Research on emotion analysis of Ctrip hotel reviews and user demand identification based on big data

Jianan Yu, Sinuo Han, Jia Ni, Jianzhen Wang, Jingyang An, Shengzhen Jia*

Tarim university, Xinjiang, China

ABSTRACT

With the rise of e-commerce economy, more and more people choose to book hotel travel online. As Ctrip is in a leading position in the e-commerce travel platform industry, this paper takes Ctrip hotel review as the representative for research. First, this article selects the top 40 hotels in the Beijing recommendation page to collect user reviews and related comments. A total of 34,988 comments were obtained after data consolidation. Then, the text data is cleaned, word segmentation, stop word and other preprocessing work. Finally, the processed data were classified. The emotion classification adopted in this paper mainly includes the emotion dictionary and the machine learning algorithms. The classification of the emotion dictionary is used to define the general classification of the text data and to visualize the display, then the word vector is constructed and the data set is divided using the TF-IDF algorithm[7]. According to the corresponding classification algorithm to train and test the training set and test set respectively, the classifier with the best emotion classification is the logistic regression model. It can be seen from the classification results of emotional polarity that the overall emotional tendency of users to Beijing hotel experience is positive. It is best to visually analyze the conclusions and results obtained, which not only has certain reference value for the hotel merchants and other subjects, but also provides theoretical reference for users.

In addition, this paper proposes a user demand identification and evolution analysis model based on online comment mining based on the data of Beijing hotel reviews on Ctrip APP. Using the Kano model and the LDA model, the review underwent classification, identification, feature emotion pair analysis as well as time series analysis. The results show that according to the emotional trend prediction, the emotional value of type 1, type 2 and type 3 of users showed an upward trend, while the emotional value of type 4 showed a downward trend. Users mainly focus on hotel services and environmental experience. The research has improved the time dimension analysis method and model of online reviews, and provided a reference value for analyzing user needs and predicting the emotional trend of hotel selection.

Keywords: Ctrip Hotel Review; Sentiment analysis; Kano model; LDA model; User requirements identification

1 INTRODUCTION

The number of users of the Internet has been very large. Most people have the habit of using the Internet to express personal ideas and obtain useful information, such as articles, videos, pictures and comments, etc. These information have hidden emotional tendencies, which has high research value. Therefore, it is of high reference significance to obtain users 'comments published on the Internet, to dig out the hidden emotions of users in the text, and to analyze the factors affecting the Internet businesses, media and government departments to affect the ups and downs of users' emotions. From the enterprise's point of view, obtain the Internet application platform consumer comment data compared to visit the type of market research is more efficient and high quality, but too complex information still brings some interference and inconvenience, in the field of hotel, the hotel surge of comments with the passage of time, screening effective information becomes tedious and time-consuming[1].At the same time, the diverse expression of opinions is highly challenging for text research. As a kind of subjective text, the comment text is based on the individual emotion and intention. In the text description, the use of words has a strong randomness, and even in the use of sentence patterns will appear relatively irregular and colloquial[2].

In text analysis, accurately locating key information is crucial. Subsubjective and objective emotion analysis of comment text is a fundamental and crucial step in understanding user interests. By studying the emotional tendency of users and applying it to the prediction of related fields, it can effectively assist decision and judgment, so it has attracted much attention in recent years. As the core link of emotion analysis, emotion classification is of great significance[1]. Emotion classification is a kind of tendency classification, and the research object is the subjective tendency expressed by the author. The result of the classification is generally positive emotion or derogatory emotion, or even complex and multi-level emotion. At present, there are many classification methods, and this paper will study the two methods: emotion dictionary and machine learning classification algorithm.

2 LITERATURE REVIEW

2.1 Research contents

This paper is based on ctrip —— domestic well-known online travel platform research, first use the web crawler from the network platform analysis of the Beijing recommended page before 40 hotel review data set, then the data pretreatment, including data cleaning, symbol characters, cut words, to stop words, speech, etc. Then, the emotion classification work is carried out based on the emotion Dictionary of CNKI, and the positive, negative emotion word cloud and correlation heat map are displayed.

On the other hand, TF-IDF feature extraction of text data is converted into a language that the machine understands, and three machine learning classification algorithms are constructed, namely naive Bayes, decision tree and logical regression model, so as to realize the classification and prediction of comment emotions.

2.2 Desearch thought

The research flow of this paper is shown in Figure Figure 1:

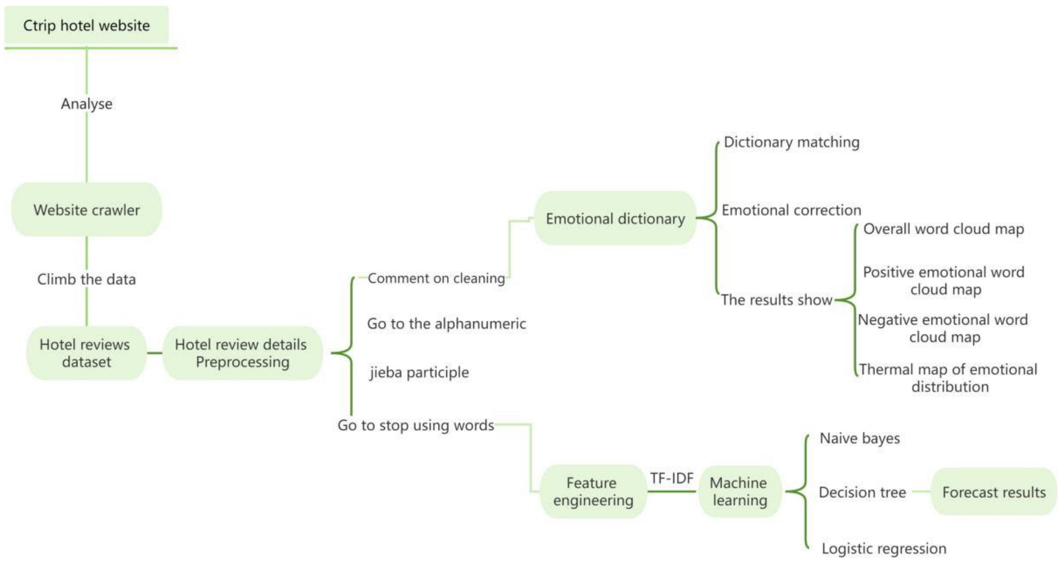

Figure 1: Flow chart of the study

2.3 Data sources

According to the similar data collection method of the data mining course, this paper obtained the data set required from the official website of Ctrip, which is the review data set of 40 hotels in Beijing (including 31 in total. The csv file), the total data of the hotel includes hotel name, hotel name, hotel name, total score, environment score, health score, service score, facility score, total number of comments, and the comment details of each hotel include "hotel name, real name, user name, personal score, comment date, and comment details". These hotels are located in different areas of Beijing, including Chaoyang Park, Wanfeng Road, Deshengmen Gulou Street, Guomao SKP, Yizhuang Economic Development Zone and other places. The 40 hotels range from 3.9 to 5.0, most with total scores above 4.5. For environment, most hotels scored 4.6 to 4.8 for health, 4.2 for health, 4.4 to 4.8 for service, and 4.1 to 4.7 for facilities. The total number of reviews ranged from the least of 3 to the most of 129. In general, these hotels perform well in all aspects, and customers have a high evaluation of their environment and service. Therefore, in the subsequent analysis, it is good that the score above 4.0 and below 4.0.

Due to the small number of total ratings for hotels, the 34,988 detailed reviews for the 40 hotels are described below, and the total number of reviews for each hotel is shown in Table 1 below:

Table 1: Total number of user reviews per hotel

Hotel name	Hotel name
Hilton North Tongzhou, Beijing	2910
Jingxi Hotel (Beijing International Trade SKP Store)	2480
Le Hejia Service Hotel (Beijing Xizhi Store)	2290
Floating HOME Chain Hotel (Beijing Guomao SKP Store)	1648
Zhe Hotel (Tiantan Park Store, Beijing Railway Station)	1585
Bao Hotel (Beijing Wangfujing Union Hospital Store)	1540
Qiuguo Hotel Jingxing (Wangfujing Store of Peking Union Medical College Hospital)	1482
Beijing Wangfujing Donghua Hotel	1327
Yishang Hotel (Beijing Global Liyuan Subway Station Store)	1310
YunLi Hotel (Beijing Chaoyang Park)	1290
THE HUMBLE Houju Hotel (Bird's Nest Store, Beijing National Convention Center)	1226
The Yongguang Hotel in Wangfujing	1111
Beijing Zimei Hotel	1074
Beijing Rock Hotel (Deshengmen Gulou Street Subway Station Store)	1070
Beijing 18 HOW Front Door Grape Yard	1037
Greenhao Tai (T3 Xingang Store, Beijing Capital Airport)	1031
Speed 8 Hotel (Beijing Fangzhuang Subway Station Cancer Hospital Store)	1018
Zhe Hotel (Yongding Gate Subway Station, Beijing South Railway Station)	1004
Ba Hotel (Lize Business District, Beijing West Railway Station)	962
7 Days Hotel Chain (Beijing)	880
Donghai Conde Si Hotel, Beijing Capital Airport	759
Lianjie Hotel (Changying Sky Street Store, North Chaoyang Road, Beijing)	672
Bona Hotel (Beijing West Railway Station)	654
Speed 8 Select Hotel (Beijing Wanfeng Road store)	650
White Deer House Hotel (New Exhibition, Beijing Capital Airport)	643
Guan Yunxuan Hotel (Beijing West Railway Station Financial Street Store)	501
Beijing Baita Light Hotel	470
Beijing Shimao Garden Hotel	373
Through the Modern Hotel (Beijing National Exhibition Store)	338
Yujing South Gong Hotel (Houhai Gulou Street Store)	248
Saihan Hotel (302 Hospital Store, Beijing)	247
Baoli Xinxing Hotel (Qianmen Dashilan Store, Tian'anmen Square, Beijing)	211
Furun Jingyun Hotel (Beijing Qianmen Dashilan Store)	195
Yunchuan Hotel (Beijing International Trade Store)	191
Home Inn (Beijing Yonghegong Heping Li East Street Forestry Bureau Store)	188
Beijing Chaoyang Gate Industrial Sports Manshe Hotel	115
Beijing Hejia B & B (Liangxiang University Chengbei Subway Station Store)	88
Speed 8 Hotel (Beijing Xueyuan Road North Beach Subway Station Store)	81
Carnival Hotel (Beijing Yizhuang Economic Development Zone Store)	59
No.10 Reserve Apartment (Chengshousi Subway Station, Beijing)	30
amount to	34988

The visualization results of the above user total comment data are shown in Figure 2 below:

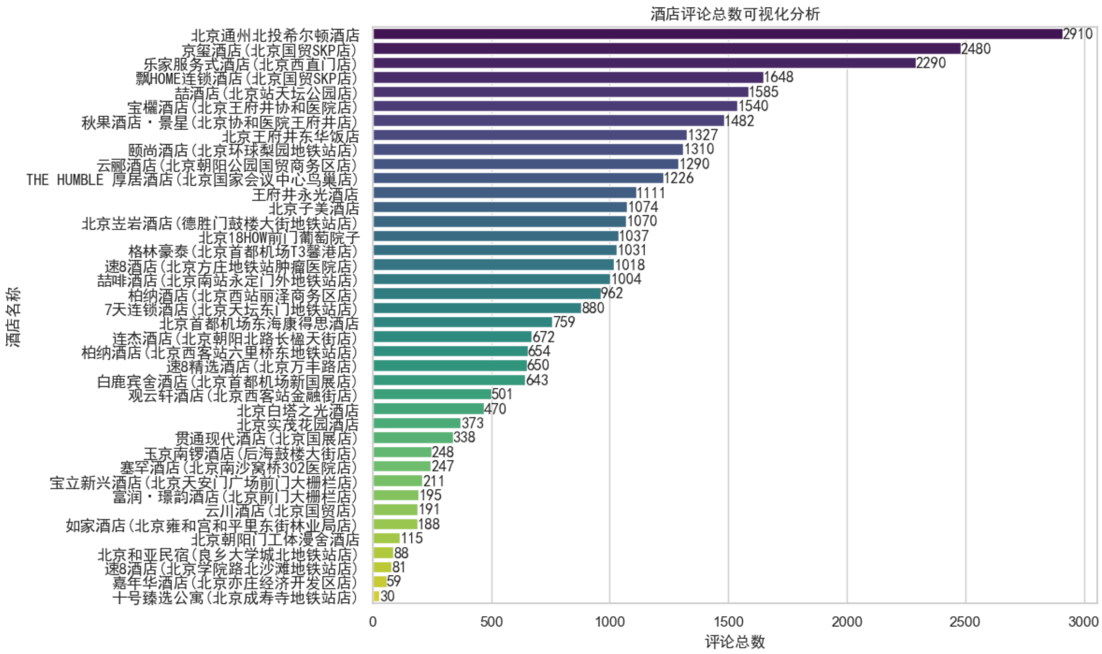

Figure 2: Total number of reviews per bar bar

From the above bar chart, the Hilton Hotel in Tongzhou, Beijing has the most reviews of 2910, and the 10 Reserve Apartment (Chengshusi Subway Station in Beijing) has the least reviews of 30. The total number of comments in most hotels is more than 1000, so the results of the analysis of the data obtained in this paper are highly supportive and reliable.

3 DATA PREPROCESSING

Data pre-processing is divided into three parts, namely, the data cleaning of missing values and repeated values of the combined data, and the visualization of the total comments.

3.1 Data cleaning

First, we combined the collected data set (40 hotel review details) together and saved them in the new one. Under the csv file, during the merger process to facilitate subsequent analysis, the hotel name (number) in the first column was changed to the number starting from 1, and the combined data set contained a total of 34,988 comments. Secondly, delete the same comment sent by the same user; 5 user names are repeated, but the user name is not the key information in this paper, so it is not processed temporarily.

After observing the data, the comments contained a large number of numbers, letters, and other characters. However, these data are not of substantial aid for the current mining goals. Therefore, we plan to remove the numbers and letters when handling the comment details. In addition, the review text mainly focuses on the quality of the hotel, which includes words such as "Beijing", "hotel", "room", "Ctrip" and so on. However, these words did not have a substantial impact on our analysis objectives. Therefore, before partitioning, we intend to remove these common words from the text.

3.2 Comment participle

The jieba library is a popular Chinese word segmentation tool for Chinese text processing. The posseg module provides the part-of-speech annotation of each word after word segmentation. Because it is difficult for the computer to analyze a sentence, but by breaking the sentence into words, you can analyze and compare the words, so the posseg module in the jieba library is used to make Chinese word segmentation and part of speech annotation, such as through the emotion represented by the words themselves, the judgment is good or bad. By matching the stop word dictionary and the mood word dictionary, remove some meaningless conformance and words, such as punctuation marks, stop words, mood words and so on. Words, parts of speech, and corresponding scores are packaged into one line for storage as special comments (storage-comment content). The first three lines are shown in Table 2 below:

Table 2: Storage-Comment on the first 3 lines

index_content	word	nature	content_type	index_word
0	installation	n	5	0
0	complete	nr	5	1
0	no	d	5	2
0	transport	v	5	3
0	Ado	n	5	4
0	pillow	n	5	5
0	Pillow	v	5	6
0	super	b	5	7
0	comfortable	a	5	8
0	environment	n	5	9

Finally, the processed comments will copy copies of the corresponding index to the overall data (storage-overall results).

3.3 Word cloud map

The word cloud map highlights the visual effect through the high-frequency vocabulary in the details of the statistical comments. Make word frequency statistics on the text, select the top 200 high-frequency words, and use the WordCloud module to make word clouds. The effect of word segmentation is shown in Figure 3 below.

Figure 3: Overall evaluation of the word cloud map

From a macro point of view, the following four types of keywords can be clearly identified from the above:

Hotel service: service, reception desk, check-in, enthusiasm

Hotel facilities: environment, facilities, clean, sanitary

Around the hotel: location, transportation, subway station, travel

Other keywords: breakfast, free, comfortable, comfortable, satisfied, neat, experience, next time, cost-effective

From the above analysis, we can see that customers pay more attention to these aspects. It can also be seen that "good, good, good, also, special" and other positive comments, indicating that the quality of most hotels in this region is in line with guests' expectations.

4 EMOTIONAL CLASSIFICATION METHOD

The processed data is classified by using the emotion dictionary and machine learning. First of all, the emotion dictionary classification is used to define the general classification of text data, and then divides the data set 2:8 in the form of feature value extraction[8].Then, the training set and the test set are trained and tested respectively according to the corresponding classification algorithm, to realize the emotion classification and prediction of the comments, and to analyze the experimental results.

4.1 Emotional dictionary

4. 1. 1 **Emotional dictionary introduction**

The content of user evaluation often reflects the attitude of emotional expression and criticism. The quality of emotional dictionaries has a direct impact on the effect of emotional classification. However, CNKI provides 12 documents, which are divided into English and

Chinese emotional dictionaries. Among them, the Chinese emotional dictionary covers the emotional text of evaluation, emotion, proposition, degree (positive and negative). Four of these documents were selected in this paper, including "positive words", "positive emotional words", "negative evaluation words" and "negative emotional words". According to the characteristics of the hotel corpus, "clean" and "cheap" were manually added to the positive evaluation words, and "dirty" and "noisy" were added to the negative evaluation words[3].

4. 1. 2 **Emotional classification design**

When performing emotion analysis, it is a common method to conduct dictionary matching based on the emotional dictionary. First, we need to import positive and negative comments. To simplify sentiment analysis, we can set an initial weight of 1 for positive words and-1 for negative words. In this way, we created a dictionary that contains emotional words and their weights. During the word segmentation process, the results and the positive and negative emotions are combined to facilitate the location of each emotion word and its corresponding weights. During the segmentation process, we combined the segmentation result with the positive and negative emotion word list in order to locate each emotional word and its corresponding weight.

However, in Chinese, there is double negation, which may change the semantics of the sentence. To more accurately capture the emotional information in the text, we need to consider the influence of negative words. Therefore, we introduced negative word lists to correct the direction of emotion values[5].Specifically, if there is a negative word before the emotional word, we will adjust the weight of the emotional word to the opposite value. It should be noted that when two or more negative words exist, we do not make weight adjustment to avoid overcorrection.

After completing the emotion word matching and weight adjustment, we can calculate the emotion score for each comment. By summing the weights of each emotion word, we can get the overall emotion score for one comment. Based on the positive and negative nature of the score, we can classify comments into positive or negative comments. This emotion analysis method can not only help to understand the emotional tendencies of users, but also can be used to automatically process a large amount of text data in business scenarios, extract the emotional information contained in it, and provide reference for further decisions.

4. 1. 3 **Display of classification results**

4. 1. 3.（1）Related thermal maps

First, the classification results are counted: the number of positive comments is 11679; the number of negative comments is 2450, see Figure 4.

From the statistical results, we can see that even if the score is around 1.0, there are still positive comments, and even if the score is around 5.0, there are still negative comments. Among them, for extreme cases, such as positive comments of 1.0, outlier treatment. But on the whole, it reflects the score and text emotional tendency, but it is not completely linked.

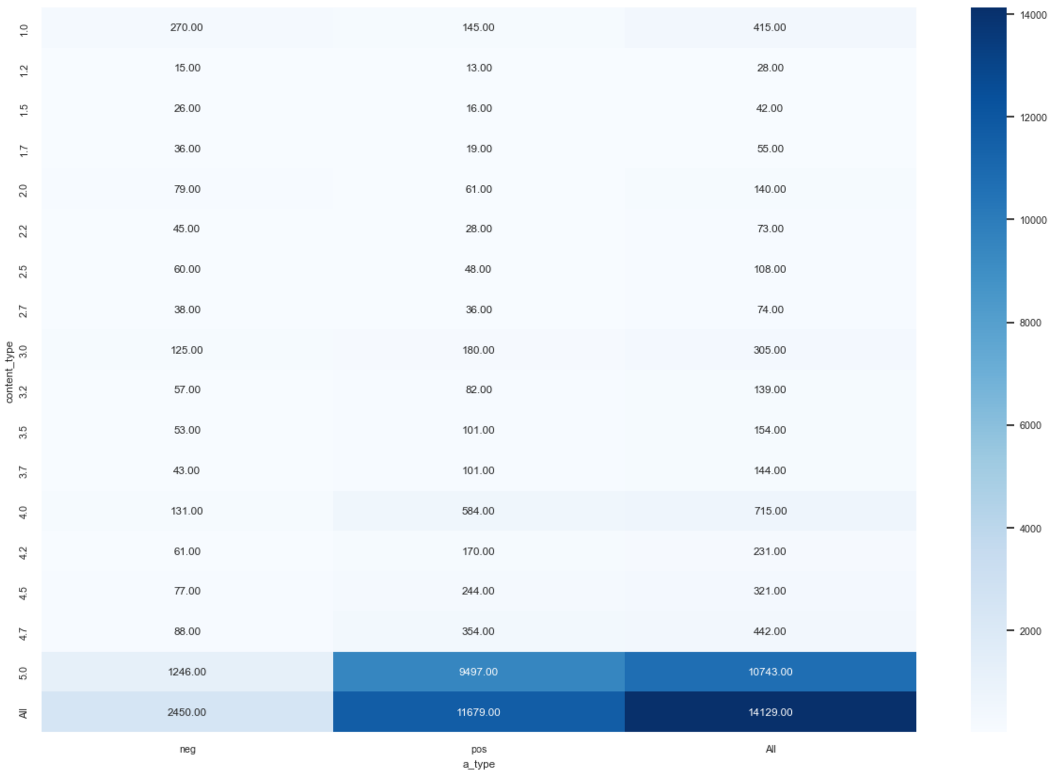

Figure 4: Heat map of the correlation between actual ratings and textual emotional tendencies

Combined with the practical analysis, the possible reasons are as follows:

① The emotional dictionary is flawed. Network texts are diversified and colloquial, and many new words, such as "bizarre", may not be included in the network emotional dictionary. Meanwhile, hotel comments have hotel context, such as "imported facilities" are positive, but not reflected in this general emotional dictionary, so there is still room for adjustment in the identification process.

② The guest's comment text cannot express the emotion completely. For example, although the guest gave a high score of more than 4 points, he emphasized the unsatisfactory aspect in the comments, without emphasizing the overall satisfaction degree, so the two are easy to contradict each other[6].

③ The score is inflated. Some users may not take the score particularly seriously, and if they do not select the score, the system will default to five-star reviews[6].

4. 1. 3.（2）Classification word cloud display

Figure 5: Positive emotional word cloud map

Figure 6: Negative emotional word cloud map

It can be seen from Figure 5 and Figure 6 positive and negative emotional comments that the comments of the two emotions are well separated, so it can be concluded that the emotion classification based on the emotion dictionary can better extract the emotional comments out.

From the analysis results, it can be seen that the hotels in this area have a high evaluation in terms of hotel services such as breakfast and clean environment, but there is still room for improvement in the service attitude of the hotel front desk, health facilities and some service personnel. Hotel management can strengthen the attention in this aspect, and can also be used as the competitiveness with other hotels.

4.2 Machine learning emotion classification

The emotion analysis methods of machine learning fall into two categories: supervised learning and unsupervised learning. Supervised learning can be subdivided into four types: probabilistic classification, support vector machine, decision tree, and deep learning. Specifically, the probabilistic classification includes naive Bayes and Bayesian networks, while deep learning includes convolutional neural networks and recurrent neural networks. These methods view sentiment analysis as a classification problem and share similar fundamental processes. First, the machine learning algorithm is trained using the training samples to build the emotion classification model. Then, the model performance was evaluated by testing the samples. Eligible models can be used for future emotion classification. Finally, the text to be classified is entered, and the model outputs the emotion classification results.

4. 2. 1 **Model select**

In this paper, machine learning methods such as naive Bayes, decision tree and logistic regression are selected for model training to classify and predict the emotional tendency of comments.

4. 2. 1.（1）Naive bayes

The naive Bayes algorithm is a classification method based on the Bayes theorem and feature condition independent assumptions, which is often used for text classification, spam filtering, emotion analysis, and multi-category classification. The core idea is to classify them by calculating the conditional independence between features for a given category. This algorithm performs well on a variety of tasks and is particularly suitable for processing large-scale text data. For sentiment analysis, text data is often represented as a bag of word model where each word is a feature.

The steps of the naive Bayes algorithm are as follows:

① Data preprocessing: convert text data into a bag of words model, calculate the frequency of each word in the document or using other text representation methods.

② Calculate the prior probability: calculate the prior probability for each emotion category.

③ Calculate conditional probability: conditional probability for each word under a given emotion category.

④ Prediction: For the new text, the posterior probability of each emotion category is calculated based on Bayes' theorem, and the category with the maximum probability is selected as the prediction result.

The naive Bayes classifier is based on the Bayes theorem, using the assumption of conditional independence between features (naive assumption). For sentiment analysis, text data is often represented as a bag of word model where each word is a feature.

4. 2. 1.（2）Decision tree model

Decision tree model is a decision model based on tree structure that classifies or predicts data through a series of rules (such as if-then) and conditions. In sentiment analysis, features can be lexical, phrases, or other textual features. The decision tree learning steps are feature selection, decision tree generation, and pruning of the decision tree.

The decision tree algorithm steps are as follows:

① Select the best feature: Select the features that best distinguish the different categories based on an evaluation index (such as information gain or Gini coefficient).

② Division data: divide the data according to the selected features to generate child nodes.

③ Recurrence: Repeat the above procedure for each child node until the stop condition is reached (if the maximum depth is reached or the number of samples contained in the node is less than the threshold).

④ Prediction: New text samples are predicted based on the structure of the decision tree.

4. 2. 1. (3) Logic regression model

Logistic regression is a model for binary classification that maps linear combinations of features to probability values by applying the logistic function between 0 and 1. In sentiment analysis, features can be word frequency in the pouch model or other textual features. It is essentially a nonlinear model, assuming that the function is usually a Logistic function (the Sigmoid function): $h_\theta\left(x\right)=\frac{1}{1+e^{-(\theta_0+\theta_1x_1+\theta_2x_2+\cdots+\theta_nx_n)}}$, where $x_1,x_2,\cdots,x_n$ is the feature and $\theta_1,\theta_2,\cdots,\theta_n$ is the model parameter.

The steps of the logistic regression algorithm are as follows:

① Linear combination: Calculate linear combinations of input features, through weight and bias.

② Applying the logistic function: By applying the result of the linear combination to the logistic function, you can map the output between 0 and 1, indicating the probability.

③ Loss function: Define a loss function, usually using a logarithmic loss function (log loss).

④ Gradient descent: minimize the loss function and adjust the weight and bias through the optimization algorithm such as gradient descent.

⑤ Prediction: For new text samples, the probability is calculated according to the learned weight and bias, and the probability is converted into categories according to the threshold.

4. 2. 2 **Model prediction and evaluation**

First, the independent and dependent variables were identified and treated. Comment text was used as independent variables, and text information was converted into a vector by feature extraction using the TF-TDF algorithm. TF indicates the frequency of keywords, and TDF represents the frequency of the inverse document, which is mainly used to indicate the importance of each keyword in the document. The score (1.0-5.0) was used as the dependent variable.

However, due to the imbalance of data sets, that is, more favorable (4.0-5.0) and less bad reviews (below 4.0), the multi-classification effect is very poor, and the test score of the model is basically 0. Therefore, the model tuning turns the multi-classification problem into a two-classification problem.

Table 3: Parameter representation of the dependent variable 'score'

Score	Values
score <=4.0	0
Score>4.0	1

Then divide the data set. Divided the data set into training set and test set, where the training set contains 80% comments and the test set contains 20% comments, and processed the data to normalization.

Start the model training. This paper uses three types of models for experiments, all machine learning models based on extracted text features, including naive Bayes classification, decision tree, and logistic regression.

Finally, the model evaluation was performed. In this paper, the accuracy (precision precision), recall (recall recall) and F1 value (F1-score) are commonly used to evaluate the prediction results[4].

4. 2. 3 **Model prediction results**

According to the text content of hotel reviews, the prediction effect of favorable reviews and bad reviews is shown in Table 4. On the whole, these three machine learning models all have good classification effect and prediction effect. Among them, the logistic regression model performs the best, and its Precision, Recall and F1-score are all 0.833. Therefore, therefore, the classifier with the best emotion classification should choose the logistic regression classifier.

Table 4: Emotional dichotomy prediction results

Classification models / metrics	Precision	Recall	F1-score
naive bayes	0.826	0.826	0.826
decision tree	0.804	0.804	0.804
logistic regression	0.833	0.833	0.833

5 MODEL OF USER NEEDS IDENTIFICATION AND EVOLUTION ANALYSIS

5.1 Establishment and solving of the Kano model

The Kano model was used to classify the requirements of online reviews. First, comments were divided into both positive and negative categories by sentiment analysis. Those with an emotional value less than 0.5 were converted to 0 (negative), and those greater than 0.5 to 1 (positive). Secondly, the attention of comments is analyzed, and the specific value is obtained by adding the number of likes and replies. The lower limit, the median and the upper limit of attention are determined by using the boxplot, and divided into two levels: low (0) and high (1).

The association of the Kano model and online reviews is shown in Table 5:

Table 5: Relation of the Kano model to online reviews

Kano model requirement type	Kano model requirement type	Kano model requirement type
Changes in user satisfaction	Changes in user satisfaction	Changes in user satisfaction

Characterization of the online reviews	Characterization of the online reviews	Characterization of the online reviews
Basic type requirements	Basic type requirements	Basic type requirements
When provided, the change of user satisfaction is not significant;	When provided, the change of user satisfaction is not significant;	When provided, the change of user satisfaction is not significant;

Emotional polarity reflects the user's attitude to the hotel experience, positive polarity satisfied, negative polarity dissatisfied. Emotional polarity and attention are demand classification indicators, which is measured by the sum of the number of likes and responses. Demand can be divided into four combinations based on polarity and attention.

① Basic needs: negative emotional polarity, high attention, need to pay attention to. Users are not satisfied with the hotel experience, but the attention is high, and there may be potential problems to be solved.

② Expectation needs: positive emotional polarity, high attention. Users' expectations can be met, and they can have strong positive emotions for the hotel experience. Negative emotional polarity, low attention, unmet needs, and low emotional degree of users.

③ Charm needs: low attention, and users have strong positive emotions after being satisfied. Users pay less attention, but generate positive feedback after demand are met.

In addition, after the early processing of the comment data, the text has been stressed, compressed, removed the stopped words and short sentences, without considering irrelevant comments. Requirements classifications based on online reviews are shown in Table 6 below:

Table 6: Demand classification based on online reviews

Emotional polarity	Emotional polarity	Emotional polarity
attention rate	attention rate	attention rate
Demand type	Demand type	Demand type
negative direction	negative direction	negative direction
tall	tall	tall

By classifying the review data, and then using topic models and dependent syntactic analysis, different types of user needs are deeply analyzed. By identifying the specific hotel characteristics that users follow, targeted hotel arrangement improvement strategies can be formulated.

According to the emotional polarity and attention level of comments, each data set can be divided into different demand types, including basic, expected and attractive needs.

5.2 The LDA model building and solution

For the LDA subject model analysis, you can simplify to the following steps:

① Conform the comment text into a bag of words model (BoW model) representation.

② The optimal number of motifs was determined using the cosine similarity.

③ Model training was performed using the LDA module of the Gensim tool, adjusting parameters, and output the subject word distribution.

When applying the LDA model, we introduce the cosine distance method to obtain the best topic analysis results by increasing the number of topics. Using cosine similarity to determine the best number of topics, in P7000P model 1 data, for example, emotional polarity for the positive expectations demand data optimal theme number is 4, emotional polarity for the negative expectations demand data optimal theme number is 3, basic demand data the best theme number is 3, charm demand data best theme number is 3. The relationship between cosine similarity and subject number is shown in Figure 7 below:

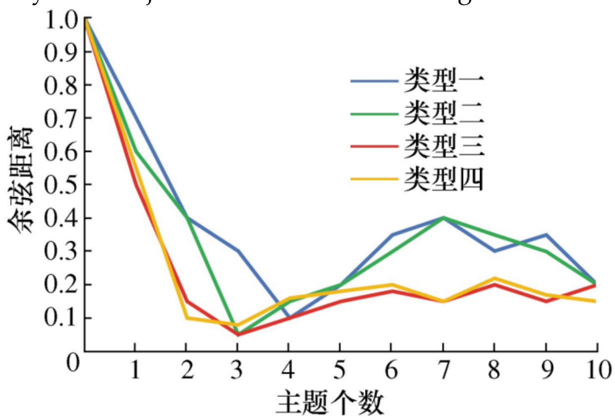

Figure 7: Best number of LAD models

5.3 Emotional value time series prediction

The time series for studying the affective values are usually the stationary ones. To make prediction, simple averaging, moving average, exponential smoothing, and autoregressive models can be considered. Simple and moving averaging rely mainly on past averages, and therefore it may be difficult to accurately predict future fluctuations. In contrast, the exponential smoothing method is more advantageous in the short-term trend prediction. It combines the characteristics of the whole period average and the moving average, and gradually weakens the influence of the past data.

The exponential smoothing method includes primary, secondary and tertiary exponential smoothing methods, which can be selected according to the actual situation. For the emotional time series of hotel reviews in type 1, we applied these methods and autoregressor processing separately, and the resulting image results are shown in Figure 8. In the figure, the red line represents the raw data, while the blue line indicates the trend after stabilization.

The choice of these methods depends on the specific context and data characteristics of the study, while the graphical results can help to observe the treatment effect and provide a basis for further analysis.

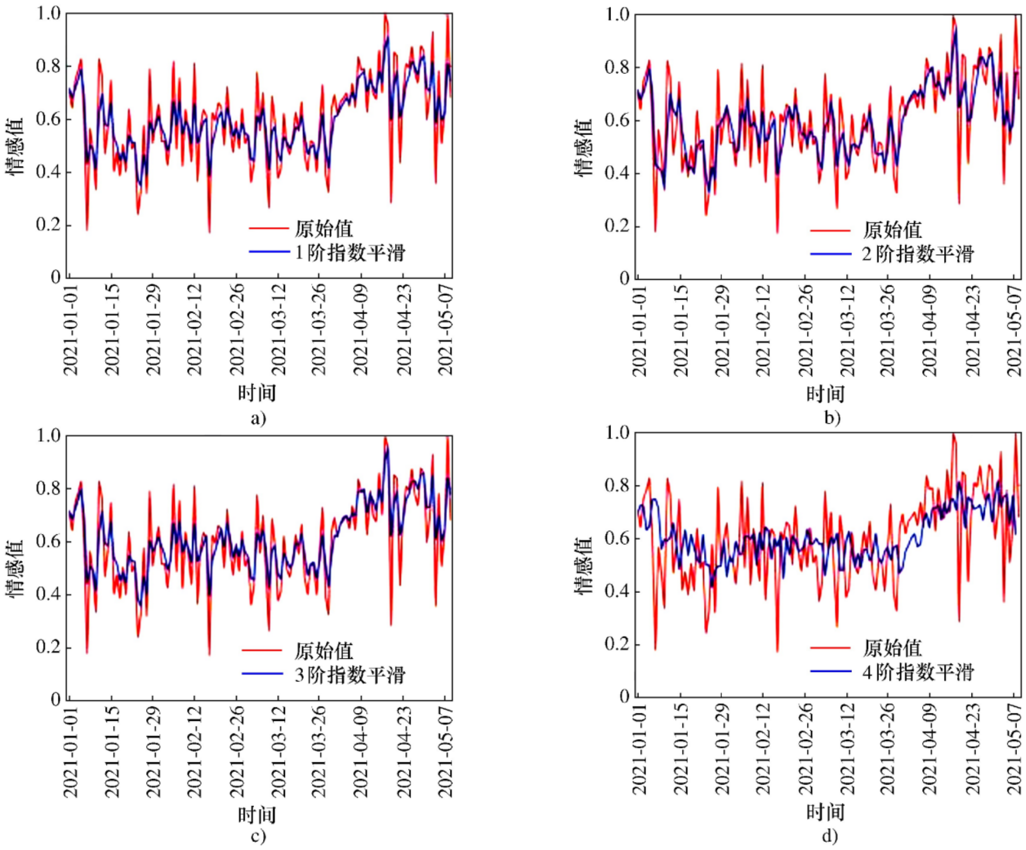

Figure 8: Time series of the 4 methods fitted to Fig

The mean variance of the fitted and actual data from the four models are shown in Table 7 below:

Table 7: Mean variance of the fitted data and the actual data

An index smooth	An index smooth	An index smooth	An index smooth
MSE 0. 011 6	0. 008 0	0. 007 8	0. 024 0

According to the comparison results of the prediction method, it is found that the tertiary exponential smoothing method performs the best in the historical data fitting, which can better reflect the changing trend of the data. Next, we will use tertiary exponential smoothing to predict the affective value time series of 4 hotel reviews. The results are shown in Figure 9 below:

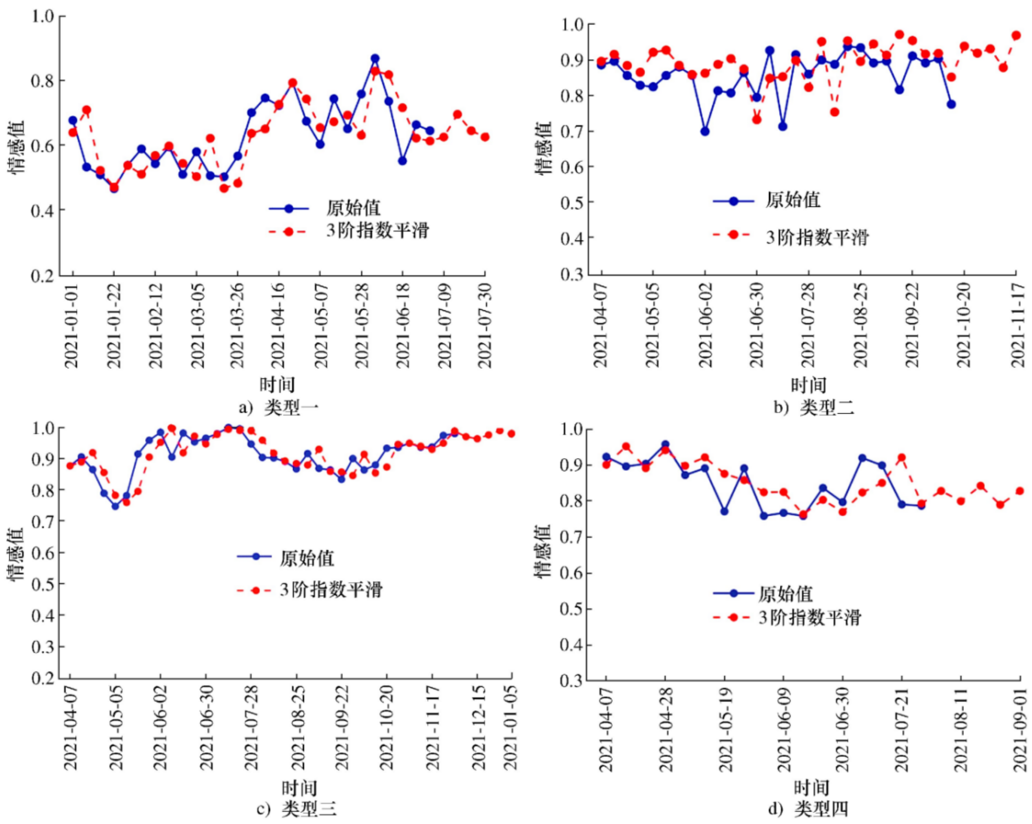

Figure 9: Emotional mean distribution over time and exponentially smoothing the prediction results

With the exponential smoothing prediction results in Figure 9, we observed different types of emotional trends. The emotional values of Type 1, Type 2 and Type 3 are rising, indicating that these hotels are increasingly popular among customers. Concontrast, the emotional value of type 4 showed a downward trend. Overall, the emotional average of Type 4 has remained around 0.8, indicating that shoppers still have a positive attitude towards the hotel. Although the emotional mean of type 1 has increased, it is constantly approaching 0.8, which indicates that type 4 still has a strong competitive advantage over type 2. However, the emotion value of type 2 and type 3 keeps approaching 1, especially that of type 3, and the emotion value fluctuates less. Therefore, from the perspective of emotion value, the competitive advantages of the four types of hotels are arranged as types, three,> types, two> types, four> types and one. This conclusion is consistent with the previous descriptive statistical analysis.

6 MODEL EVALUATION

6.1 Effect summary analysis

Based on the platform of Ctrip, this paper studies the emotional tendencies in hotel reviews. First, this paper designs the corresponding web crawler according to the characteristics of the website to obtain the data required for the experiment; second, this paper cleans the climbing data and pretreatment, and visually displays the data with word cloud and heat map; then this paper designs the experiment, realizes the emotion classification, displays the classification and prediction results, uses the word cloud and heat map etc., and analyzes the experimental results, obtaining some new conclusions, with certain reference value to the hotel merchants. Overall, both have a good results. From the emotion dictionary classification method, we get three conclusions:

① On the whole, the guests are satisfied with the hotel in the area, mostly positive comments;

② There is a certain correlation between the actual score and the emotional tendency, but not completely linkage, and makes a possible explanation for this;

③ The keywords and aspects are extracted from the positive and negative emotional word cloud map and the overall word cloud map, which is helpful to the management of the hotel.

From the machine learning classification method, we know that the two classification effect of the three models is better, in which the logical regression is optimal.

At the same time, there are some deficiencies in the research process, which needs further research and improvement. Since the details of the hotel comments are Chinese network text, the Chinese language itself has rich meanings and various emotions, and the network language is updated rapidly, so the emojis contained in the text also have rich connotation. In order to ensure that the comments can be fully utilized, some network terms should be discussed more deeply.

This paper presents a user demand identification model based on LDA topic analysis and an improved Kano model. Combined with dependent syntactic analysis and time series model, the changing trend of users' demand for hotel selection is studied, and the emotional value prediction is conducted to provide the basis for market hotel construction. Compared with the existing methods, the model can consider the time factor, deeply analyze the change of users' emotional value and the evolution of demand theme, and analyze the user needs dynamically, and provide guidance for business management strategy, hotel improvement and innovation. However, there are still some limitations. Emotion values were only divided into positive and negative aspects, and neutral emotion was not considered. Future studies can further analyze the impact of user-neutral sentiment on the development of e-commerce.

6.2 Future prospects

In the in-depth analysis, we can propose the following three prospects to provide more possibilities for future research and applications:

① The comprehensive expansion of the emotional dictionary and the network language attention

This paper mentions that the emotional dictionary can be expanded more comprehensively, with special attention to the network language that contains the users' feelings. In addition to the traditional text expression method, online communication often involves non-literal elements such as emojis, and these elements can also convey rich emotional information. Therefore, in future studies, collecting and analyzing various emoticons and labeling their corresponding emotions to achieve a more comprehensive emotional understanding is considered.

② In-depth study of multilevel emotion classification

This paper looks at dichotomizing texts as positive and negative, however, human emotions are varied and can further classify emotions in a more nuanced way. Future studies could explore multilevel emotion classification methods that subdivide the emotion into multiple categories to more precisely capture the emotion information in the text. Such research can facilitate a more comprehensive understanding of users' emotional experience, and provide more in-depth emotional analysis support for areas such as personalized recommendation and user research.

③ Regional subdivision and individual hotel sentiment analysis

This paper mainly focuses on the overall research of hotels in this region of Beijing, Xinjiang province. In the future work, we can consider the more detailed emotional analysis of hotels in different regions. At the same time, the emotional classification of individual hotels is also an important research direction. By deeply exploring the reviews of a single hotel, we can provide more targeted suggestions for improvement, and further improve the service quality and user experience.

REFERENCES

[1] Shi Xin. Based on the emotional dictionary in the hotel field [D]. Hebei University, 2014.

[2] Ding Wei. Emotion analysis based on a combination of dictionary and machine learning [D]. Xi'an University of Posts and Telecommunications, 2017.

[3] Ma Zirui. Recommendation algorithm for tourist attractions based on the emotional dictionary [J]. Computer programming skills and maintenance, 2022, (04): 20-21 + 37.

[4] The Hao Glacier. Research on text classification algorithm based on corpus features [D]. Yanshan University, 2019.

[5] Peng Mei, Hu Bibo. Emotion analysis of e-commerce user comments based on big data and artificial intelligence [J]. Computer Programming Skills and Maintenance, 2022, (06): 123-126.

[6] Xu Dehua, Yang Zhiling. Correlation study between emotional tendency and ratings in online product review text [J]. Library and Information Guide, 2022,7 (02): 59-65.

[7] Liu Haiou, Yao Sumei, He Xutao, etc. Deep learning-based online health community depression user portrait study [J]. Small Microcomputer system, 2021,42 (03): 572-577.

[8] Zhang Cui, Zhou Mengjie. A text emotion classification method based on the fusion of CNN and bidirectional LSTM [J]. Computer Age, 2019, (12): 38-41.

[9] Yang Deqing, Zhang Jing, Guo Wei, et al. Kanano Model Building based on the online product community [J]. Mechanical Design, 2018,35 (03): 12-19.

[10] Wang Roihuan. Analysis and improvement of hotel customer satisfaction based on text mining [D]. Dongbei University of Finance and Economics, 2021.