so that the logistics enterprises can identify their own shortcomings and carry out targeted rectification measures for different aspects. At the same time, the weakly supervised lda model used in this article is based on the preset theme. In the case of semi-supervision, using seed words to analyze the data, this method is more suitable for use when a corresponding evaluation system is preset. The shortcoming of this article is that the evaluation system used is not yet complete and can be continued to supplement In order to more comprehensively reflect the results of data analysis, in addition, the weakly-supervised lda model used in this paper is mentioned less in the current research, and its accuracy and precision can still be improved, which needs to be further improved.

References

[1] State Post Office of the People's Republic of China. 2018 China Express Development Index Report [R] .Beijing: State Post Office of the People's Republic of China, 2018, from http: // www.spb.gov.cn/xw/dtxx_15079/201904/t20190417_1814716.html.

[2] Zheng Na, Zhang Jia, Ma Li. Development status and countermeasures of e-commerce logistics in China [j] .China Business, 2011, (11): 124-125.

[3] Wu Anbo, Sun Linhui, Liu Zhenyu. Discussion on the Storage and Distribution Mode of Fresh Agricultural Products in the E-commerce Environment [j] .Commercial Economics Research, 2017, (24): 92-94.

[4] Qi Lei, Sun Wenxian. Research on Factors Influencing Consumers' Online Shopping Based on Customer Satisfaction [j] .Jiangsu Commercial Theory, 2016, (1): 25-28.

[5] Yang Yongqing, Yu Benhai. An Empirical Analysis of Consumers' Post-buy Behavior Based on Logistics Services [j] .Chinese Journal of Management, 2014, (3): 434-438.

[6] Mei Hu, Wen Zhian. Effects of Online Logistics Review Information on Consumers' Purchase Intent [j] .Logistics Technology, 2016, (1): 51-54.

[7] Mentzer J T, Flint G J,Kent J L. Logisties Service Quality as a segment- Customized Process [J]. Journal of Marketing,2001,65(4):82-104.

[8] Li Aiguo, Deng Zhaohui, Mao Bingjie. The Impact of Online Negative Reviews on Sales of Experiential Products-Based on the Perspective of Merchant Response [j] .Business Research, 2016, (7): 138-144.

ISSN: 2688-9323

[9] Gronroos C.A Service Quality Model and It s Marketing Implication[J].European Journal of Marketing, 1982,18(4):36-44.

[10] Gronross C. Applied Service Marketing Theory[J].European Journal of Marketing,1984,16 (7):30-41.

[11] Mentzer JT,Flint DJ, Hu It GTM.Logistics service quality as a segment-customized process [J].Journal of marketing.2001,65(4):82-104.

[12] Mentzer J T,Flint D J. Logistics Service Quality as a Segment-Customized Process[J].Journal of Marketing,2013,65(4):82-104.

[13] Wei Fuxian. Evaluation and Management of Service Quality [m] .Beijing: People's Posts and Telecommunications Press, 2005.

[14] CAROL C B. An expanded Model of logistics Service Quality: Incorporating Logistics Information Technology[J].International Journal of Production Economics,2007(3):205-222.

[15] Chen Fang. Research on b2c e-commerce logistics service quality evaluation from the perspective of customer satisfaction [d] .North China University of Water Resources and Hydropower, 2018.

[16] Zhou Zhengsong, Shi Guohong. Research on Service Quality Evaluation of Logistics Enterprises Based on servqual and lsq Models [j] .Science and Technology Management Research, 2012,32 (06): 27-29 + 34.

[17] Sun Liwei, He Guohui, Wu Lifa. Research on web crawler technology [j] .Computer Knowledge and Technology, 2010,06 (15): 4112- 4115.

[18] Chen Yi. Research on the Application of Text Mining in Online Shopping User Reviews [d] .Zhejiang Gongshang University, 2018.

[19] Zhang Qiyu, Zhu Ling, Zhang Yaping. Review of Research on Chinese Word Segmentation Algorithms [j] .Information Research, 2008 (11): 53-56.

[20] Lao Xin. Research on the Impact of Online Reviews on Product Sales [d] .Beijing University of Posts and Telecommunications, 2017.

[21] Wu Weijun. Research on the usefulness of comments based on semantic sentiment information [d] .Chongqing University, 2016.

[22] ZHOU Hongqing, WU Yangyang. Extracting and Clustering Features of Chinese Customer Comment Objects [j] .Microcomputer & Application, 2014,33 (15): 69-71.

[23] Blei D M,Ng A Y, Jordan M I. Latent Dirichlet allocation[J].Journal of Machine Learning Research, 2003, 3:993-1022.

[24] Tang Xiaobo, Xiang Kun. Hotspot mining based on lda model and Weibo popularity [j] .Library and Information Service, 2014,58 (05): 58-63.

[25] Wang Peng, Zhang Shaozhong. Research on feature mining of express service based on semantic analysis [j] .Journal of Zhejiang Wanli University, 2018,31 (05): 80-87.

[26] B. Palese, A. Usai. The relative importance of service quality dimensions in E-commerce experiences [J]. International Journal of Information Management,2018,40.

Author Contribution Statement:

Zhou Xue: conducting experiments, drafting and revising thesis;
Kaiqing: collect, clean and analyze data;

Conflict of Interest Statement:

All authors declare no conflict of interest.

Supporting Data:

The supporting data is stored by the author, E-mail: 168192985@qq.com.

ISSN: 2688-9323

[1] Zhou Xue.data.json. The complaint case dataset of the State Post Office complaint website.
[2] Zhou Xue. Seed Words.xlsx. Theme model seed word chart.
[3] Zhou Xue. Theme extraction results.xlsx. Use weakly supervised LDA for topic extraction results.