Wang *et al. Computational Urban Science* (2022) 2:10
https://doi.org/10.1007/s43762-022-00042-1

Computational Urban Science

ORIGINAL PAPER

Open Access

Emotional characteristics and theme mining of star-rated hotels from the perspective of social sensing: a case study of Nanchang City, China

Jingbo Wang[1], Yu Xia[1,2*] and Yuting Wu[1]

Abstract

Mining hotel social sensing data and analyzing its spatial and temporal characteristics can provide decision support for hotel managers. Present research on this topic is limited to the overall hotel industry and text mining. Here, we first obtain POI and reviews for star-rated hotels in Nanchang from 2018 to 2021. Secondly, the hotel POI (Point of Interest) is combined with the sentiment value of customer reviews. Finally, comparative analysis and topic mining of Spatio-temporal aspects of text reviews of different star-rated hotels are conducted using sentiment analysis, spatial analysis, and thematic social network analysis. Results show that: (1) Hotel star rating and hotel review sentiment value are significantly positively correlated. The seasonal trends of different star rating hotel sentiment values are similar, but are highest in summer and lower in autumn; (2) The highest sentiment value is seen for friends' outings and the lowest is for business trips; (3) Customer reviews of star-rated hotels focus on three aspects: facilities, service, and location. Three-star hotels focus on the stay experience, while four-star hotels focus on the breakfast situation. Exploring hotel social sensing data can intuitively illustrate hotel selection's behavioral patterns and spatial-temporal characteristics. The methods of this study can expand the application of social sensing data in different fields, such as the tourism and restaurant industries.

Keywords: Social sensing, Star hotel, Spatio-temporal features, Sentiment analysis, Topic mining, POI

1 Introduction

Customer evaluation information helps potential consumers make purchasing decisions and helps managers improve hotel services' quality. Customers generally cannot judge the quality of products and services provided by hotel companies until they experience them, while the opinions expressed online by other customers are a more independent and trustworthy source of information than hotel advertisements (Muhammad et al., 2021; Wang et al., 2020). An increasing number of customers browse review data before travel to arrange they are itinerary and choose hotels (Chang et al., 2017; Valdivia et al., 2017). Therefore, how to analyze and study customer evaluation information is a key concern for hotel managers.

Currently, social sensing data mining for the hotel industry is mainly focused on the exploration of review data. Sentiment analysis of online reviews can yield customer preferences for hotels (Zhang, Ma, et al., 2021). Stringam et al. explored what factors increase hotel consumer ratings by evaluating consumer ratings and online review data (Stringam & Gerdes, 2010). Yen et al. examine the impact of hotel attributes on consumer behavior in posting online reviews (Yen & Tang, 2018). Meanwhile, various research methods for online review of big data such as sentiment analysis (Jia & Chen, 2020;

*Correspondence: geoxy@126.com

[2] Key Laboratory of Poyang Lake Wetland and Watershed Research, Ministry of Education, Jiangxi Normal University, Nanchang 330022, China
Full list of author information is available at the end of the article

Tan et al., 2021a) and topic mining (Luo & Yi, 2019; Wang et al., 2014) have provided new ideas and tools for hotel management. Sentiment dictionary (Rao et al., 2014), machine learning (Tang et al., 2019), and sentiment analysis tools (Tao et al., 2019) are primary methods of conducting sentiment analysis. Word frequency statistics (Reagan et al., 2017), semantic analysis (Han, & Chiu,, & Cheng., 2020), topic probability models (Blei et al., 2001), and machine learning (Rusanov et al., 2018) can be used for topic mining. There are natural differences in the hardware facilities and service environment of different-star hotels. Thus, the service expectations of customers are also different (Li et al., 2020). However, existing studies are inadequate in terms of both their content and methodology, specifically: on the one hand, most of the existing results study hotel customer evaluations as a whole, lacking a comparison of the differences in the influencing factors of different star-rated hotels; on the other hand, they generally focus on the analysis and research of review texts and have not yet fully explored the hotel location information. In recent years, various online booking platforms have provided a wealth of valuable socially-aware data. Proper exploitation of Spatio-temporal big data of the hotel industry to provide a theoretical and practical basis for social sensing research is a primary focus of current research.

This study uses hotel POIs and customer reviews web-crawled under the Ctrip hotel platform as a database and then processes them to obtain hotel social sensing data. First, the BERT (Bidirectional Encoder Representation from Transformers) model is used to classify the reviews for sentiment, and then the hotel sentiment value is calculated by combining the hotel information to carry out data analysis in spatial and temporal dimensions. Second, kernel density analysis of the distribution characteristics and hotel sentiment values according to hotel POIs was conducted to study the spatial layout differences and sentiment distribution characteristics of different star hotels. Finally, theme mining of hotel review data by the LDA (Latent Dirichlet Allocation) theme model is combined with a social network analysis method to analyze the similarities and differences in customer attention given to different star hotels. Therefore, the research results can be used to provide theoretical guidance and practical reference for the effective management of different-star hotel managers.

2 Data sources and research methodology

2.1 Study area and data sources

Nanchang City is the administrative and cultural center of Jiangxi Province and has a large number of hotels, clearly-defined spatial distribution, and many customer reviews, which can reflect the current situation of the local hotel industry (Fig. 1). The Ctrip hotel platform is the domestic hotel booking industry benchmark and the online accommodation booking market (Wang et al., 2021). Its user base is large, hotel customers are active in their online reviews and the online information on the website reflects the basic profile of hotel users (Maomao et al., 2021). We use Python scripts for web crawling to collect online reviews and hotel POIs on the Ctrip hotel platform from September 2018 to September 2021 and select three-star and above-star hotel reviews in Nanchang. We focus on hotels with three or more stars since lower-rated hotels have few reviews. Since the number of reviews varies from hotel to hotel and some hotels have a very low number of reviews, hotels with less than 20 reviews were filtered in this study. In total, 362 (29 five-star, 79 four-star, 254 three-star) hotels and 168,613 customer text reviews were obtained (Table 1). Each piece of data includes hotel ID, hotel name, review text, hotel coordinates, review time, and travel type label (Table 2). Among them, travel type is a label classified by the Ctrip hotel platform according to customers. As the subject of this study is hotels in Nanchang, China, non-Chinese reviews were filtered to ensure the accuracy of the findings.

3 Research methods

3.1 BERT-based sentiment classification

The BERT model enhances its semantic representation with the masked language model and next sentence prediction tasks (Acheampong et al., 2021; Tan et al., 2021b). The core idea is the same as that of the Transformer model, which effectively removes distance constraints by combining the interconnections between words in the text to fully display the dependencies between the current word and the rest of the words in the sentence, thus better identifying the semantic information of the sentence. We use the BERT model to classify the sentiment of hotel review texts and use probability values [0, 1] (Zhang, Ma, et al., 2021) output from the model as sentiment scores. Where the closer the value to 1, the more positive the comment sentiment, and the closer to 0, the more negative the comment sentiment.

The encoding process of the model input is the summation of three vectors, the inputs of which are in the form of a vector representation of the corresponding word for each word in the text, the encoding of location information, and the marking of paragraph information (Fig. 2). At the same time, two special symbols, “CLS” and “SEP” are added. “CLS” is generally added to the first part of the text, where the feature can be extracted for use in the classification model. The “SEP” symbol

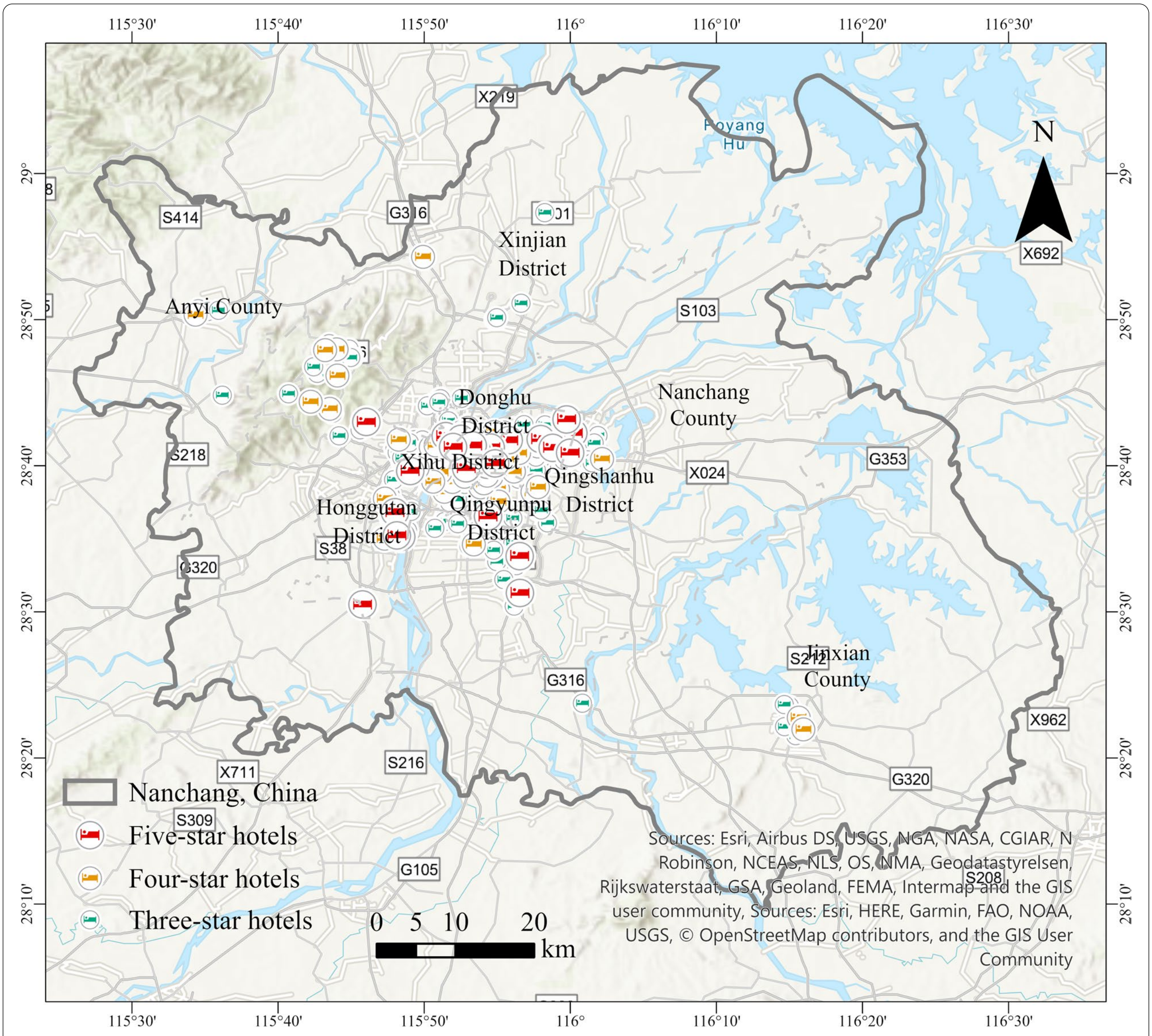


Fig. 1 Distribution of star hotels in Nanchang

Table 1 Information on the number of reviews for each star hotel

Hotel stars	Number of hotels	Total reviews	Maximum reviews	Minimum reviews	Average reviews
Three-star	254	81,848	1900	22	323
Four-star	79	51,084	3880	53	646
Five-star	29	35,681	3201	113	1230

indicates a clause break, which is used to break two sentences in the input text.

The BERT base model structure is built by a 12-layer Transformer architecture, with the encoding dimension set to 768 dimensions, while the multi-head is set to 12 and the model has 110 million parameters. Due to the small comment corpus and to avoid parameter overload, the open-source shared BERT sentiment classification

Table 2 Sample data example

Comment text	Hotel ID	Hotel coordinates	Date	Type of travel
The hotel facilities are good in all respects and the service is particularly welcoming and attentive.	811,119	115.817318, 28.660664	2021-08-14	Family-friendly
The manager on duty was very good and welcoming and the check-in was very skilled and professional.	716,162	116.003495, 28.703978	2021-07-21	Family-friendly
I think it's great! Superior, clean, and tidy! Gave me a very comfortable feeling!	1,498,523	115.979817, 28.687144	2020-03-23	Traveling alone
The waiter service was very good, and the hotel service was very good.	1,479,634	115.8579, 28.700409	2019-08-01	Business trips

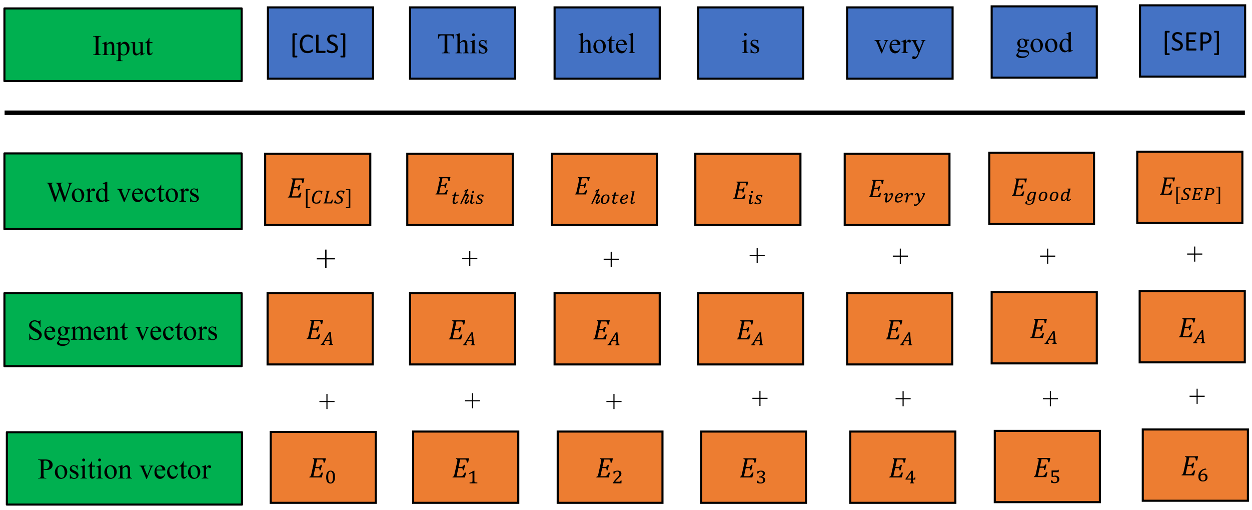


Fig. 2 Input sequence diagram

model from the GitHub web platform is used here.[1] The model was built by a 6-layer Transformer architecture, with the coding dimension set to 384 dimensions, the multi-head set to 12, and is trained using the Chinese Wikipedia corpus. By manually labeling and sorting the emotional tendency of hotel review data, 5000 positive and negative hotel reviews are obtained as the review corpus, which is divided into a training set (95%) and a test set (5%) (Yang et al., 2019). The pre-trained model was fine-tuned to train the final sentiment classification model.

The training process of the model and the main parameters are as follows. First, each epoch is trained with the training set and the model is evaluated using an AUC (area under the curve) method (Berrar, 2014). Second, the model uses a dynamic learning rate and early termination, with the initial learning rate set to 1e-6 and the Batch size set to 24. The process is as follows: after the current Epoch is trained, the current training result is measured using the test set and the AUC of the Epoch is noted, and if the current AUC does not improve over the previous Epoch, then the learning rate is reduced (by 20%) until the AUC of all 10 Epoch test sets did not improve, then training was stopped. In the end, the model was trained for 418 Epochs. To solve the overfitting problem, the model sets the dropout to 0.4 and uses the sigmoid function as an activation function to predict the target classification. The AUCs of both the training and test sets of the model were greater than 0.95 (Fig. 3).

To further investigate the accuracy of the model sentiment classification, this study compared the results of manual classification of a corpus of 200 hotel reviews by both groups of researchers with the model classification results (Table 3). First, two groups of four people, who were not in contact with each other, were separately identified for the sentiment of the hotel review corpus. Secondly, the researchers classified hotel reviews into two categories of affective polarity (positive and negative) based on the proportion of positive and negative emotion words appearing in the reviews and the overall context of the reviews, where the emotion words were referenced to

[1] GitHub(2019)aespresso [Source code].https://github.com/aespresso/a_journey_into_math_of_ml.

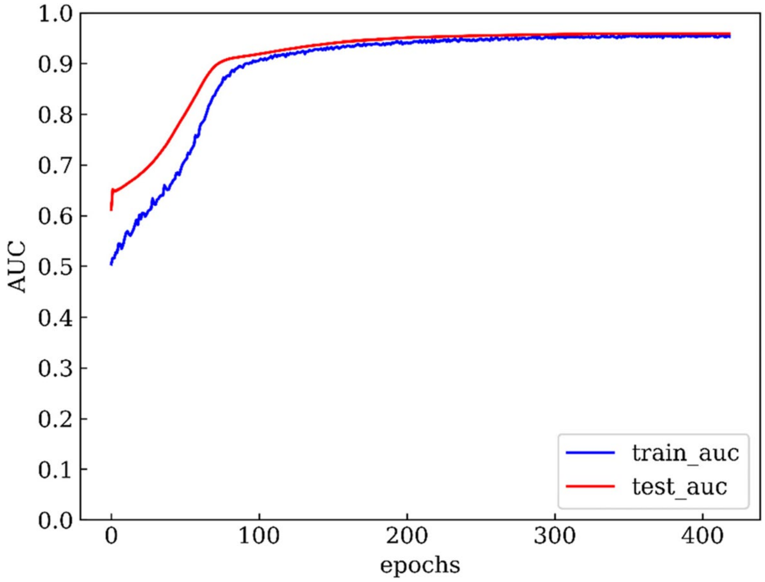


Fig. 3 Training and test set AUC

Table 3 Fits manual classification and the BERT model classification

Test	Research Group 1	Research Group 2	BERT model
Research Group 1	100%	86%	89%
Research Group 2	86%	100%	91%
BERT model	89%	91%	100%

the Chinese emotion dictionary How-Net. At the same time, the same comment corpus was classified for sentiment using the BERT model. Finally, the results of the manual classification were compared with the results of the BERT sentiment classification. It was found that the model classification results agreed with both sets of manual classification results by more than 85%. Therefore, the BERT model used in this study can discriminate the sentiment polarity of hotel review texts correctly and can be used to classify the sentiment of the review corpus for this study.

3.2 Topic-based social network analysis of commentary texts

Traditional topic mining methods mainly include word frequency analysis, co-word analysis, and citation analysis (He et al., 2018). These have a wide range of applications but are more subjective (Choi & Park, 2019). A more objective technique proposed by Blei for determining common themes in a set of texts to extract the distribution of themes and the distribution of subject words in the text uses the implicit Dirichlet distribution model (LDA) (Blei et al., 2001). It is a 3-layer Bayesian generative model of "document-topic-word", which can efficiently mine the topic information contained in a large number of documents and has no strict limitation on the text length. This approach can effectively and significantly discover the topic features of short texts (Zhao et al., 2020). We thus use the LDA topic model to mine the topics of hotel review texts, conceptualize the topics and their relationships as nodes and edges in the social network graph, and mine the core nodes and tight sub-clusters through graph mining methods to effectively identify text topics and their correlation structures (Maomao et al., 2021). In turn, we analyze the similarities and differences in customers' concerns about different-star hotels.

The specific approach used here is as follows:

(1) Data pre-processing: The comment text is first de-segmented, and the deactivated word database is constructed by selecting the deactivated word list of HIT and the deactivated word list of Baidu for deactivation and integration. Then, the Jieba package in Python is used to complete the word separation process of the comment text.

(2) LDA topic modeling: The classification topics of the text are established in the clustering results of the LDA model. Since the effect of LDA, topic extraction is directly related to the number of topics, we determine the optimal number of topics for each star hotel

Table 4 Topic - topic co-occurrence matrix example

Topics	Facilities	Breakfast	Location	Services	Check-in
Facilities	58,793	23,510	36,348	32,145	44,983
Breakfast	23,510	58,843	34,957	23,510	23,510
Location	36,348	34,957	90,698	23,510	59,939
Services	32,145	23,510	23,510	73,564	55,374
Check-in	44,983	23,510	59,939	55,374	106,361

Note: This table is based on the review data of four-star hotels on the Ctrip hotel platform

review (the optimal number of topics for 3-star is 4, for 4-star is 5, and for 5-star is 3) based on the topic distribution through LDAvis visualization mapping (Goloshchapova et al., 2019). To ensure clear boundaries between topics, feature words that are unclear and appear in multiple topics (e.g., "hotel", "room", etc.) were removed, and six words with relatively high frequency were selected as topic representatives, and the names of topic descriptions were further confirmed based on the semantic relationships of feature words.

(3) Thematic social network analysis: First, the feature words of LDA are organized and summarized, and the feature words under the same topic are used as topic feature identifiers, to construct the external co-occurrence matrix of topic-topic (Table 4). The elements on the non-diagonal line of the co-word matrix are the number of times two keywords appear in the same comment, and the elements on the diagonal line are the number of times the word appears in all comments. Second, an internal co-occurrence matrix was constructed based on the feature word-feature word co-occurrence relationship to reveal the feature word association relationship under a single topic (Table 5). Finally, the Gephi software was used to visualize the thematic social network of each star hotel review.

3.3 Spatio-temporal analysis

The kernel density analysis method can be used for surface density calculation and empirical analysis of aggregation class research. It calculates and estimates the data aggregation through the sample data and measures the density change of research elements through the established distance decay function, to explore the distribution and change characteristics of hot spots in the spatial region (Dong & Xu, 2020; Fang et al., 2021). We obtain the hotel location coordinates, count the distribution of each star hotel, spatialize the hotel sentiment value, and conduct kernel density analysis in ArcGIS software. Meanwhile, to explore the spatial distribution of the sentiment value of star-rated hotels in Nanchang, this study divided star-rated hotels into positive and negative hotels, where those with sentiment values higher than the average of star-rated hotels were positive hotels and vice versa (200 positive hotels and 162 negative hotels). The distribution of positive and negative hotels was subjected to kernel density analysis separately. The kernel density bandwidth was set at 2000 m according to the scope of this study area. Meanwhile, to explore the differences in emotions in the temporal order of each star hotel, this study conducted a comparative analysis of each star hotel by season (spring, summer, autumn, and winter). Among them, spring is March, April, May, summer is June, July, August, autumn is September, October, November, and winter is January, February, and December.

4 Analysis of results

4.1 Timing analysis

The number of reviews for each star hotel is broken down by seasons (Fig. 4 left). The number of customer reviews for each-star hotel follows a similar seasonal trend, showing a gradual increase in spring, summer, and autumn, and a decrease in winter. At the same time, this study analyses the sentiment value of each star hotel statistically by season (Fig. 4 right). The results found that the change in sentiment values with the seasons was consistent across the star hotels: The same trend in sentiment values for four and five-star

Table 5 Feature - feature co-occurrence matrix example

Feature Words	Facilities	Air-conditioning	Bad	Soundproofing	Restrooms	General
Facilities	4837	76	56	60	75	294
Air-conditioning	76	891	46	25	32	41
Bad	56	46	746	100	14	47
Soundproofing	60	25	100	676	11	55
Restrooms	75	32	14	11	607	36
General	294	41	47	55	36	2195

Note: This table is based on the data of the "general facilities" theme feature of 4-star hotels

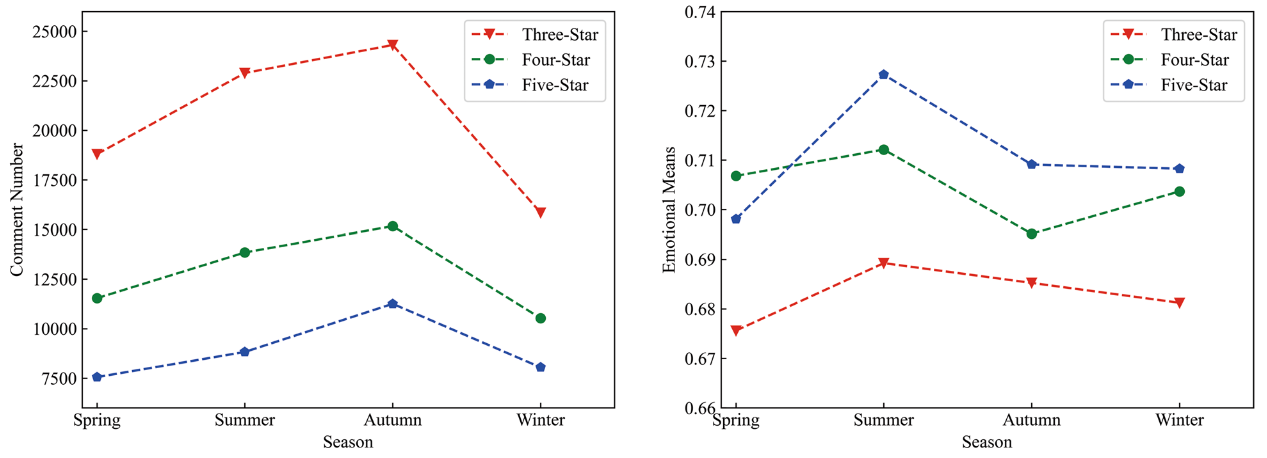


Fig. 4 Seasonal statistics on the number of reviews and mean sentiment values

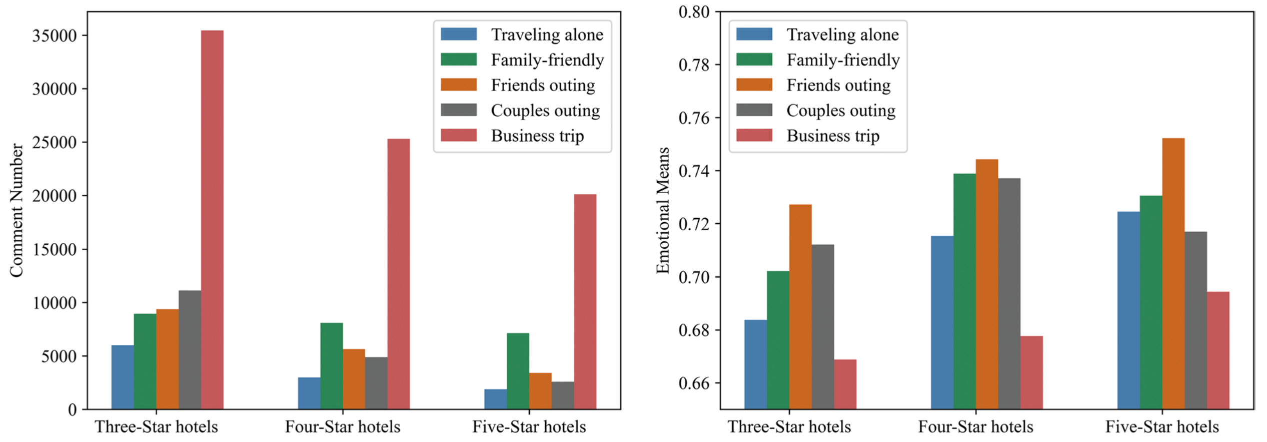


Fig. 5 Average number of people and emotions of different star travel types

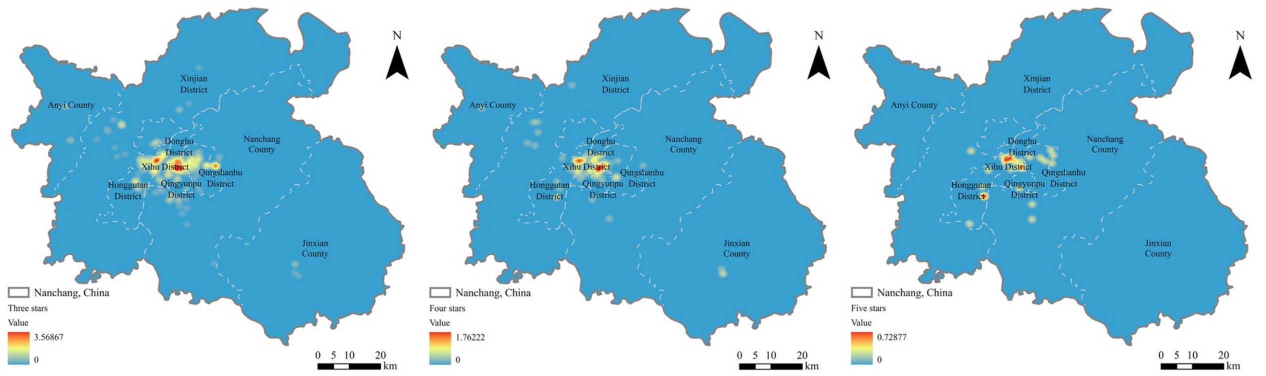
hotels with the seasons, with the highest in summer and lower in autumn; Seasonal sentiment averages for 3-star hotels, are highest in summer and lowest in spring.

We also analyze the number of reviews by trip type and the sentiment value by star-rated hotels (Fig. 5). The number of reviews was essentially the same across all star categories, with business trips being the most frequent and solo travel the least frequent. In the comparison of sentiment values, the differences are broadly similar between the different types of travel, with leisure trips of friends being the highest and business trips the lowest in all star-rated hotels.

4.2 Spatial analysis

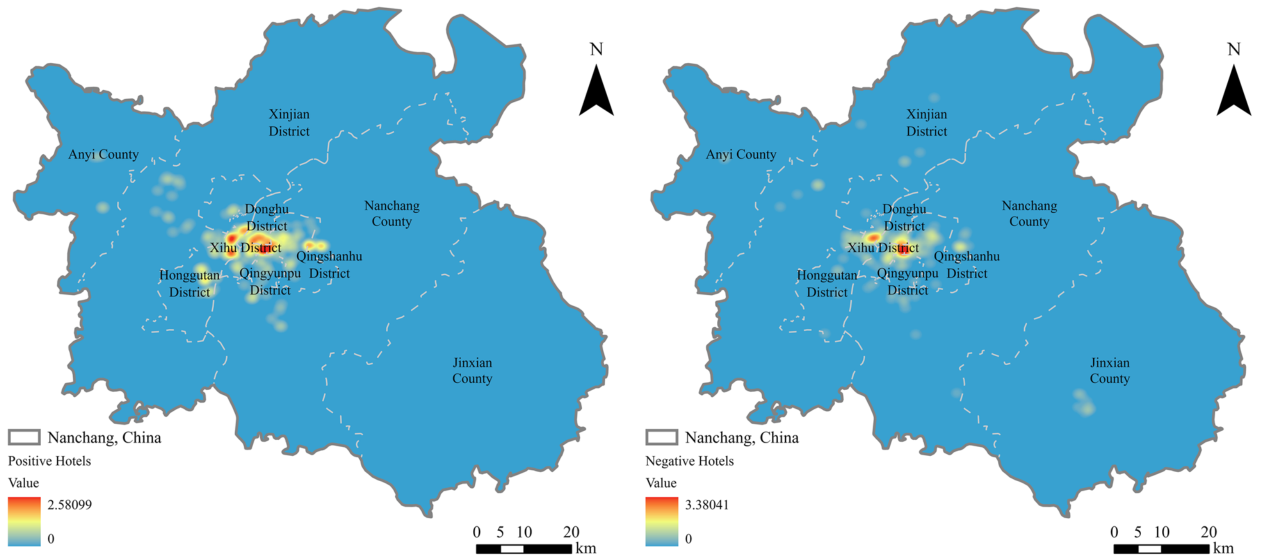
Each star-rated hotel is mainly distributed in the main urban area of Nanchang, showing an overall state of aggregation. A nuclear density analysis of the distribution of each star hotel (Fig. 6) shows that the spatial distribution range of different star hotels varies significantly, with its distribution range of three-star > four-star > five-star. The high-density area of the distribution of each star-rated hotel is similar, mainly distributed in the central part of Honggu Tan along the Ganjiang River and the border area of Qingshan Lake, Xihu District, Donghu District, and Qingyunpu District. In terms of distribution focus, the higher the star rating of the hotel, the more the high-density area is skewed towards the central part of Honggu Tan along the Ganjiang River.

This study conducted a kernel density analysis of the spatial distribution of positive and negative hotels (Fig. 7). It was found that negative hotels were more spatially clustered than positive hotels, with high-density areas mainly located in the border areas of Qingshan Lake District, West Lake District, East Lake District, and Qingyunpu District. Positive hotels are more widely



(a) Three-star spatial distribution (b) Four-star spatial distribution (c) Five-star spatial distribution

Fig. 6 Spatial distribution of the nuclear density of each star hotel



(a) Positive hotel distribution (b) Negative hotel distribution

Fig. 7 Spatial distribution of positive and negative hotel kernel density

distributed than negative hotels, with high-density areas mainly located along the Ganjiang River in the central part of Honggu Tan and the border areas of Qingshan Lake, Xihu, Donghu, and Qingyunpu districts.

4.3 Topic mining

The results of the LDA model for star-rated hotels (Table 6), with the three-star hotel themes being average facilities, warm service, convenient location, and comfortable stay. Four-star themes are average facilities, warm service, convenient location, comfortable stay, and good breakfast; Five-star hotel themes are average facilities, warm service, and convenient location.

Social networks were further constructed based on the results of the LDA model to explore the association between the textual themes of different star-rated hotel reviews. The Gephi software was used to explore and visualize the relationship between topic-to-topic associations and the co-occurrence of feature words under a single topic (Fig. 8).

The size of the nodes in the graph is proportional to the number of topic occurrences, which means that the larger the node, the more user attention the topic will have. The general theme of facilities accounts for the largest proportion of themes in all-star hotels, indicating that

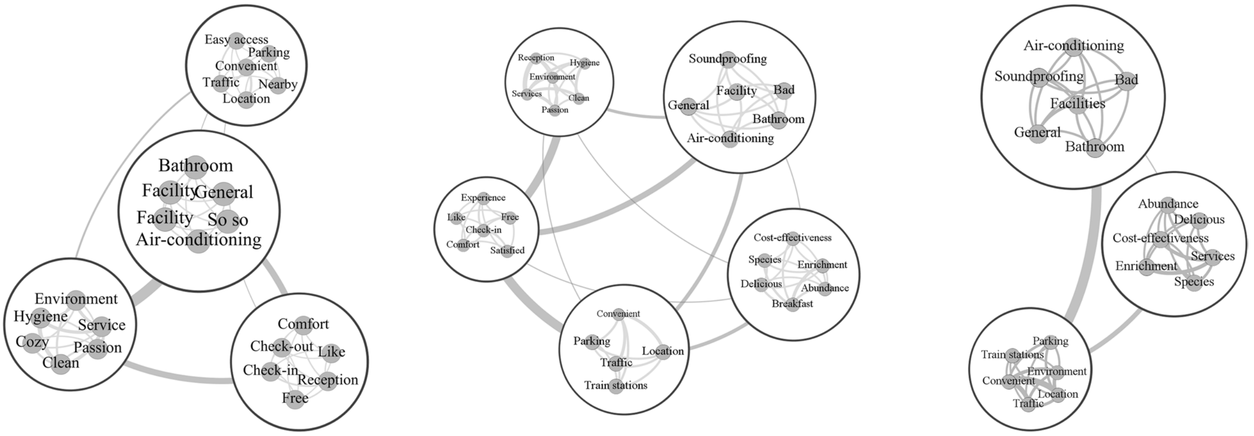
Table 6 LDA topic model results

Topics	3-star	4-star	5-star
Average facilities	Fair So-so Air Conditioning Facilities Bathroom Smell	Facilities Air conditioning Bad Soundproofing Bathroom Average	Facilities Average Air Conditioning Bad Soundproofing Bathroom
Passionate service	Service Clean Environment Hygiene Comfortable Enthusiasm	Service Front Desk Enthusiasm Clean Environment Hygiene	Service Abundance Value for money Variety Good food Abundance
Convenient location	Convenient Location Traffic Parking Convenient Nearby	Convenience Location Transportation Parking Train Station Environment	Convenience Location Transportation Parking Train Station Environment
Stay comfortable	Front Desk Check-in Like Check-out Free Comfort	Stay Like Comfortable Free Satisfaction Experience	
Rich breakfast		Breakfast Rich Value for money Variety Good food Plenty	

customers in all-star hotels are more concerned about the general theme of facilities. The themes common to all-star hotels are general facilities, warm service and convenient location, the theme common to three- and four-star hotels are comfortable to stay, and the theme specific to four-star hotels is rich breakfast. This suggests that facilities, location, and service are aspects that are common concerns to all-star hotels.

The thickness of the connecting lines between themes in the figure is proportional to the number of co-occurrence of themes in the corresponding nodes. This is reflected in the strong linkage between the three themes of average facilities, warm service, and comfortable stay in the social network map of the three-star hotels. The four-star thematic social network map shows strong links between the four themes of a comfortable stay, convenient location, average facilities, and warm service. For five-star hotels, the facilities are generally closely linked to the convenience of the location.

There is little variation in feature words under the same theme across star hotels. Most themes have a relatively even thickness of connecting lines between feature words, and individual themes have strong internal social network connections. This is evidenced by the fact that of the three and four-star hotel themes, the "passion for service" theme has the strongest internal social network links. Of the five-star hotel themes, the location facilitates closer social network connections. Words that clearly express emotions appear, such as "average, warm, bad", indicate the customer's emotional evaluation of various aspects of the hotel. This is evident in the general theme of facilities, where the presence of feature words such as "average, okay, and bad" indicate that customers do not rate hotel facilities highly. The presence of words such as "warm and comfortable" in the theme of enthusiasm for service indicates that customers are mostly positive about the hotel's service. In the convenience theme, the presence of words such as "convenient and handy" indicate that customers are satisfied with the location of the hotel.



(a) Three-star topics (b) Four-star topics (c) Five-star topics

Fig. 8 Topic social network visualization

5 Discussion of results

5.1 Main contributions

The contribution of this study is mainly in the research methodology: this study combines hotel POI and hotel review information to calculate hotel sentiment value and conducts a comparative analysis of three-star to five-star hotels. Specifically, hotel customer evaluations are analyzed from the perspective of time and space respectively, and the Spatio-temporal dynamic characteristics of customer evaluations of hotels with different stars are studied. Previous literature mainly studied hotels as a whole, lacking the analysis of differences for different star-rated hotels. At the same time, most of them analyze for a single review text and have not fully explored the hotel location information.

Meanwhile, this study has practical implications, mainly by capturing and processing information from Ctrip's hotel platform to obtain hotel social sensing data, and then using sentiment analysis, spatial analysis, thematic social network analysis, and other methods to carry out a comparative analysis of different star-rated hotels in terms of space and time, as well as fine-grained mining of text reviews, ultimately providing decision support for hotel managers, as well as providing a reference for customers to choose hotels.

5.2 Practical insights

Our findings have practical implications for hotel managers:

(1) Analyses of the changes in the number of occupants and the sentiment value of each star hotel with the seasons show that the number of customers staying in each star hotel has a similar trend as the seasons, showing a gradual increase in spring, summer, and autumn, and a decrease in winter. In addition, the sentiment value of each star hotel has a similar trend as the seasons, with an increase from spring to summer, a decrease from summer to autumn, and an increase from autumn to winter. Our results can thus provide hotel managers with the opportunity to implement corresponding management adjustments in response to seasonal changes, such as strengthening management and improving service quality in response to the situation that the number of occupants in each star hotel is the highest in autumn and the customer evaluations of the hotel are relatively low.

(2) Statistical analysis of the sentiment value of different types of customers in each star hotel showed that the sentiment value of different types of customers in each star hotel is the highest for the "friends trip" type and the lowest for the "business trip" type. Therefore, hotel managers can make corresponding service improvements for different types of customers, such as focusing on the needs of business travel-type customers.

(3) Through the spatial distribution of each star hotel and the nuclear density analysis of positive and negative hotels, it is found that each star hotel is mainly distributed in the main urban area of Nanchang, and basically in the city center. The high-density areas of each star-rated hotel distribution are similar, mainly distributed in the central part of Honggu Tan along the Ganjiang River and the border areas of Qingshan Lake, Xihu, Donghu, and Qingyunpu districts. In terms of distribution focus, the higher the star rating of the hotel, the more the high-density area is skewed towards the central part of Honggu Tan along the Ganjiang River. At the same time, the high-density area of positive hotels is more inclined towards the river area in Honggu Tan than the high-density area of negative hotels. Therefore, the difference in the distribution of different star hotels and the difference in the distribution of positive hotels and negative hotels can provide a reference for hotel site selection. For example, there are many high-star hotels along the riverside of Honggu Tan and the sentiment value of the hotels is relatively high. Therefore, hotel managers need to focus on the competitive influence of the surrounding star-rated hotels when selecting a site.

(4) In this study, by subject mining the hotel review texts and performing social network analysis on the results, we found that the results of different star hotel subjects have similarities and differences, and the differences of feature words within each subject are not significant. Customers often rate multiple aspects of a hotel at the same time. Therefore, the customer evaluations for different-star hotels are improved accordingly. For example, when customers rate the hotel's facilities low but rate the service and location relatively high, hotel managers can focus on improving the facilities aspect of the hotel. At the same time, hotel managers can enhance the management of the aspects that customers focus on according to the differences in the themes that customers focus on in different-star hotels. For example, four-star hotel customers pay more attention to breakfast, therefore, four-star hotel managers need to pay attention to the service and management of breakfast

6 Conclusion

We performed web-crawls and information processing from the Ctrip hotel platform to obtain hotel social sensing data. We then use sentiment analysis, spatial analysis, and thematic social network analysis to carry outa comparative analysis of different star-rated hotels in terms of space and time and topic mining of text reviews. We explored the similarities and differences in the Spatio-temporal dynamics of customer evaluation information and textual review themes of different star-rated hotels and provide theoretical guidance and practical reference for the effective management of different star-rated hotels. However, we did not analyze the influencing factors behind them here. Future research will consider the influencing factors of sentiment distribution with the help of relevant urban environments, socio-economic, and humanistic data.

Acknowledgments
The authors thank Prof. Adam T. Devlin, Dr. Jian Xu, and Dr. Qiyue Li from Jiangxi Normal University for their constructive suggestions on the article, which led to a great improvement in the quality of the article. Also, many thanks to Dr. Wang of Wuhan University for his advice on this paper.

Code availability
Not applicable.

Authors' contributions
All authors contributed to the study's conception and design. Material preparation, data collection, and analysis were performed by Jingbo Wang, Yuting Wu, and Yu Xia. The first draft of the manuscript was written by Jingbo Wang and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding
This research is partially funded by the Science and Technology Research Project of Jiangxi Provincial Department of Education, No. GJJ170211.

Availability of data and materials
The authors confirm the data and software used in this research are available.

Declarations

Competing interests
The authors declare that no conflicting interest in this research.

Author details
[1]School of Geography and Environment, Jiangxi Normal University, Nanchang 330022, China. [2]Key Laboratory of Poyang Lake Wetland and Watershed Research, Ministry of Education, Jiangxi Normal University, Nanchang 330022, China.

Received: 5 March 2022 Accepted: 8 May 2022
Published online: 20 May 2022

References

Acheampong, F. A., Nunoo-Mensah, H., & Wenyu, C. (2021). Transformer models for text-based emotion detection: a review of best-based approaches. *Artificial Intelligence Review, 54*(8), 5789–5829. https://doi.org/10.1007/s10462-021-09958-2

Berrar, D. (2014). An empirical evaluation of ranking measures concerning robustness to noise. *AI Access Foundation.* https://doi.org/10.1613/jair.4136

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research, 3*(Jan), 993–1022. https://sci-hub.wf/10.1162/jmlr.2003.3.4-5.993

Chang, Y. C., Ku, C. H., & Chen, C. H. (2017). Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor. *International Journal of Information Management, 48*, 263–279. https://doi.org/10.1016/j.ijinfomgt.2017.11.001

Choi, H. J., & Park, C. H. (2019). Emerging topic detection in a twitter stream based on high utility pattern mining. *Expert Systems with Applications, 115*, 27–36. https://doi.org/10.1016/j.eswa.2018.07.051

Dong, X., & Xu, S. (2020). Spatial evolution characteristics of urban and rural settlements in Inner Mongolia. *Arabian Journal of Geosciences, 13*(22), 1–11. https://doi.org/10.1007/s12517-020-06167-z

Fang, Y., Mao, J., Liu, Q., & Huang, J. (2021). Exploratory space data analysis of spatial patterns of large-scale retail commercial facilities: The case of Gulou District, Nanjing, China. *Frontiers of Architectural Research, 10*(1), 17–32. https://doi.org/10.1016/j.foar.2020.02.002

Goloshchapova, I., Poon, S. H., Pritchard, M., & Reed, P. (2019). Corporate social responsibility reports topic analysis and big data approach. *The European Journal of Finance, 25*(17), 1637–1654. https://doi.org/10.1016/j.foar.2020.02.002

Han, C., & Chiu, & Cheng. (2020). Application of support vector machine (SVM) in the sentiment analysis of the twitter dataset. *Applied Sciences, 10*(3), 1125. https://doi.org/10.3390/app10031125

He, W., Xie, H., & Feng, G. (2018). Review on latent Dirichlet allocation model. *Journal of Information Resources Management*. https://doi.org/10.1177/1729881420904213

Jia, F., & Chen, C. C. (2020). Emotional characteristics and time series analysis of internet public opinion participants based on emotional feature words. *International Journal of Advanced Robotic Systems, 17*(1), 1729881420904213. https://doi.org/10.1177/1729881420904213

Li, H., Liu, Y., Tan, C. W., & Hu, F. (2020). Comprehending customer satisfaction with hotels: Data analysis of consumer-generated reviews. *International Journal of Contemporary Hospitality Management*. https://doi.org/10.1108/IJCHM-06-2019-0581

Luo, X., & Yi, Y. (2019). Topic-specific emotion mining model for online comments. *Future Internet, 11*(3), 79. https://doi.org/10.3390/fi11030079

Maomao, C., Meiyu, P., & Weijun, W. (2021). A cross-platform comparative study of reviews on sharing accommodation and hotels reservation platform: Combined with LDA-SNA and sentiment analysis. *Library and Information Service, 65*(2), 107.

Muhammad, P. F., Kusumaningrum, R., & Wibowo, A. (2021). Sentiment analysis using word2vec and long short-term memory (lstm) for Indonesian hotel reviews. *Procedia Computer Science, 179*(6), 728–735. https://doi.org/10.1016/j.procs.2021.01.061

Rao, Y., Lei, J., Wenyin, L., Li, Q., & Chen, M. (2014). Building emotional dictionary for sentiment analysis of online news. *World Wide Web, 17*(4), 723–742. https://doi.org/10.1007/s11280-013-0221-9

Reagan, A. J., Tivnan, B., Williams, J. R., Danforth, C. M., & Dodds, P. S. (2017). Benchmarking sentiment analysis methods for large-scale texts: A case for using continuum-scored words and word shift graphs. *Ep Data ence, 6*(1), 28. https://doi.org/10.48550/arXiv.1512.00531

Rusanov, A., Miotto, R., & Weng, C. (2018). Trends in anesthesiology research: A machine learning approach to theme discovery and summarization. *JAMIA Open, 1*(2), 283–293. https://doi.org/10.1093/jamiaopen/ooy009

Stringam, B. B., & Gerdes, J. (2010). An analysis of word-of-mouse ratings and guest comments of online hotel distribution sites. *Journal of Hospitality Marketing & Management, 19*(7), 773–796. https://doi.org/10.1080/19368623.2010.508009

Tan, X., Zhuang, M., Lu, X., & Mao, T. (2021a). An analysis of the emotional evolution of large-scale internet public opinion events based on the BERT-LDA hybrid model. *IEEE Access, 9*, 15860–15871. https://doi.org/10.1109/ACCESS.2021.3052566

Tan, X., Zhuang, M., Lu, X., & Mao, T. (2021b). An analysis of the emotional evolution of large-scale internet public opinion events based on the BERT-LDA hybrid model. *IEEE Access, 9*, 15860–15871. https://doi.org/10.1109/ACCESS.2021.3052566

Tang, J., Hao, S., & Qu, W. (2019). Sentiment analysis of online Chinese comments based on statistical learning combined with pattern matching. *Concurrency and Computation: Practice and Experience, 31*(10), e4765. https://doi.org/10.1002/cpe.4765

Tao, Y., Zhang, F., Shi, C., & Chen, Y. (2019). Social media data-based sentiment analysis of tourists' air quality perceptions. *Sustainability, 11*(18), 5070. https://doi.org/10.3390/su11185070

Valdivia, A., Luzon, M. V., & Herrera, F. (2017). Sentiment analysis in TripAdvisor. *IEEE Intelligent Systems, 32*(4), 72–77. https://doi.org/10.1109/MIS.2017.3121555

Wang, F., Wang, Z., Wang, S., & Li, Z. (2014). Exploiting description knowledge for keyphrase extraction. In *Pacific Rim International Conference on Artificial Intelligence* (pp. 130–142). Springer. https://doi.org/10.1007/978-3-319-13560-1_11

Wang, J., Zhao, Z., Liu, Y., & Guo, Y. (2021). Research on the role of influencing factors on hotel customer satisfaction based on BP neural network and text mining. *Information, 12*(3), 99. https://doi.org/10.3390/info12030099

Wang, L., Wang, X. K., Peng, J. J., & Wang, J. Q. (2020). The differences in hotel selection among various types of travelers: A comparative analysis with a useful bounded rationality behavioral decision support model. *Tourism Management, 76*(Feb.), 103961.1-103961.16. https://doi.org/10.1016/j.tourman.2019.103961

Yang, K., Lee, D., Whang, T., Lee, S., & Lim, H. (2019). Emotionx-ku: Bert-max based contextual emotion classifier. arXiv preprint arXiv:1906.11565. https://doi.org/10.48550/arXiv.1906.11565

Yen, C., & Tang, C. (2018). The effects of hotel attribute performance on electronic word-of-mouth (new) behaviors. *International Journal of Hospitality Management, 76*, 9–18. https://doi.org/10.1016/j.ijhm.2018.03.006

Zhang, C., Ma, X., Zhou, Y., & Guo, R. (2021). Analysis of public opinion evolution in COVID-19 pandemic from a perspective of sentiment variation. *J Geo-Inf Sci, 23*(02), 341–350.

Zhang, H., Dong, J., Min, L., & Bi, P. (2020). A bert fine-tuning model for targeted sentiment analysis of Chinese online course reviews. *International Journal on Artificial Intelligence Tools, 29*(07n08), 2040018. https://doi.org/10.1142/S0218213020400187

Zhang, J., Lu, X., & Liu, D. (2021). Deriving customer preferences for hotels based on aspect-level sentiment analysis of online reviews. *Electronic Commerce Research and Applications, 49*, 101094. https://doi.org/10.1016/j.elerap.2021.101094

Zhao, F., Ren, X., Yang, S., Han, Q., Zhao, P., & Yang, X. (2020). Latent Dirichlet allocation model training with differential privacy. *IEEE Transactions on Information Forensics and Security, 16*, 1290–1305. https://doi.org/10.1109/TIFS.2020.3032021

Publisher's Note