Article

Analysis of Customer Satisfaction in Tourism Services Based on the Kano Model

Kailin Zhou and Zhong Yao *

School of Economics and Management, Beihang University, Beijing 100191, China; klzhou1972@126.com
* Correspondence: iszhyao@buaa.edu.cn

Abstract: Understanding customer needs is of great significance to enhance service quality and competitive advantage. However, for the tourism industry, it is still unclear how to mine service improvement strategies from tourist-generated online reviews. This paper aims to develop a data-driven approach to conduct a fine-grained dimension analysis of customer satisfaction with tourism services. First, this paper uses Latent Dirichlet Allocation to explore the key dimensions of tourist satisfaction from online reviews. Next, based on the Chinese sentiment dictionary, tourists’ emotional attitudes towards each service dimension can be identified. Then, the backpropagation neural network is used to measure the complex relationship between tourists’ sentiment orientations towards different dimensions and their satisfaction. Finally, according to the improved Kano model, multidimensional attribute classification is realized to support the strategic analysis of tourism service quality improvement. The proposed method is empirically verified through a real tourism review dataset. The results exhibit the theoretical and practical implications of our method.

Keywords: customer satisfaction; Kano model; online review; backpropagation neural network; online reviews

1. Introduction

Customer satisfaction is a psychological state achieved by customers based on their subjective judgment of the degree to which products or services fulfill their requirements [1]. The inherent characteristics of tourism services determine that customer satisfaction is an important reference to evaluate its quality. Firstly, tourism services are experience-oriented services, and the quality of service largely depends on the experience and feelings of tourists themselves, which is difficult to be evaluated by consistent testing standards [2]. Secondly, tourists participate in and experience the whole process of the service, and have a tangible perception of its advantages and disadvantages. Thirdly, customer satisfaction has a strong correlation with recommendation intention [3], and improving customer satisfaction is an important driving force for the economic development of tourism enterprises. Some existing studies have shown that effective product improvement or Research and Development analysis can be conducted by modeling or measuring customer satisfaction, supplemented by other relevant models [4,5]. Consumer surveys or experiments are a typical way of measuring customer satisfaction [6]. However, this approach requires rigorous design and appropriate procedures to ensure the quality of participants’ responses. Undoubtedly, this is costly in money and time, and the data can become quickly outdated. For the tourism service industry, tourists’ consumption has the characteristics of one-time, mobility and hedonism, which makes it more difficult to carry out customer satisfaction surveys. Therefore, how to use the accumulated online tourism data resources to measure the customer satisfaction of tourism services is worth further research.

User-generated content (UGC) is the public content spontaneously generated by users in the Web2.0 environment, which includes various forms of data such as audio, video, text, pictures, etc. UGC contains a wealth of people’s attitudes, opinions and other information, and the research on its mining is becoming deeper and deeper. Among them, text data
have become one of the most important elements reflecting market sentiment, as well as one of the main forms of tourism’s Big Data [7]. Online travel reviews objectively reflect the tourists’ real perception of the attractions and services of the tourist destination, which is one of the most important ways of online word-of-mouth communication and is the most authentic portrayal of tourists’ views, feelings, and sentiments. It is also the true expression of tourists, which contains the tourists’ attitude towards the tourism service elements of tourism destinations. Therefore, travel online reviews contain rich information, which is of great value for managers and researchers to understand customer satisfaction [8]. In addition, compared with survey research or experimental research, online review data have the characteristics of public availability, low cost, spontaneous generation, great insight, and large number of participants [9], which make online review data more suitable for constructing comprehensive customer satisfaction models. With the above advantages in mind, online travel reviews are a potentially powerful data resource for understanding customer satisfaction with tourism services.

Some studies in the field of tourism have confirmed that customers’ travel sentiments have a significant impact on their satisfaction. For example, a survey on customer satisfaction of diving services shows that high customer satisfaction is closely related to sentiments, such as excitement, pleasure, awe, surprise, etc. All themes in the theoretical framework of diver satisfaction determined in this study are regulated by sentiments [10]. Geetha et al. [11] studied the relationship between customer sentiment and customer ratings in online reviews of the hotel industry, and found consistency between customer ratings and customer feelings for advanced and economic hotels, and that customer sentiment explained significant differences in customer ratings for the two hotel categories. In fact, the overall satisfaction of customers is a comprehensive consideration of the various tourism services provided. Similar to the research of Bi et al. [7], this paper regards customer satisfaction with tourism services as the result of the comprehensive effect of the perception (emotion) of customers in the multiple dimensions of tourism services they experience, which is consistent with our intuition. However, it is still unclear how to design service improvement strategies according to the association between customer satisfaction and multiple dimensions of service quality perception. This paper focuses on developing a data-driven approach to conduct a fine-grained dimension analysis of customer satisfaction with tourism services. We will achieve this aim by answering three sub-issues:

1. how to explore the key quality factors affecting customer satisfaction from the online travel review data of tourist attractions?
2. how to measure the complex impact of tourists’ emotional attitude in each key factor on overall customer satisfaction?
3. based on the influence of each key factor obtained, how to classify their attribute types and gain insight into how these key factors play a role in improving customer satisfaction, especially when faced with severe challenges, such as the COVID-19 pandemic?

To solve these problems, this paper establishes a framework for constructing a customer satisfaction model from online travel reviews. In this framework, firstly, the Latent Dirichlet Allocation (LDA) model is used to extract important tourism service features that customers care about from online reviews of tourist attractions. Then, based on the Chinese sentiment dictionary, emotional attitudes (positive or negative) about these important characteristics can be identified in the review data. Then, to measure how customer sentiment of each feature affects customer satisfaction, a measurement method based on BPNN is proposed. On these bases, combined with the Kano model, the dimension of customer satisfaction is classified. Finally, this paper uses the review data of tourist attractions in Guizhou province, since the outbreak of COVID-19, to verify the proposed method and analyze the customer satisfaction of tourism services in Guizhou. Notably, Guizhou is a leading tourism province in China, and rich tourist review resources have accumulated on online platforms, influencing its choice as a typical example for this paper. Similarly, tourism services in other regions can be analyzed according to our method guideline, which exhibits generalization.
The rest of this paper is organized as follows. Section 2 reviews the relevant literature. Section 3 presents our proposed research method. Section 4 explains our method through empirical research. Section 5 demonstrates the theoretical and practical implications, and presents the limitations of our study with directions for future research.

2. Literature Review

In this section, we will review the research on customer satisfaction measurement, then introduce the Kano model.

2.1. Research on Customer Satisfaction Measurement

With the accumulation of massive online review data, many studies are emerging to analyze customer satisfaction directly or indirectly from online reviews. These studies can be mainly divided into three main streams: (1) to explore the key attributes affecting customer satisfaction from online reviews and conduct sentiment analysis; (2) research on the relationship between product/service attribute performances and customer satisfaction; (3) research on customer satisfaction model based on online reviews.

2.1.1. Attribute Extraction and Sentiment Analysis Based on Online Reviews

With the rapid development of the Internet, massive UGC has accumulated continuously. UGC contains rich and real customer perspective information and is regarded as a data resource with a strong potential for understanding and managing customer demands [12]. However, online review data exist in the form of free text. As a kind of unstructured data, text data cannot be directly analyzed. Therefore, the conversion of text review data into structured data that can be directly utilized is the basis for subsequent analysis. Customer opinion mining from online reviews is one of the most pursued areas of research, which is mainly carried out based on attribute extraction and sentiment analysis [13].

Attribute extraction refers to extracting topics that users pay frequent attention to from online reviews, as well as keywords related to topics. There are two main categories of attribute extraction methods: (1) The statistical model-based methods, such as association rule mining [14], Hidden Markov Model [15], LDA model [8], etc. Among them, the LDA model becomes one of the most widely used models. For example, the study of Tirunillai and Tellis used the improved LDA model to propose a unified framework for extracting tourism service attributes from online reviews [16]. Another study utilized the LDA model to identify the key dimensions of customer satisfaction from 266,554 pieces of online review data [8]. (2) The rule-based method, which formulates the corresponding extraction rules according to the characteristics of the review text and the research goal to realize the extraction of attributes. For example, Kang and Zhou [17] proposed an unsupervised rule-based approach to identify subjective and objective characteristics from online reviews. Rana and Cheah et al. [18] defined a rules-based sequential pattern for online review mining and proposed a rules-based two-stage extraction model for dimensional extraction.

Sentiment analysis can mine the sentiment information hidden in online reviews to help understand customers’ emotional attitudes toward product and service attributes. Sentiment analysis methods are mainly divided into lexical sentiment analysis, such as dictionary-based sentiment analysis and corpus-based sentiment analysis [19] and sentiment analysis based on machine learning, such as support vector machine (SVM), Naive Bayes, unsupervised machine learning [20], etc.

2.1.2. Research on the Influence of Attribute Performance on Customer Satisfaction

On the basis of attribute and sentiment mining, the subsequent research mainly focuses on analyzing the impact of tourism service attribute performances on customer satisfaction in online reviews. It is worth noting that since online reviews express customers’ views and feelings, attribute performances in related studies refer to customers’ feelings on product/service attributes, which are usually expressed by customers’ emotional attitudes...
towards attributes. In the field of tourism, the several present studies mainly focus on the hotel industry. For example, a recent study examined how cultural traits affect the role of attribute-level experience on tourist satisfaction, and used a deep learning algorithm to propose an attribute-level sentiment analysis model to extract tourists’ attribute-level experience from online reviews. An empirical study based on nearly 50,000 online reviews collected by TripAdvisor found that positive/negative attribute experience has different impacts on customers with different cultural traits [21]. In addition, to understand the customers’ demands of five-star hotels, Bi et al. [7] proposed an online review mining method from the perspective of attribute importance–performance analysis (IPA). In this method, LDA is first used to find several useful hotel attributes from online reviews, then SVM is used to analyze customers’ performance feelings on these attributes in the reviews, and then an integrated neural network model is used to calculate the importance of attributes. Finally, an IPA diagram is constructed according to the results to analyze customer demand. From the perspective of service improvement, Zhang et al. [22] based on the existing research on the relationship between service performance and customer satisfaction, and considering the influence of consumer expectations and subjective opinions of management, proposed an online review-driven method to determine the priority of hotel service resource allocation. In this method, the LDA model is used to extract service attributes, and the recursive neural tensor network is used to divide the attribute sentiments involved in reviews into five categories. Then, the traditional PRCA model is improved to analyze the asymmetric relationship between attribute performances and customer satisfaction. On this basis, the customer mention frequency is calculated, the customer satisfaction function is constructed, and finally, the improvement strategy analysis is realized under the framework of the Kano model.

These studies have made outstanding contributions to the tourism field to gain insight into customer satisfaction from online reviews. However, there are still limitations in these research methods. First of all, for the research that extracts service attributes from online reviews for performance analysis, there is no quantitative measurement of the impact of these attributes on customer satisfaction, such as the following study [7]. Second, even if the influence of service attributes on customer satisfaction is measured, these studies are usually based on the assumption that customer satisfaction (online ratings) follows a Gaussian distribution, and the relationship between satisfaction and all attribute tendencies follows additive independence, such as the following study [22]. In fact, in many real situations, customer satisfaction follows a positive skew, asymmetric, bimodal (or J-shaped) distribution [23,24]. At the same time, because the attributes automatically mined from reviews are not as rigorous as those in a well-designed questionnaire, there may be more complex multilinear or nonlinear relationships between different attributes and customer satisfaction. Third, some studies do not pay attention to the categories of service attributes, such as the following study [21]. However, some studies have confirmed that service attributes can be divided into different categories, and the attributes of different categories will affect customer satisfaction in different ways [25,26]. For example, performance attributes will cause dissatisfaction when they are not implemented, but satisfaction when they are implemented; however, a higher degree of realization of the reverse attributes will lead to an increase in tourist dissatisfaction. Recently, the empirical study of Xu et al. [27] also showed that the attribute type moderates the impact of perceived attribute experience on overall satisfaction. Therefore, identifying the category of tourism service attributes is helpful to provide a clearer improvement direction for promoting tourism satisfaction and realize a more effective allocation of tourism service resources, which is necessary for further research.

In summary, the extant related literature does not provide an effective scheme to design tourism service improvement strategies from abundant UGC. We will develop a comprehensive approach including multiple techniques to model the complex relationship between customer satisfaction and multiple dimensions of service quality perception, and to classify the types of service attributes.
2.2. Kano Model

The Kano model is a two-dimensional demand analysis model that classifies product/service attributes into different categories. According to the realization degree of attributes and their impact on customer satisfaction, Kano divides into five types of attributes, as shown in Figure 1 [25]:

Figure 1. Traditional Kano model [25].

(1) Performance attributes: The realization of such attributes is positively related to tourist satisfaction. When such attributes are not implemented, tourists will be dissatisfied; when implemented, they will bring satisfaction to tourists;
(2) Excitement attribute: This kind of attribute will make tourists satisfied when it is realized, but will not lead to dissatisfaction when it is not realized;
(3) Must-be attributes: the realization of such attributes is taken for granted by tourists, and will not improve satisfaction; however, not being realized will lead to dissatisfaction of tourists;
(4) Reverse attributes: the higher the realization of such attributes, the more dissatisfied tourists will be;
(5) Indifferent attributes: the realization degree of such attributes has nothing to do with tourist satisfaction, or has only a very small impact.

3. An Approach to Modeling Tourist Satisfaction in Online Travel Reviews

In this section, we propose our methodological framework, and then detail each part in the following.

3.1. Methodological Framework

This section introduces a methodological framework for modeling tourist satisfaction from online travel reviews, as shown in Figure 2. For a clearer introduction to the method framework, first define some basic concepts covered in this article.
Tourists usually evaluate tourism services based on their perception of some important attributes of the tourism services they experienced. Similar to the research of Guo et al. [8], this paper defines these important attributes of tourism services as TSDs;  

1. Online travel reviews: An online travel review is text generated by a tourist, including the tourist’s opinions on the tourism services he/she experienced;  
2. Tourist satisfaction: Tourist satisfaction is a kind of psychological state, resulting from tourists’ overall subjective evaluation of tourism services based on their expectations and actual performance [7]. In previous studies, online ratings of customers are usually used to represent customers’ satisfaction with products and services [8,16,28,29]. Following these studies, this paper will use online ratings of tourists to express their satisfaction levels with the tourism services they experienced;  
3. Tourist satisfaction dimension (TSD): Tourists usually evaluate tourism services based on their perception of some important attributes of the tourism services they experienced. Similar to the research of Guo et al. [8], this paper defines these important attributes of tourism services as TSDs;  
4. Classification of TSDs: In this paper, under the classification framework of the Kano model, TSDs are divided into five categories, namely, performance TSD, excitement TSD, must-be TSD, reverse TSD, and indifferent TSD, whose meanings correspond to the five basic attributes of Kano model.

Based on Figure 2 and an explanation of the basic concepts, let’s look at the three main parts of the framework.
3.1.1. Mining Tourists’ Sentiments toward TSDs from Online Reviews

The collected online review data exist in the form of free text, which cannot be directly analyzed. In this part, through various text processing technologies, tourists’ positive or negative sentiments on different TSDs are mined from online text reviews, so as to transform unstructured online text reviews into a structured data matrix for modeling tourist satisfaction. Specifically, this part consists of two stages: (1) mine TSDs from online reviews using the LDA model; (2) identify the sentiment orientations of the review data for each TSD based on the sentiment dictionary.

3.1.2. Identifying the Category of Each TSD

Tourist satisfaction is expressed by the online rating given by tourists. The process of tourists giving the overall rating is a comprehensive evaluation of the performance of the tourism services in various aspects. Thus, tourist satisfaction can be viewed as a complex combination of tourist sentiments regarding the multiple TSDs covered in their reviews. Based on the structured review data obtained in the last stage, as well as overall tourist satisfaction (online ratings), this paper uses the BPNN network to depict the influence of tourists’ positive and negative sentiments of TSDs on their satisfaction, and then puts forward the calculation method of effect according to the weight parameters obtained by model training.

3.1.3. Measuring the Influence of Each TSD Sentiment on Tourism Service Satisfaction

Identifying the corresponding attribute category of each TSD under the Kano framework is conducive to improving the tourism services more effectively and thus improving tourist satisfaction. According to the importance calculated in the previous step, we can use the effect-based Kano model (EKM) to convert the extracted TSDs into five categories (performance TSD, excitement TSD, must-be TSD, reverse TSD, and indifferent TSD). Finally, the identified TSDs category plays an important role in the improvement strategy of tourism services.

Let’s go through each of these sections in detail based on the overall framework in Figure 2.

3.2. Excavating Tourists’ Sentiments on TSDs

3.2.1. Extracting TSDs Based on LDA

Previous studies have proved that LDA is an effective topic extraction method for online reviews [8,30]. LDA is a three-level Bayesian model, which assumes that each item in the set contains a finite number of topics. In the LDA model, words, topics, and documents are three important concepts. The word is a basic unit. Each document consists of multiple words and contains one or more topics. In this context, each online review can be viewed as a document, so the words in the document are the words in the review. According to the frequency of each word in each review, we can obtain the topic distribution of the review, and the word distribution of the topic by training the LDA model.

Extracting TSDs from travel reviews using LDA mainly includes two steps: preprocessing text reviews and extracting TSDs from review data:

1. Preprocessing text review data

In the Chinese travel text review data, there are not only words related to the required TSDs, but also a large number of noise words and irrelevant words, which will not become the target CSDs and will aggravate the data sparsity problem. Therefore, it is necessary to improve the effect of the LDA model through the pretreatment process. First, we divide the Chinese text review data into words. Then, we filter the corresponding words in the sentences according to the stop word dictionary (HIT Stop Word List, Chinese Stop Word List), negation dictionary, degree adverb dictionary, and sentiment dictionary. Let $R = \{r_1, r_2, \ldots, r_M\}$ denote the online review dataset, where $r_m$ ($m = 1, 2, \ldots, M$) denotes the $M$th text review, and $m$ is the number of reviews in the dataset. By counting the occurrence
frequency of each word in each preprocessed text review, we can obtain the review-word matrix $X_{M \times N}$, where $N$ denotes the number of words appearing in all the preprocessed review data;

2. TSDs extraction based on LDA

By using the obtained $X_{M \times N}$ matrix as input, the LDA model can be trained. The output of the trained LDA model has three parts, including review-topic matrix, topic-word matrix, and topic list. Because there are noise words in the obtained topics, and some subject words may have similar meanings, we can manually filter the noise words and merge subject words with similar meanings to obtain more reasonable results. Then, we select the appropriate topics from the results and assign the tag to each topic. As in some existing studies, each thematic term can be regarded as a TSD [7, 16]. Let $I$ denote the number of TSDs (i.e., topics), each consisting of multiple frequent words, and let $C_i$ denote the number of frequent words under the $i$th topic, so that the $i$th TSD can be denoted as $t_i = \{\text{word}_i^1, \text{word}_i^2, \ldots, \text{word}_i^{C_i}\}$, where $\text{word}_i^j (i = 1, 2, \ldots, I; j = 1, 2, \ldots, C_i)$ denotes the $j$th frequent word in the $i$th TSD.

3.2.2. Dictionary-Based TSD Granularity Sentiment Recognition

1. Review text decomposition

Typically, a single online review may contain several sentences related to different TSDs. Table 1 shows several examples of online travel review text. At the same time, the TSDs mentioned in different reviews may be different. To identify tourists’ emotional attitudes towards different attributes, it is necessary to extract the sentences containing each attribute from the original reviews, that is, decompose the original reviews into components under different TSDs. First, we divide the online reviews in $R$ into clauses according to punctuation marks.

<table>
<thead>
<tr>
<th>No.</th>
<th>Review Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Huangguoshu Waterfall lives up to its name and is spectacular.</td>
</tr>
<tr>
<td>2</td>
<td>This is a very huge water cave, beautiful scenery, can see very high, the cave is very cool, different sky, lights a dozen colorful, especially good looking, the ticket is not expensive. Welcome everyone here.</td>
</tr>
<tr>
<td>3</td>
<td>Although the service is not great, I have to say that the scenery is good, especially the waterfalls and the river valley are very commendable, but it is recommended to visit during the off-season.</td>
</tr>
</tbody>
</table>

Then, according to the obtained $t_i = \{\text{word}_i^1, \text{word}_i^2, \ldots, \text{word}_i^{C_i}\}$, we extract sentences from $R$ which contains $\text{word}_i^j (i = 1, 2, \ldots, I; j = 1, 2, \ldots, C_i)$, and obtain the review set $R_i$ related to the topic $t_i(TSD_i)$. In particular, if more than one sentence in a review is related to a TSD, the sentences are merged into one. If a review contains no sentence related to a TSD, the corresponding component is null.

Let $R_i = \{r_{i1}, \ldots, r_{im}, \ldots, r_{iM}\} (i = 1, 2, \ldots, I)$ denote the review set related to the $i$th TSD in the total review data $R$, $r_{im}$ denote the $m$th review in $R_i$, and $M$ denote the number of reviews in the dataset. To show this process more clearly, we take the reviews in Table 1 as an example. Through the above processing, we can extract the clauses related to TSDs from the three reviews. The specific results are shown in Table 2.
Table 2. Examples of short sentence extraction about TSDs.

<table>
<thead>
<tr>
<th>No.</th>
<th>TSDs</th>
<th>Ticket</th>
<th>Scenery</th>
<th>Overall</th>
<th>Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Huangguoshu Waterfall lives up to its name</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>the ticket is not expensive</td>
<td></td>
<td>beautiful scenery</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>especially the waterfalls and the river valley are very commendable</td>
<td>-</td>
<td>the scenery is good</td>
<td>the service is not great</td>
<td></td>
</tr>
</tbody>
</table>

2. Review text decomposition

In order to identify the sentiment orientations of each review in $R_i$, this paper will use a Chinese sentiment dictionary for sentiment analysis. Dalian University of Technology Chinese Emotion Vocabulary Ontology Database (called EMO_DIC) covers commonly used sentiment words in Chinese context, which is a widely used sentiment dictionary in the field of Chinese sentiment analysis [31,32]. Referring to these studies, this paper uses EMO_DIC as the basic sentiment dictionary, which contains a total of 27,466 Chinese sentiment words. It divides Chinese sentiment into 7 categories and 21 sub-categories. Specifically, they include “joy”, “good”, “anger”, “sorrow”, “fear”, “evil”, and “shock”. As shown in Table 3, the dictionary classifies each sentiment word into a specific sentiment category and provides the sentiment polarity of the word (1 for positive, 2 for negative). The parts of speech include adjectives, nouns, etc.

Table 3. Examples of Dalian University of Technology Chinese Emotion Vocabulary Ontology Database.

<table>
<thead>
<tr>
<th>Words</th>
<th>Part of Speech</th>
<th>Classification of Emotion</th>
<th>Polarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>dirty</td>
<td>adj</td>
<td>NN</td>
<td>2</td>
</tr>
<tr>
<td>rescue</td>
<td>verb</td>
<td>PH</td>
<td>1</td>
</tr>
<tr>
<td>forces</td>
<td>adj</td>
<td>NN</td>
<td>2</td>
</tr>
</tbody>
</table>

Based on the sentiment orientations of each review in $R_i$, we can obtain the sentiment orientations of each review $r_m$ about $t_i$ (that is, the $i$th TSD), $i = 1, 2, \ldots, I, m = 1, 2, \ldots, M$, and convert the results into nominal encoded data, as shown in Table 4. The missing values in Table 4 indicate that there is no review about $t_i$ in the review $r_m$, or that there is no sentiment orientations shown in the review about $t_i$. Let $F = \{Pos, Neg\}$ denote the sentiment orientations of review $r_m$ about $t_i$. Through Formula (1), we can convert nominal encoded data into structured data, as shown in Table 5. As can be seen from Table 5, if $r_m$’s sentiment orientations about $t_i$ are positive, $(E^{Pos}_{im}, E^{Neg}_{im}) = (1, 0)$; If $r_m$’s sentiment orientations about $t_i$ are negative, $(E^{Pos}_{im}, E^{Neg}_{im}) = (0, 1)$; If sentiment orientations are missing, $(E^{Pos}_{im}, E^{Neg}_{im}) = (0, 0)$.

$$E^F_{im} = \begin{cases} 1, & \text{If the sentiment orientations } \in F \\ 0, & \text{others} \end{cases}$$

Table 4. Nominal coded data for each TSD in online travel reviews.

<table>
<thead>
<tr>
<th>Online Reviews</th>
<th>TSDs</th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_1$</td>
<td>Positive</td>
<td>Null</td>
<td>...</td>
<td>Null</td>
</tr>
<tr>
<td>$r_2$</td>
<td>Null</td>
<td>Positive</td>
<td>...</td>
<td>Negative</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$r_M$</td>
<td>Null</td>
<td>Null</td>
<td>...</td>
<td>Positive</td>
</tr>
</tbody>
</table>
Table 5. Structured online travel reviews.

<table>
<thead>
<tr>
<th>Online Reviews</th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>…</th>
<th>$t_I$</th>
<th>$r_1$</th>
<th>$r_2$</th>
<th>…</th>
<th>$r_M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_1$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$r_2$</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>$r_M$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>…</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

3.3. Evaluate the Impact of Each TSD’s Sentiment Orientation on Tourist Satisfaction

In this section, we propose a method based on backpropagation neural network (BPNN) to measure how tourists’ sentiments towards TSDs affect their satisfaction.

Most current studies modeling tourist satisfaction from online review data follow the following assumptions [28,29]: (1) assume that tourist satisfaction (i.e., online rating) follows a Gaussian distribution; (2) at the same time, it is assumed that tourist satisfaction is a linear combination of tourists’ emotional attitudes towards TSDs; (3) in addition, the multicollinearity between different TSDs is low. However, in many practical problems, these assumptions cannot be satisfied. In practice, the TSDs mined from online reviews and the online ratings of tourists usually have the following characteristics: (1) tourist satisfaction (online ratings) usually follows the positive skew, asymmetric, bimodal (or J-shaped) distribution [23]; (2) tourist satisfaction may be a nonlinear combination of sentiments toward TSDs; (3) there may be a multicollinearity relationship between the different attributes automatically mined from the reviews, and there may be a complex nonlinear relationship between the attributes and tourist satisfaction. In fact, tourist satisfaction is a complex union of their emotional attitudes towards the full range of TSDs involved in the reviews. Therefore, considering the above characteristics, this section proposes a new method to measure the impact of tourists’ TSD sentiments on tourist satisfaction.

Neural network (NN) is an effective prediction method. In some complex data environments (such as non-normal data, nonlinear relationship, multicollinearity, etc.), NN is significantly better than the multiple regression model because it is not affected by collinear independent variables and does not require the linear assumption of multivariate input variables and dependent variables [7,33]. Although NN is proposed to be used for prediction tasks, some studies have shown that it can also be used to determine the weight information of input variables [34]. For example, artificial NNs were utilized in the study [35] to evaluate the relative importance of the influence factors of consumer acceptance of behavioral-targeted advertising services. Therefore, NN is undoubtedly a competitive alternative method for measuring the influence of positive and negative TSDs sentiments of tourists on their satisfaction. BPNN is one of the most popular NN models, which will be used as the importance measurement technique in this paper. Specifically, this paper builds a BPNN model with three layers of network structure, including input layer, hidden layer, and output layer, as shown in Figure 3 below. Generally, BPNN includes two processes of forward information propagation and reverse error propagation to train the model. In the forward process, the input node transmits the tourists’ sentiment information about each TSD to the hidden node, and then the hidden node transmits the corresponding information to the output layer through the activation function. In the reverse process, according to the error between the model calculation results and the real results, the gradient descent method is used to minimize the error, and the model parameters (namely the weight) are updated [36].
The following describes in detail the method of calculating the impact of tourist’ TSD sentiment on tourist satisfaction based on the BPNN training.

1. Let \( b \) denote the trained BPNN model. Let \( w_{ih}^{\text{Pos}} \) and \( w_{ih}^{\text{Neg}} \) denote the weight between the input nodes \( E_{i}^{\text{Pos}} \) and \( E_{i}^{\text{Neg}} \) in BPNN \( b \) and the \( h \)th hidden node, respectively (the blue line in Figure 3), where \( i = 1, 2, \ldots I, h = 1, 2, \ldots, D \). Let \( w_{h} \) denote the weight between the \( h \)th hidden node in \( b \) and the output node (yellow line in Figure 2), \( h = 1, 2, \ldots, D \). Let \( W_{i}^{\text{Pos}} \) and \( W_{i}^{\text{Neg}} \), respectively, denote the influence of positive and negative sentiments of tourists towards each TSD on tourist satisfaction. \( W_{i}^{\text{Pos}} \) and \( W_{i}^{\text{Neg}} \) can be obtained from Formulas (2) and (3), respectively.

\[
W_{i}^{\text{Pos}} = \frac{\sum_{h=0}^{D} (w_{ih}^{\text{Pos}} \times w_{h})}{\sum_{i=0}^{l} \sum_{h=0}^{D} |w_{ih}^{\text{Pos}} \times w_{h}| + \sum_{i=0}^{l} \sum_{h=0}^{D} |w_{ih}^{\text{Neg}} \times w_{h}|}, \quad (i = 1, 2, \ldots, I), \tag{2}
\]

\[
W_{i}^{\text{Neg}} = \frac{\sum_{h=0}^{D} (w_{ih}^{\text{Neg}} \times w_{h})}{\sum_{i=0}^{l} \sum_{h=0}^{D} |w_{ih}^{\text{Pos}} \times w_{h}| + \sum_{i=0}^{l} \sum_{h=0}^{D} |w_{ih}^{\text{Neg}} \times w_{h}|}, \quad (i = 1, 2, \ldots, I). \tag{3}
\]
(2) In order to reduce the overfitting problem and enhance the reliability of the results, we conduct 10-fold cross-validation on the dataset. According to Equations (2) and (3), we calculate the sentiment weight information of the 10 trained BPNN models, respectively, and take their average value as the final required result, denoted as $\overline{W}_i^{\text{Pos}}$ and $\overline{W}_i^{\text{Neg}}$, $i = 1, 2, \ldots, 10$.

(3) Based on the calculated $\overline{W}_i^{\text{Pos}}$ and $\overline{W}_i^{\text{Neg}}$, we can evaluate the total impact (relative importance weight) of the $i$th TSD on tourist satisfaction. Let $R_i$ denote the range of influence of the $i$th TSD on tourist satisfaction. $R_i$ can be calculated by Formula (4):

$$R_i = \left| \overline{W}_i^{\text{Pos}} - \overline{W}_i^{\text{Neg}} \right|, \ (i = 1, 2, \ldots, 1).$$

(4) Let $TE_i$ denote the total impact (importance weight) of the $i$th TSD on tourist satisfaction, $i = 1, 2, \ldots, 10$. Then, according to the studies [37,38], the formula for calculating $TE_i$ is (5):

$$TE_i = \frac{R_i}{\sum_{i=1}^{10} R_i}, \ (i = 1, 2, \ldots, 1).$$

3.4. TSD Category Recognition Based on Kano Model

According to the obtained $\overline{W}_i^{\text{Pos}}$ and $\overline{W}_i^{\text{Neg}}$, as well as the basic principle of the Kano model, we propose a model based on the effect of Kano (Effect-Based Kano (EKM)), which can identify the category of each TSD from the perspective of tourists. The core idea of EKM is shown in Figure 4.

![Figure 4. Effect-based Kano Model (EKM).](image)

(1) In Figure 4a, positive sentiment is considered as the performance of the TSD achieves the requirements of tourists (that is, the green rectangle in Figure 4a); in contrast, negative sentiment is considered to be when the performance of the TSD does not achieve the requirement of tourists (the red rectangle in Figure 4a). In addition, the online ratings of tourists indicate the overall satisfaction of tourists with the tourism services they enjoyed.

(2) In Figure 4b, with the introduction of $\overline{W}_i^{\text{Pos}}$ and $\overline{W}_i^{\text{Neg}}$, the traditional Kano model framework is divided into two parts. Among them, the right side is the part related to positive sentiments, that is, the requirements for the TSD are fulfilled. Meanwhile, $\overline{W}_i^{\text{Pos}}$ can be regarded as the influence of $t_i$ on the overall satisfaction of tourists when TSD $t_i$ is satisfied; accordingly, the left side of Figure 4b is the part related to negative sentiments, that is, the requirements of the TSD are not fulfilled. Meanwhile, $\overline{W}_i^{\text{Neg}}$ can be regarded as the influence of $t_i$ on the overall satisfaction of tourists when TSD $t_i$ is not satisfied. At this point, the detailed meanings of $\overline{W}_i^{\text{Pos}}$ and $\overline{W}_i^{\text{Neg}}$ are as follows:

(i) $\overline{W}_i^{\text{Pos}} > 0$ indicates that the overall satisfaction of tourists will increase when their requirements for $t_i$ are satisfied;

(ii) $\overline{W}_i^{\text{Neg}} < 0$ indicates that the overall satisfaction of tourists will decrease when their requirements for $t_i$ are not satisfied.
(ii) $W_{i}^{Pos} \leq 0$ indicates that the overall satisfaction of tourists will not increase when their requirements for $t_i$ are satisfied;

(iii) $W_{i}^{Neg} \geq 0$ indicates that the overall satisfaction of tourists will not decrease when their requirements for $t_i$ are not satisfied;

(iv) $W_{i}^{Neg} < 0$ indicates that the overall satisfaction of tourists will decrease when their requirements for $t_i$ are not satisfied.

In Figure 4c, $W_{i}^{Pos}$ and $W_{i}^{Neg}$ are denoted as the horizontal axis and vertical axis, respectively. Therefore, the TSD represented as a curve in Figure 4b can be converted into a point in Figure 4c. Thus, according to the basic principles of the Kano model, combined with $W_{i}^{Pos}$ and $W_{i}^{Neg}$, TSDs can be divided into five types in Figure 4c, the detailed explanation is as follows:

(i) If both $|W_{i}^{Pos}| < \left(\frac{1}{10^{27}}\right)$ and $|W_{i}^{Neg}| < \left(\frac{1}{10^{27}}\right)$, this indicates that $t_i$ has a very small effect on the overall satisfaction of tourists, and $t_i$ is an indifferent attribute. It is worth noting that this means the threshold that determines whether a CSD is an indifferent property; the classification conditions for other cases are as follows, (ii)–(v);

(ii) If $|W_{i}^{Pos}| \leq \left(\frac{1}{10^{27}}\right)$ and $|W_{i}^{Neg}| < \left(\frac{1}{10^{27}}\right)$, then $t_i$ is a must-be attribute, that is, when the requirements of tourists for $t_i$ are satisfied, the overall satisfaction of tourists will not increase. When they are not satisfied, the overall satisfaction of tourists will decrease;

(iii) If $|W_{i}^{Pos}| \leq \left(\frac{1}{10^{27}}\right)$ and $|W_{i}^{Neg}| \geq \left(\frac{1}{10^{27}}\right)$, then $t_i$ is a reverse attribute, that is, when the requirements of tourists for $t_i$ are satisfied, the overall satisfaction of tourists will not increase. When they are not satisfied, the overall satisfaction of tourists will not decrease;

(iv) If $|W_{i}^{Pos}| > \left(\frac{1}{10^{27}}\right)$ and $|W_{i}^{Neg}| < \left(\frac{1}{10^{27}}\right)$, then $t_i$ is a performance attribute, that is, when the requirements of tourists for $t_i$ are satisfied, the overall satisfaction of tourists will increase. When they are not satisfied, the overall satisfaction of tourists will decrease;

(v) If $|W_{i}^{Pos}| > \left(\frac{1}{10^{27}}\right)$ and $|W_{i}^{Neg}| \geq \left(\frac{1}{10^{27}}\right)$, then $t_i$ is an excitement attribute, that is, when the requirements of tourists for $t_i$ are satisfied, the overall satisfaction of tourists will increase. When they are not satisfied, the overall satisfaction of tourists will not decrease.

4. Empirical Research Based on Online Review Data

This section will verify the method proposed in Section 3 by online reviews posted by tourists. This study is carried out on a PC, which is configured with 8 GB memory space, 1.8 GHz dual-core Intel Core i5 processor, and an iOS operating system. Data processing and method implementation uses Python. The following chapters first introduce the experimental data, then explain the experimental procedures and some important experimental results.

4.1. Data Source

The tourism industry is a pillar industry in Guizhou, China. For this study, we crawled 10,307 online reviews of tourist attractions in Guizhou between 1 January 2020, and 31 May 2021, from Ctrip (www.ctrip.com, accessed on 8 June 2021), one of the most popular online travel platforms in China. The time span of the dataset includes the initial period of the COVID-19 outbreak to the period of the normal management of the pandemic, so it can reflect the views of tourists on the services of tourist attractions in Guizhou during the COVID-19 period. Figure 5 is an example of the data collected. As can be seen from the figure, the collected data include the review text of tourists, the ratings of tourists (overall satisfaction), the date of the reviews, etc.
The following chapters first introduce the tourism industry is a pillar industry in Guizhou, China. As can be seen in Table 6, \( C_i(i = 1, 2, \ldots, 18) \) is the number of frequent words included in \( t_i \) (i.e., TSD), and the total frequency is the total number of times the word \( \text{word}_{ij} \) in \( t_i \) appears in all reviews. As can be seen from the table, ‘Huangguoshu Waterfall’ and ‘sightseeing transportation in tourist attraction’ are the two TSDs about which the tourists are most concerned, with each TSD appearing more than 7000 times in the total review data. Secondly, two TSDs, ‘general feeling’ and ‘ticket service’, appear 6400 times and 4800 times, respectively. Next, the four TSDs of ‘Pingba Cherry Garden’, ‘Zunyi Red Culture’, ‘Fanjingshan’, and ‘Minority Culture’ all appear more than 3000 times.

**Table 6.** The extracted tourist satisfaction dimension based on LDA.

<table>
<thead>
<tr>
<th>No.</th>
<th>TSD</th>
<th>Total Frequency</th>
<th>No.</th>
<th>TSD</th>
<th>Total Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t_1 )</td>
<td>Natural scenery</td>
<td>11</td>
<td>( t_{10} )</td>
<td>Ticket Service</td>
<td>12</td>
</tr>
<tr>
<td>( t_2 )</td>
<td>Night view of Maotai Town</td>
<td>14</td>
<td>( t_{11} )</td>
<td>Zunyi Red culture</td>
<td>21</td>
</tr>
<tr>
<td>( t_3 )</td>
<td>Landmark buildings of Guiyang City</td>
<td>10</td>
<td>( t_{12} )</td>
<td>Ancient town tourism</td>
<td>8</td>
</tr>
<tr>
<td>( t_4 )</td>
<td>The Hmong village</td>
<td>9</td>
<td>( t_{13} )</td>
<td>Huaxi Wetland Park</td>
<td>7</td>
</tr>
<tr>
<td>( t_5 )</td>
<td>Sightseeing traffic of tourist attractions</td>
<td>13</td>
<td>( t_{14} )</td>
<td>Minority Cultures</td>
<td>15</td>
</tr>
<tr>
<td>( t_6 )</td>
<td>Pingba Cherry Blossom Garden</td>
<td>12</td>
<td>( t_{15} )</td>
<td>Huangguoshu Waterfall</td>
<td>10</td>
</tr>
<tr>
<td>( t_7 )</td>
<td>Karst cave scenery</td>
<td>15</td>
<td>( t_{16} )</td>
<td>Overall feeling</td>
<td>12</td>
</tr>
<tr>
<td>( t_8 )</td>
<td>Qianling Mountain</td>
<td>14</td>
<td>( t_{17} )</td>
<td>Fanjing Mountain</td>
<td>11</td>
</tr>
<tr>
<td>( t_9 )</td>
<td>Adverse environmental factors</td>
<td>14</td>
<td>( t_{18} )</td>
<td>Tour guide and hotel</td>
<td>4</td>
</tr>
</tbody>
</table>

A major attraction with excellent overall management. One is that the arrangement of playing vehicles is good, and you can choose which one to play first, which helps to divert traffic. Another is that there are many places for dining in the scenic area. It is necessary to visit a scenic spot for 6 hours, and after seeing the affordable and high-quality products, KFC can meet the needs of different tourists without having to spend a lot of time in quality products. KFC can meet the needs of different tourists without having to spend a lot of time.
According to Section 3.1.2, the sentiment orientation of 10,307 reviews with respect to 18 TSDs can be decided. The statistical results for the sentiment orientation of each TSD across all reviews are shown in Figure 6. As can be seen from the figure, for all TSDs, the number of positive reviews outnumbers the number of negative reviews. Among them, \( t_{16}, \) i.e., the dimension of ‘overall feeling’, shows the largest number of positive sentiments, which is much higher than the other dimensions. On the contrary, \( t_5, \) i.e., the dimension of ‘scenic tourism transportation’ has the most negative sentiments. Next, we convert the nominal encoded data into structured data according to Formula (1).

\[
\text{Statistical results of TSD sentiment orientations in the review data.}
\]

4.3. Measuring the Influence of CSDs Sentiment on Tourist Satisfaction

We used the obtained structured data to train the BPNN model (the number of nodes in the hidden layer is 37), where the network parameters are set as the learning rate = 0.6, maximum allowable error = 0.01, and the number of iterations = 1000. In order to overcome the overfitting problem and enhance the reliability of the results, we performed 10-fold cross-validation on the dataset and recorded the weight parameter information of the 10 trained BPNN models. The RMSE values of the 10 models are shown in Figure 7. It can be seen from the figure that the RMSE values of the 10 models are all small, which can obtain better prediction performance. Finally, according to Formulas (2) and (3), we can calculate the importance information \( W_{i}^{\text{Pos}} \) and \( W_{i}^{\text{Neg}} (i = 1, 2, \ldots, 18) \) of 10 BPNN models about each TSD, respectively, and calculate their average value as the final \( W_{i}^{\text{Pos}} \) and \( W_{i}^{\text{Neg}} \) of each TSD, as shown in Table 7.

\[
\text{Table 7. The final } W_{i}^{\text{Pos}} \text{ and } W_{i}^{\text{Neg}} \text{ for each TSD.}
\]

<table>
<thead>
<tr>
<th>TSD</th>
<th>( t_1 )</th>
<th>( t_2 )</th>
<th>( t_3 )</th>
<th>( t_4 )</th>
<th>( t_5 )</th>
<th>( t_6 )</th>
<th>( t_7 )</th>
<th>( t_8 )</th>
<th>( t_9 )</th>
<th>TSD</th>
</tr>
</thead>
</table>
| Pos | -0.0034 | -0.0077 | -0.0018 | -0.0087 | -0.0051 | -0.0168 | 0.0013 | 0.0008 | -0.0087 | Neg | -0.0069 | -0.0096 | -0.0076 | -0.0030 | -0.0042 | -0.0070 | -0.0010 | 0.0023 | -0.0040 | -0.0069 | 0.0231 | -0.0036 | -0.0196 | -0.0027 | -0.0196 | -0.0034 | 0.0053 | -0.0059 | -0.0040 | -0.0079 | 0.0468

Figure 6. Statistical results of TSD sentiment orientations in the review data.
According to the classification results of TSDs, we can determine the TSD priority dimensions and their satisfaction. Finally, according to the improved Kano model, namely the EKM attribute classification process in Figure 8, all TSDs of most Guizhou tourism services belong to the must-be attribute, and no TSD is identified as the excitement attribute. Specifically, indifferent TSDs include ‘Huangguoshu Waterfall’, ‘Karst cave scenery’, ‘Huaxi Wetland Park’, ‘Zunyi Red Culture’, and ‘tourist traffic of tourist attractions’. Must-be TSDs include ‘Fanjing Mountain’, ‘Miao Village’, ‘natural scenery’, ‘Guiyang landmarks’, ‘tour guides and hotels’, ‘ticket services’, ‘night views of Maotai Town’, ‘Pingba cherry blossom Garden’, ‘ethnic culture’, and ‘ancient town tours’; performance TSDs include ‘Qianling Mountain’, ‘the overall feeling’; reverse TSDs are adverse environmental factors.

4.4. Measuring the Influence of CSDs Sentiment on Tourist Satisfaction

Based on the results in Table 7 and the EKM attribute classification process in Figure 4, the TSD categories of Guizhou tourism services are obtained, as shown in Figure 8. In the figure, 0.0056 (that is, 1/(10 × 18)) is the threshold that determines whether a TSD is an indifferent TSD (that is, the blue area in Figure 8). As can be seen from the figure, the TSDs of most Guizhou tourism services belong to the must-be attribute, and no TSD is identified as the excitement attribute. Specifically, indifferent TSDs include ‘Huangguoshu Waterfall’, ‘Karst cave scenery’, ‘Huaxi Wetland Park’, ‘Zunyi Red Culture’, and ‘tourist traffic of tourist attractions’. Must-be TSDs include ‘Fanjing Mountain’, ‘Miao Village’, ‘natural scenery’, ‘Guiyang landmarks’, ‘tour guides and hotels’, ‘ticket services’, ‘night views of Maotai Town’, ‘Pingba cherry blossom Garden’, ‘ethnic culture’, and ‘ancient town tours’; performance TSDs include ‘Qianling Mountain’, ‘the overall feeling’; reverse TSDs are adverse environmental factors.

![Figure 7. RMSE for 10-fold cross-validation.](image)

![Figure 8. Classification results of Guizhou TSDs.](image)
According to the classification results of TSDs, we can determine the TSD priority order in the formulation of the tourism service promotion strategy. First of all, the above 10 must-be TSDs represent the tourists’ basic requirements for Guizhou tourism services. When these requirements are not fulfilled, tourists will feel very dissatisfied. Therefore, relevant tourism managers should first examine must-be TSDs and try to fulfill the requirements associated with them. Then, since tourist satisfaction is directly proportional to the level of performance TSDs, relevant managers should consider fulfilling tourists’ requirements related to performance TSDs on the premise of fulfilling the must-be TSDs. Finally, for reverse TSDs, managers should carefully examine these factors to avoid such situations as much as possible.

5. Conclusions

Online tourism platforms provide an open and convenient channel for tourists to share their experiences in travel. At the same time, for the tourism industry, these online travel reviews also contain valuable information related to the quality of tourism services. This paper focuses on how to measure the key factors of customer satisfaction from the online travel review data and then classifies the key factors to provide the basis for the tourism service quality improvement strategy. To this end, this paper proposed a framework for modeling customer satisfaction from free review texts. First, this paper used the LDA model to explore the key dimensions of tourist satisfaction from online reviews. Next, based on the sentiment dictionary, we identified tourists’ sentiment attitudes towards all dimensions from these reviews; then, we used the deep learning network to measure the complex relationship between tourists’ sentiment orientations to different dimensions and their satisfaction. Finally, according to the improved Kano model, namely EKM, multi-dimensional attribute classification was realized to support the analysis of tourism service quality improvement.

5.1. Theoretical Contribution

The contribution of this paper has three aspects. First, this paper contributes to the research of online product/service reviews in the field of Information Systems (IS). Most of the related literature studies the impact of online reviews on customer purchases [39,40], explores the antecedents of the usefulness [41,42] of product reviews, or studies the influencing factors of online review behavior [43]. The research of this paper focuses on the mining of customers’ experience of service in online reviews, expands the scope of IS research, and puts forward an effective and feasible method to transform a large amount of online review data into useful business intelligence. Specifically, compared to [7,21–24], we model customer satisfaction more accurately through BPNN, and classify the types of service attributes through quantitative calculation process of EKM.

Secondly, this study contributes to the literature in the field of tourism. This study provides an effective methodological framework for analyzing tourist satisfaction from the perspective of customer demand. Generally, customer satisfaction is usually evaluated by conducting market surveys or experiments to collect and analyze customer preference data. However, this method of customer preference survey often requires a lot of time and cost. The method framework provided in this study can be refined into a dimensional evaluation of tourism service satisfaction based on online travel review data. Especially for the influence calculation of tourist satisfaction dimension based on BPNN, this method can model more complex data situations and does not require the assumption of Gaussian distribution of satisfaction (online rating) and the additive independence between all attribute sentiments, which is more realistic.

Thirdly, this study contributes to the literature in the field of market analysis. This study proposes an Effect-based Kano model (EKM), which is oriented by tourist satisfaction and realizes the classification and prioritization of tourists’ requirements in the dimension of service attributes. By perfectly integrating the influence of customer sentiment on satisfaction in service dimensions with the traditional Kano model, EKM can analyze and
visualize customer requirements for tourism services. Using the dataset in this paper, the validity of EKM can be empirically explained.

5.2. Practical Contribution

This article also contributes several valuable enlightenments to the tourism management industry. Comprehensively using the proposed method, this paper mines four types of information from online travel reviews, including TSDs of tourism service, tourist sentiments towards these TSDs, the impact of tourist sentiments towards TSDs on their satisfaction, and the category of each TSD. The four kinds of mined information can provide helpful information for real tourism industry managers, and further provide reference analysis methods for other product/service industries.

Firstly, the TSDs mined from online reviews usually represent the aspects that tourists care most about tourism services. Therefore, according to the extracted TSDs, managers can know which tourism service dimensions are of most concern to tourists.

Secondly, tourists’ views or sentiments on tourism service TSDs reflect their multidimensional perception of tourism services and can be regarded as the actual performance of the tourism service dimension provided. Therefore, managers can evaluate the performance of tourism services in each dimension based on customers’ views or sentiments on the service dimensions. On this basis, managers can further understand the advantages and disadvantages of tourism services.

Thirdly, the proposed method enables managers to assess the influence of tourists’ TSD sentiment on their satisfaction, which can be regarded as aggregated customer preferences, which is important for planning and decision-making related to service quality optimization.

Fourthly, the method in this paper can enable managers to understand the different categories of tourism service TSDs, which is crucial for managing tourist demands and deciding service optimization strategies. Specifically, the service characteristics of tourism are divided into five categories, including performance type, excitement type, must-be type, reverse type, and indifferent type. This understanding can help decision makers of the tourism industry to decide the priority of TSD development or promotion and, make more effective development plans for tourism service quality.

5.3. Limitations and Future Research Directions

There are also several limitations in this study, which deserve further attention. Firstly, this study assumes that all online reviews obtained from tourism social platforms are authentic. However, spam reviews, which may be manipulated or fake, are common on online platforms. At present, fake review recognition is receiving more and more research attention [44]. Therefore, it is a necessary and interesting research line to combine effective fake review recognition methods with this study. Secondly, a potential problem of this study is the representativeness of reviewers on online tourism platforms, that is, it is widely recognized that online reviewers may not be effective representatives of all target groups. Therefore, it is not reasonable to use online tourism reviews to infer the tourism service demands of all tourists for the representativeness of the reviewers and online reviews adopted. To reduce this concern, one possible direction for future research is to incorporate review data from multiple sources (multiple online tourism platforms) to reduce bias issues related to a single platform. Alternatively, we can collect some characteristic information of reviewers and adopt some statistical methods to extend the self-selection questions in online review data. Thirdly, this study analyzes all tourists as a whole. In fact, different consumer types often have different requirements for service attributes. In the future, we can further consider the customer type factor for analysis based on the current study. Fourthly, a potential research direction is to study other potentially more effective methods for analyzing tourist sentiment, such as building a professional sentiment dictionary in the field of tourism, so as to improve the accuracy of analysis results.


44. Ahmed, H.; Traore, I.; Saad, S. Detection of online fake news using n-gram analysis and machine learning techniques. In *Intelligent, Secure, and dependable Systems in Distributed and Cloud Environments*; ISDDC; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2017; p. 10618. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.