Research on Emotional Tendency of Catering Service Evaluation Based on LDA Topic model

Xiaojun Xu*

College of mechanical and electrical engineering, Nanjing Tech University Pujiang institute, Nanjing, China, 211222

*Corresponding author: 1600879748@qq.com

Abstract. To enhance consumers' dining service experience, this article constructs a mathematical model for emotional tendency analysis based on the text content of dining service evaluation. Firstly, clean the data of the comment text, and then use the *jieba* library to segment words. After removing the stopping words, the positive and negative parts are obtained. Secondly, use the *counter* library to calculate word frequency. calculate the number of emotional words, degree words, exclamation marks, and negative words, and classify and summarize them. Finally, sum up to obtain the total score of the comment, and use the LDA theme model based on various emotional feedback to obtain the overall strengths and weaknesses of the catering merchant. Research has shown that the advantages of merchants who gain the most positive emotions are: good taste, affordable prices, etc. The disadvantages of catering merchants with the most negative emotions are: slow delivery speed, incomplete preparation of tableware, etc. The improvement strategy is to accelerate delivery speed, patiently serve them, and make consumers eat well.

Keywords: LDA Topic Model, Analysis of emotional tendencies, Catering Service Evaluation.

1. Introduction

In China, the development of the catering industry is in full swing, and the vigorous development of the Internet has injected new vitality into the catering industry [1]. With the continuous stimulation of people's enthusiasm for online shopping, the catering industry has also developed a "new retail" business model [2]. The networking of the catering industry has become the norm, and the accompanying competition pressure within the industry is very prominent [3-4]. Therefore, for businesses, accelerating digital transformation, grasping user reviews, and mining useful information are crucial for improving services, understanding users' personal needs and preferences, and enhancing their competitiveness. For consumers, based on the comments of other buyers before purchasing, it is easy to understand the quality of the product, evaluate cost-effectiveness, and provide reference for their own consumption decisions [5]. However, originally, regarding the issue of emotional orientation in text, the models involved in analyzing the content were mainly traditional mechanical analysis, which could only study word frequency, and the selected vocabulary was mostly on the same topic. This article uses the LDA topic model to partition high frequency vocabulary, and the classification is mainly based on its basic definitions, which helps to understand the theme trends in large-scale text datasets, quickly obtain the core content of the text, and achieve more complete text classification.

2. Data sources and analysis

This article uses the data provided by question B of the 2022 National College Student Data Analysis Competition for analysis and research. (https://www.saikr.com/vse/Analysis)

3. Data preprocessing

3.1. Data cleaning

To analyze and evaluate the content, the first step is to perform data cleaning. Read the comment text, with a total of 17953 comment data. First, remove some special characters and import them into the *re* library in Python. Call the function ‘re.sub’ to remove special characters such as carriage returns, uninterrupted white-space, “×” Symbol, "☆" symbol, and some punctuation marks such as "!" "?" ":" ";". Then remove the word 'text' from the comments, and then remove emoticons from the HTML language.

3.2. Text deduplication

Considering that in practical situations, consumers may directly copy and paste other people's comments when commenting, or the system may automatically help customers make comments, which can change the frequency of comments appearing in the research process of text analysis, resulting in useless data. As it is impossible to determine whether it is a malicious comment swiping phenomenon, it is recommended to delete completely duplicate comments. After performing the text deduplication operation, there are still 17941 comments left, indicating that 12 duplicate comments have been deleted.

3.3. Segmentation and removing Stop words

Import *jieba*, a database used for word segmentation, and split the comment statement into several words. After word segmentation, a large number of auxiliary words such as "is, the" appeared in the comment. Therefore, it is necessary to treat these types of words as stop words and delete them. Here, stop words will be organized and summarized in the "stoplist. txt" file. The parts listed here are shown in Table 1.

Table 1. Partial stop words list

Number	words
0	a
1	able
2	above
3	&
...	...
736	yet
737	yes
738	yours
739	yourself

This is filtering the results of the previous segmentation step.

3.4. Word Frequency Statistics

Firstly, it is necessary to count the number of occurrences of the segmentation results implemented in the previous step. The word frequency can be placed in an empty dictionary data and implemented using two for loops. The core idea of the implementation is to use the word as a key in the dictionary if it appears for the first time, and assign the key a value of 1. If the word appears more than once, add one to the value after the word.

Secondly, the field "Target" is divided into two categories based on positive and negative emotions, and stored in "data_target0.xlsx" (positive emotions) and "data_target1. xlsx" (negative emotions), respectively. The frequency of each word is calculated, and the top 10 words for positive and negative emotions are listed in Table 2.

Table 2. Statistical Table of the Top 10 Words with the Most Positive and Negative Emotions

10 words with the most positive emotions		10 words with the most negative emotions	
words	frequency	words	frequency
good	3615	eat	2221
eat	2992	really	2005
taste	2370	taste	1264
delicious	2067	disgusting	1038
like	1495	hour	960
environment	932	order	911
feel	812	bad	907
service	676	deliver	866
price	676	dish	736
really	622	poor	727

3.5. Select the merchants with the most positive and negative emotions

In "data_target0.xlsx", the merchant with the most positive emotions is classified and summarized as Merchant 1041. Figure 1 shows the statistics of the number of positive evaluations from the top 10 merchants with positive emotions.

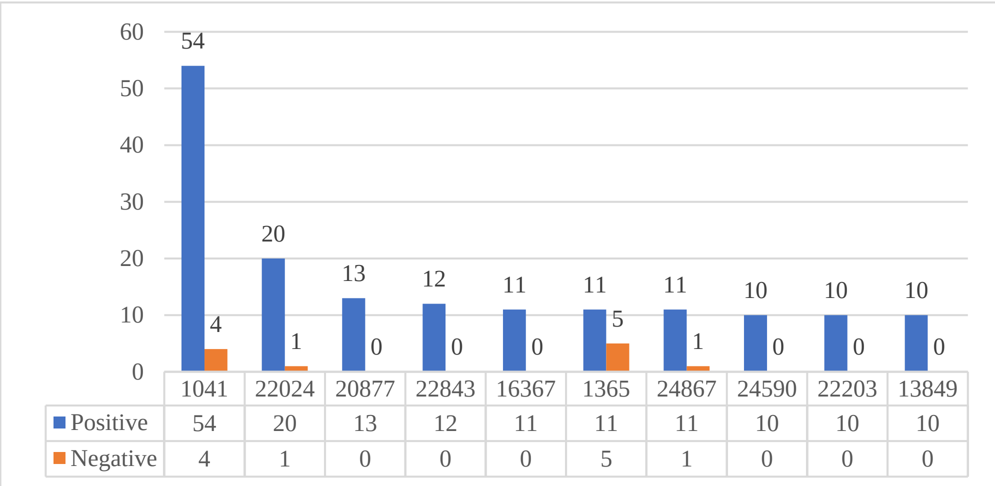

Figure 1. Statistical chart of the number of positive evaluations from the top 10 merchants with positive emotions

Similarly, in "data_target1. xlsx", the merchant with the most negative emotions is classified and summarized as Merchant 971. Figure 2 shows the statistics of the number of negative reviews from the top 10 merchants with negative emotions.

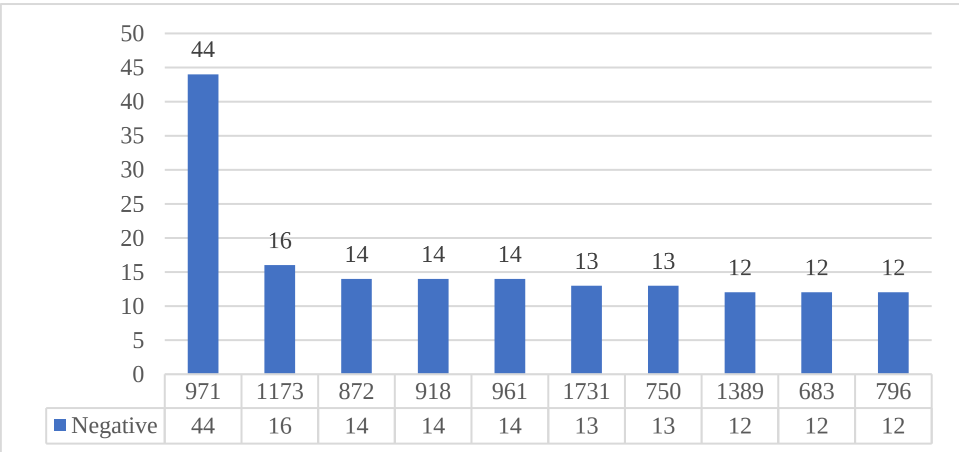

Figure 2. Statistical chart of the number of negative reviews from the top 10 merchants with negative emotions

4. Model establishment and solution

4.1. Determine the optimal number of topics

Topic coherence assessment is a method used to measure the quality of topics in the LDA (Latent Dirichlet Allocation) topic model. Its goal is to measure the semantic correlation between words in a topic, and a higher score indicates stronger correlation between topic words, which helps to understand and interpret textual data [6].

As shown in Figure 3, by calculating the coherence scores under different number of topics, it was found that the relatively high number of topics is distributed between 1-3. Considering that the information contained in the three topic numbers is relatively comprehensive, this article uses three topic numbers to construct the LDA topic model [7].

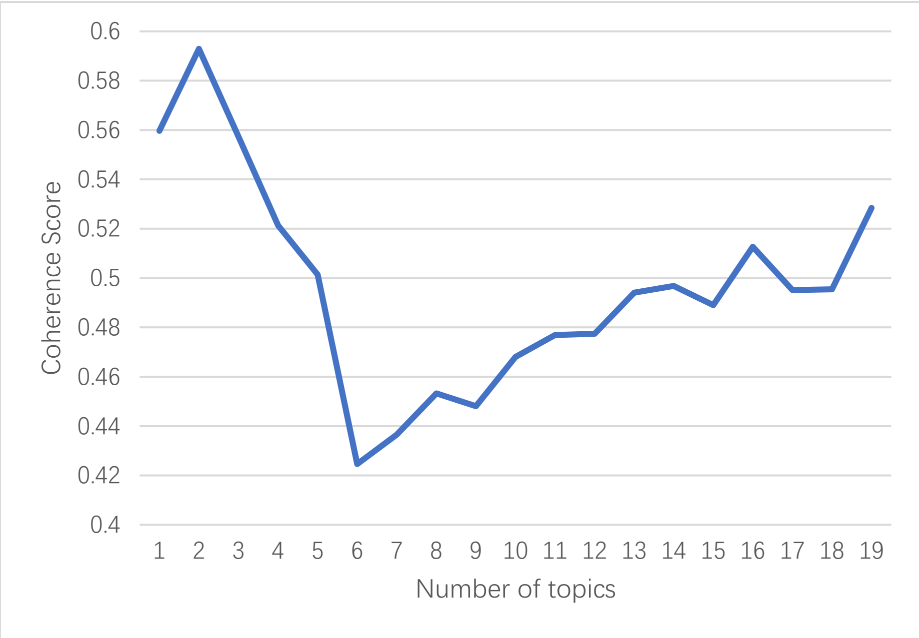

Figure 3. “topic-coherence”line chart

4.2. Analysis of merchants with the most positive emotions

Using the *LdaModel* class from the *Gensim* library to construct and train an LDA model, the top 10 keywords for each topic were printed, and the output results are shown in Table 3 and Figure 4.

Table 3. Word frequency distribution corresponding to each topic in positive comments

Topic	The weight of vocabulary in the original corpus is arranged in descending order									
	1	2	3	4	5	6	7	8	9	10
1	good	affordable	special	great	delicious	courier	very	deliver	speed	delivery
2	great	good	taste	service	speed	fast	packaging	satisfy	like	put
3	good	delicious	hardwork	clean	order	deliver	satisfy	on time	delivery	really

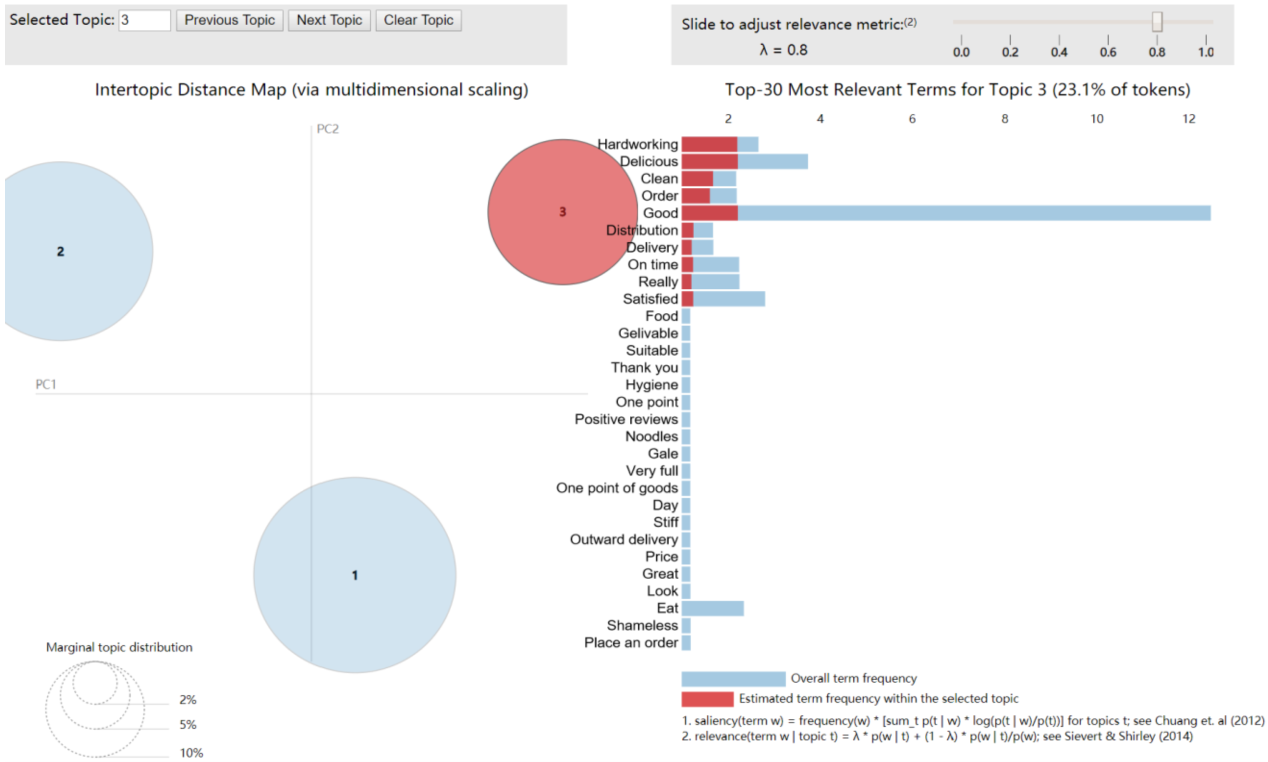

Figure 4. Visualization of LDA Theme Model Results with Positive Comments

For Topic 1, it can be summarized that the young man delivers meals quickly, at affordable prices, and is delicious; For Theme 2, it can be summarized as having a good taste, good service, and satisfactory packaging; For Topic 3, it can be summarized that the food is clean and delivered on time.

Therefore, the advantages of this merchant are summarized as follows: good taste, affordable price, good service, fast delivery speed, satisfactory packaging, and clean food.

4.3. Analysis of merchants with the most negative emotions

Similarly, the *LdaModel* class from the *Gensim* library was used to construct and train an LDA model, as shown in Table 4 and Figure 5.

Table 4. Word frequency distribution corresponding to each topic in positive comments

Topic	The weight of vocabulary in the original corpus is arranged in descending order									
	1	2	3	4	5	6	7	8	9	10
1	disgusting	dish	order	bad	smell	taste	slow	deliver	eat	poor
2	no	deliver	Chopsticks	really	hour	long	time	person	delivery	lot
3	few	attitude	poor	dish	operator	time	long	eat	express	Chopsticks
4	deliver	slow	call	no	delivery	hour	too	slow	time	cold

Regarding Topic 1, it can be summarized that the food is disgusting, tastes bad, and delivery is too slow; For Topic 2, it can be summarized that there are no chopsticks and long delivery times; For Topic 3, it can be summarized that there are few dishes, poor operator attitude, and long delivery time; For Topic 4, it can be summarized as follows: delivery is too slow, but the food has not been delivered yet when I call, and the dish is cold.

Therefore, the shortcomings of this merchant are: slow delivery speed, incomplete preparation of tableware, poor service attitude, and not only limited quantity but also poor taste of dishes. The improvement strategy is to accelerate delivery speed, not let consumers wait too long, patiently serve them, and improve the taste quality and quantity of dishes, so that consumers can eat well.

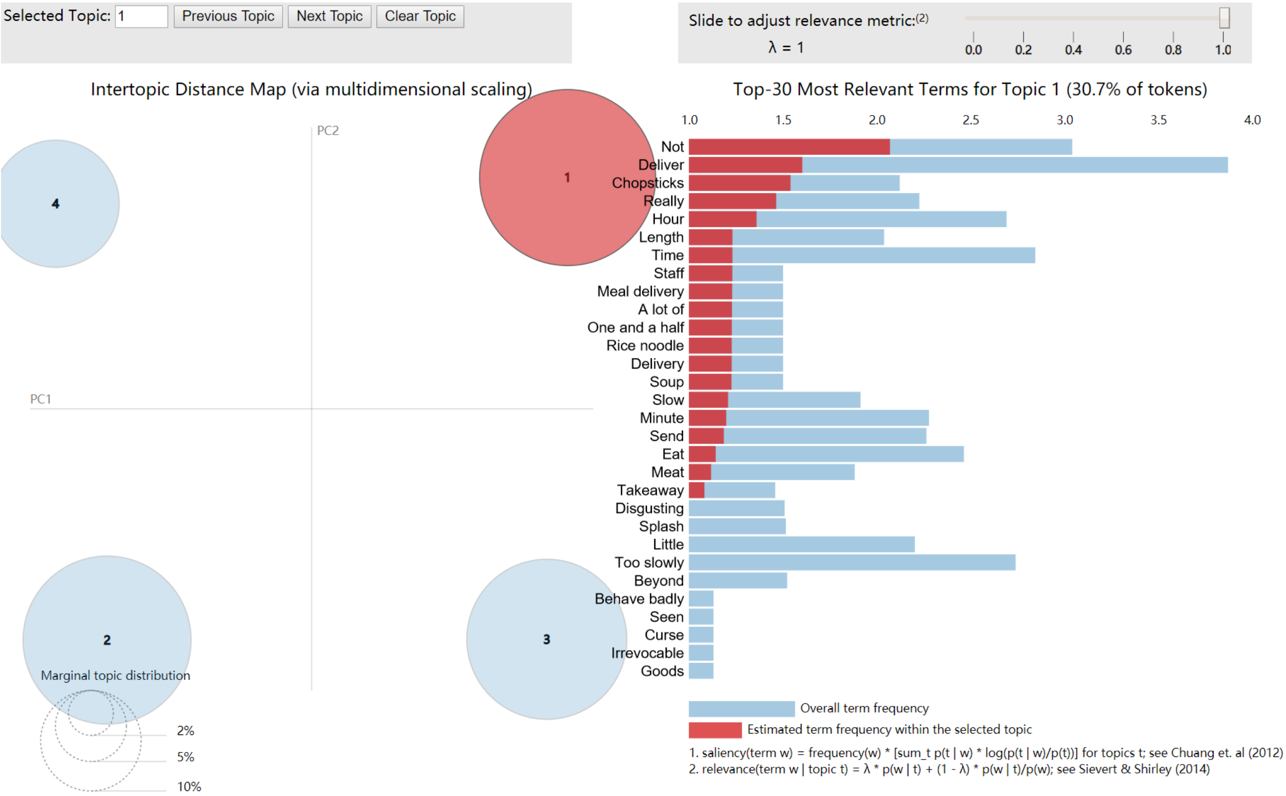

Figure 5. Visualization of LDA Theme Model Results with Negative Comments

4.4. Prediction of Customer's Emotional Tendency Analysis towards Merchants

4.4.1 Calculating Emotional Words

First, read the words in the text according to different parts of speech.

When a positive word appears in the text, add 1, and a negative word subtracts 1. For example, if "sad" appears, it means a negative word needs to be subtracted by 1. If there are two positive words, add 2, and the final score is 1.

4.4.2 Calculate degree words

When there are words such as "extremely", "very", "incomparable", and "too" in the text, then multiply the emotional score by 4.

When words such as "quite", "relatively", and "fairly" appear in the text, then multiply the emotional score by 2.

When words such as "only", "only", and "barely" appear in the text, then multiply the emotional score by 0.5 [8].

4.4.3 Calculate exclamation points

When an exclamation mark appears in the text, it indicates stronger emotions. At this point, an exclamation mark multiplied by 1.2 cannot be removed during text preprocessing.

4.4.4 Calculating Negative Words

When a negative word appears in the text, it indicates a change in the polarity of emotions. "Not delicious" is a common word in this text, so it should be multiplied by -1 to indicate the meaning of "disgusting " [9].

In addition, if you encounter a double negation, it is affirmative. Therefore, it is necessary to determine the number of negative words that appear. For odd words, multiply by -1, and for even words, multiply by 1.

4.4.5 Divide positive and negative segments

In a comment, there may be several sentences, or there may be both positive and negative comments, which means that one score cannot be used to express its emotional tendency. Therefore, for such comments, a positive score and a negative score should be given, and their sum is the final score [10].

If the final score is greater than 0, it is a positive emotion; if the score is less than 0, it is a negative emotion.

5. Conclusions

In the era of big data, the networking of the catering industry has become the norm. Extracting useful information from consumer feedback data plays an important role for both businesses and consumers. However, in the face of massive user comments, the method of manually reading reviews cannot timely examine and summarize consumers' comments after their dining experience. Therefore, based on the text content of the catering service evaluation, the article first cleans the data of the comment text, uses the Counter library to calculate word frequency, calculates emotional words, degree words, exclamation marks, and the number of negative words, and classifies and summarizes them. Finally, the total score of the comment is obtained by summing up, which can construct a mathematical model for emotional tendency analysis, obtaining positive and negative parts, And use the LDA theme model to identify the strengths and weaknesses of merchants. Research has shown that the advantages of catering merchants who gain the most positive emotions are: good taste, affordable prices, good service, fast delivery speed, satisfactory packaging, and clean food. The disadvantages of catering merchants with the most negative emotions are: slow delivery speed, incomplete preparation of tableware, poor service attitude, not only limited quantity of dishes, but also poor taste. Improvement strategies include: accelerating delivery speed, not letting consumers wait too long, patiently serving them, and improving the taste, quality, and quantity of dishes, so that consumers can eat well. In addition, the data is fully utilized by splitting words into different parts of speech, fully utilizing the language characteristics of Chinese, with high accuracy. The model preprocessing has been adjusted accordingly, using punctuation points such as "!" as a quantitative analysis of emotions, considering multiple factors and making the analysis more comprehensive. At the same time, the sentiment orientation analysis model constructed in this article can not only be used for reading comments from catering enterprises, but also for sentiment analysis of all e-commerce product comment data, with good promotion value.

References

[1] Zhang Xiaoyan. Research on the Impact of the Internet Economy on the Catering Industry [J]. Product Reliability Report, 2023 (06): 83-85.

[2] Dong Jiahang. Research on the Development Status and Upgrade Strategies of the Catering Industry under the New Retail Model [J]. China Collective Economy, 2022 (27): 85-87.

[3] Wang Yu. Exploring the Factors Influencing the Competitiveness of Catering Enterprises in the Context of "Internet plus" [J]. Mall Modernization, 2022 (15): 24-26.

[4] Xu Weiwei The impact of the COVID-19 on catering enterprises and its countermeasures [D]. Beijing Foreign Studies University, 2021.

[5] Qin Jingqiao Fine grained emotional analysis of multi-channel restaurant reviews based on deep learning [D]. Lanzhou University of Finance and Economics, 2023.

[6] Wu Mei, Deng Song, Yu Jianjun. Research on the Factors Influencing Lifelong Learning: Analysis Based on LDA Theme Model [J]. Journal of Higher Continuing Education, 2023,36 (03): 1-9+21.

[7] Wang Siyi, Wu Xinyi. A Study and Analysis of the Emotions of Young Audiences in Revolutionary History Theme Exhibitions Based on the LDA Theme Model: A Case Study of the Shanghai Sihang Warehouse Anti Japanese War Memorial Hall [J]. Southeast Culture, 2023 (03): 162-171.

[8] Yang Jiawen, Shi Yuanyuan, Yan An. Text Mining and Emotional Tendency Analysis Based on Online Reviews: A Case Study of Museums in Beijing [J]. Internet Weekly, 2023 (11): 20-23.

[9] Yang Lianzheng, Zhai Tianzhi. Research on Emotional Tendency Analysis of Video Commentary Based on Emotional Dictionaries [J]. Network Security Technology and Applications, 2022 (03): 53-56.

[10] Luo Junjie, Lei Zexin, Hu Yike, et al. Analysis of Emotional Tendency of Urban Park Tourists Based on Deep Learning: Taking Tianjin Water Park as an Example [J]. Chinese Landscape Architecture, 2021,37 (09): 65-70.