REGULAR PAPER

Dong Li · Hong Yin · Changbo Wang · Sicheng Song · Kirlin Li · Chenhui Li

Visual Recommendation for Peer-To-Peer Accommodation with Online Reviews based on Sentiment Analysis and Topic Models

Abstract Peer-to-peer accommodation is developing rapidly in the era of sharing economy, and the visual recommendation of accommodation is also an urgent problem to be solved. Meanwhile, user-generated content is critical in P2P accommodations, because they contain a wealth of information about the opinions and experiences of users, which helps understand consumer decisions and improve products and services better. However, the huge volume of reviews makes it difficult for potential customers to gain useful insights and for managers to track customer opinions. In this paper, we propose a complete pipeline for recommending personalized accommodations for consumers, while also providing insights for managers. First, we use topic modeling techniques to mining opinions from review. Second, we build a deep learning network for review sentiment analysis. Third, we perform sentiment analysis of the reviews at the aspect level to obtain the sentiment vector representation of the accommodation. Finally, we propose a personalized accommodation recommendation method based on the above work. Moreover, we design a visual analytic system with a user-friendly interface to facilitate interactive analysis. Evaluation including user and case studies demonstrates the usefulness and effectiveness of our method and system.

Keywords Visual analysis · Opinion mining · Sentiment analysis · Personalized recommendation · Deep learning

1 Introduction

The sharing economy (Puschmann and Alt 2016) is growing rapidly in the short-term shared housing rental industry. Meanwhile, with the development of the Internet, user-generated content (Krumm et al. 2008) is growing rapidly. Potential consumers increasingly rely on this word-of-mouth information to evaluate

D. Li · H. Yin (✉) · C. Wang · S. Song · K. Li · C. Li
East China Normal University, Zhongshan, China
E-mail: hyin@jjx.ecnu.edu.cn

D. Li
E-mail: 51194501141@stu.ecnu.edu.cn

C. Wang
E-mail: cbwang@sei.ecnu.edu.cn

S. Song
E-mail: clare.song@foxmail.com

K. Li
E-mail: 52184501007@stu.ecnu.edu.cn

C. Li
E-mail: chli@sei.ecnu.edu.cn

products and services because online review information help them make informed decisions (Shi and Li 2011). At the same time, managers are increasingly recognizing the importance of using this information to gather feedback and evaluation of their own and competitors' products and performance. It provides a better understanding of consumer behavior and expectations, and identifies market opportunities.

However, the volume of reviews is so large that it is difficult for potential consumers to gain useful insights, which influence them to make informed decisions. It also makes it difficult for managers to track and manage customer comments. To our knowledge, there is a lack of sentiment analysis visualization systems for customer feedback data related to the short-term rental industry.

In this paper, we propose a new complete pipeline to analyze and deal with these problems. First, we use topic modeling techniques to mine a set of topics from a large number of online reviews of short-term rental platforms. We conduct experiments and visualize the topic model and consider them as aspects of interest to users. Second, we use deep learning models for sentiment analysis, combining topic model and sentiment model to analyze the reviews at the aspect level. We then obtain the sentiment vectors of different aspects of accommodations. Compared with online ratings, the implied sentiment features are more effective for managers to analyze business and assist consumers in decision-making. Finally, we propose a well-designed visualization system for tracking and managing consumer opinions and exploring potential patterns.

Our main contributions are as follows:

- We obtain a set of topics in the review corpus for accommodation analysis by unsupervised topic mining methods.
- We combine topic model and deep learning model for aspect-level sentiment analysis to construct implicit representations of accommodations.
- We design a visual analytic system to assist users in decision-making and provide empirical results with real data to demonstrate the effectiveness of the system.

This paper is built upon our previous conference paper (Li et al. 2021), but it differs in many ways from that work, where the main extensions are summarized in the following list:

(1) Supplement opinion mining topic model with a more detailed description of model training and evaluation.
(2) Improve the sentiment analysis network model, explain the network design in more detail and add the comparative experiment to prove the effectiveness of the model.
(3) Improve the visual effects of the visualization system to achieve a better user experience, and introduces the interaction of the system.
(4) Add more details and sufficient explanations in case studies and evaluations.
(5) Add new case study on how to use the system to perform topic and sentiment analysis on spatiotemporal analysis.

2 Related work

This section discusses the prior studies in the following four categories:

Topic modeling A probabilistic topic model is an algorithm for discovering latent semantic structures from a large number of documents. The most popular of which is latent Dirichlet assignment (Blei et al. 2003). It has a wide range of applications in text analysis, data visualization, recommendation systems, information retrieval and web analysis. Kim et al. (2018) use LDA model mining on online reviews to analyze customer preferences for detailed product attributes, providing realistic information that helps companies to manage their activities. In recent years, topic mining has also been widely applied to the hotel industry. Padma et al. (2020) use word frequency analysis to find the key attributes of luxury hotel services and promote the sound development of hotel services. Ding et al. (2020) use a structured topic model to extract service quality attributes from Airbnb reviews. Hu et al. (2019) use an improved structured topic model (STM) to analyze New York City Tripadvisor's hotel reviews, and provide detailed customer insights for the hotel by comparing the topics in positive reviews and negative reviews. In our work, we use LDA method on accommodation review data to extract topics that reasonably correspond to multiple aspects of accommodation. This helps us to identify the aspects that users care about and provides the basis for later data analysis modeling.

Text sentiment analysis Sentiment analysis (Agarwal et al. 2015) uses natural language processing (NLP), text analysis and computational techniques to automatically extract or classify sentiment from reviews. Statistics in April 2013 found that 90% of customer decisions rely on online reviews (Peng et al. 2014). Thousands of users rely on online sentiment comments to make decisions, making sentiment analysis technology very popular in decision-making applications. In the past few years, there has been a lot of research devoted to applying sentiment analysis in various industries, using various types of methods. Prameswari et al. (2017) use sentiment analysis and text summarization approach to mine scenic hotel reviews in Indonesia, ultimately improving the quality of the hotel industry and supporting tourism in Indonesia. Godakandage et al. (2020) provide a system for analyzing hotel reviews; they extract important aspects of reviews and perform sentiment analysis on them, and finally provide aspect-based personalized hotel recommendations, helping users to easily find the best hotel based on their preferences. Existing sentiment analysis techniques include sentiment dictionary-based methods, machine learning methods and deep learning methods. Deep learning methods that have emerged in recent years compensate for the shortcomings of traditional machine learning methods. Deep learning models including convolutional neural network (CNN) and recurrent neural network (RNN) are widely used in the field of text sentiment analysis. In this paper, we use bidirectional encoder representations from transformers (BERT) (Devlin et al. 2019) for word embedding and bidirectional LSTM (BiLSTM) (Mousa and Schuller 2017) model for sentence sentiment classification.

Accommodation recommendation At present, many scholars have carried out related research on the accommodation recommendation problem. Lerttripinyo et al. (2015) propose ACRoSS, a system that collects data from websites and selects representative reviews on predetermined topics through a feature-based weighted nonnegative matrix factorization method. It is used to help users choose accommodation easily and has achieved high user satisfaction. Weng et al. (2018) design a visual analysis system to help customers find the ideal place to live. They mainly focused on points of interest and accessibility. Although they had good experimental results, they did not consider related comments, sentiment and other factors. Different from the selection of traditional hotels, there are more factors to be considered in the selection and recommendation of P2P accommodations. Based on the existing research, this paper comprehensively considers the factors for selection and recommendation of P2P accommodations through topic mining and sentiment analysis.

Visualization of feedback data The rapid development of the Internet has brought about many websites. Many people leave ratings and text reviews on travel and accommodation sites or shopping sites. Data mining of customer feedback content is necessary for merchandising as well as service improvement. OpinionSeer (Wu et al. 2010) is an effective visualization system that compares customer opinions of different groups by visualizing online hotel customer feedback. Oelke et al. (2009) design an interactive cyclic correlation graph for customer reviews on online stores and product reviews on websites, allowing analysts to identify strengths and problems with different products and services. In our work, we use a treemap diagram (Fig. 8) to visualize customer feedback. In the tree diagram, each topic is used as a node, and the polarity of the comments under the topic is used as sub-nodes, which is represented by different colors.

3 Approach overview

We describe the pipeline of our system in this section. As shown in Fig. 1, it consists of three modules.

The first module is the data model. The input data is the accommodation listing data and online review data. All raw data is available on Airbnb[1]. We borrowed existing Python modules and crawler methods to write a Python crawler program for the Airbnb website, and used proxy IP and multi-threading technology to crawl data on specific web pages. In general, the data crawled in this paper covers all housing listings in the Shanghai area and the time span is from 2017 to 2020. The data can be divided into three parts: housing listing data, order data and consumer data. Order data refers to basic accommodation records, including occupants, accommodation time and the listing number of the accommodation booked; listing data includes listing location information, category to which listings belong, and overall evaluation of listings; consumer data includes consumption reviews of people on the property. We perform data filtering and cleaning on raw data, and finally store them in the database.

The second module is the data analysis module. First, we use topic modeling techniques to mine accommodation reviews and obtain the best topic model through experiments. This could help us summarize the key attributes of consumer comments and mining consumers' concerns. Then, we train a sentiment analysis model using advanced deep learning methods. Finally, on the basis of the previous work, we perform topic-based sentiment analysis on the reviews. This is an important condition for us to provide consumers with personalized accommodation recommendations.

The third module is the visual design module. We provide a visual analysis system for P2P accommodation recommendation, which takes online accommodation reviews, sentiment characteristics, geographic location and price as visual objects. The system includes a set of well-designed visual views, and introduces a large number of interactive analysis functions to realize multi-dimensional display and interactive operation of accommodation sharing and online review data. The visualization system takes the computer software development technology as the core, takes the portability and maintainability as the design principle and takes the ease of use as the goal. Accommodation recommendation results and sentiment-related information on review topics are directly displayed to users through a web interface. At the same time, users can configure basic options through the navigation view, and can explore potentially interesting patterns through association analysis of multiple visualization components.

4 Data analysis

In this section, we introduce our analytical approaches used in the data analysis module.

4.1 Topic model for opinion mining

As mentioned in Sect. 3, the first step of our work is to discover the aspects that users care about, so we integrate topic modeling into our pipeline. We use all text data from customer reviews for topic modeling.

Model train There are many implementations of the LDA. In this work, we utilize the LDA method based on a variational Bayesian algorithm (Hoffman et al. 2010), as it is more efficient than the original one. It uses a stochastic adaptive strategy to update the posterior parameters of the topic.

Determining the number of topics in the LDA is a difficult problem. When the similarity between the topics is minimized, the right number of topics is found. In the LDA model training, the best indicator for evaluating the model is considered as the log perplexity:

$$Perplexity(D) = exp(-\sum_{m=1}^{M} \log_D P(w_m) / \sum_{m=1}^{M} N_m) \tag{1}$$

where D represents the test set in the corpus, M represents the number of documents, N_m represents the number of words in the document M, $P(w_m)$ represents the probability of w_m, and the perplexity reflects the predictive ability of the model. For a document d, the uncertainty of our model about which topic document d belongs to is the perplexity.

In our experiments, we extract nearly 200, 000 comments for topic word mining, and we set the number of topics k from 2 to 20. First, we calculate the model perplexity of the corpus under different number of topics. Second, we select the models with the least perplexity and check their topic distribution. Finally, the perplexity achieves the minimum when $k = 8$ as shown in Fig. 2.

Model evaluation We use the perplexity to evaluate our model. Since the output of the LDA is a set of semantically similar word sequences, it is difficult for users to understand the generated topic. Therefore, we apply LDAVIS (Sievert and Shirley 2014), a visualization method, to verify the feasibility of our model. A visualization of the model, including the most relevant terms for each topic and the relationships among each topic, is acceptable for non-technical users. We visualize the best LDA model as shown in Fig. 3.

We find that only two topics (Topic 7 & 8) partially overlap among the eight topics, and the remaining topics are independent of each other. This shows that our model works well, and the correlation between the various topics is not high. We also study the two overlapping topics and find that the terms in Topic 7 are related to subway stations and transportation, while those in Topic 8 are related to geographic location. In fact, these two aspects do have similarities, because location includes the factor of subway stations. However, the latter also includes other factors, such as the distance to attractions and the distance to shopping malls.
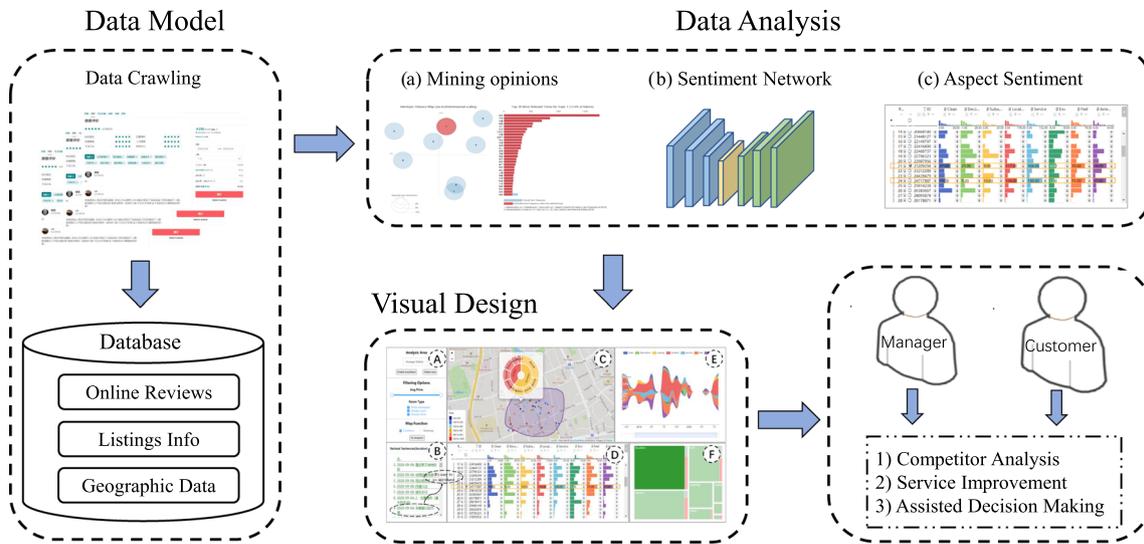
Fig. 1 The pipeline of our system, which includes data module, data analysis module, visual design module. Data module includes some work of data crawling and preprocessing, and stores the data in the database. Data analysis module includes three important parts, topic mining, sentiment network training and aspect-based sentiment analysis. Visual design module introduces the design principles of our visualization system and the use of visualization components
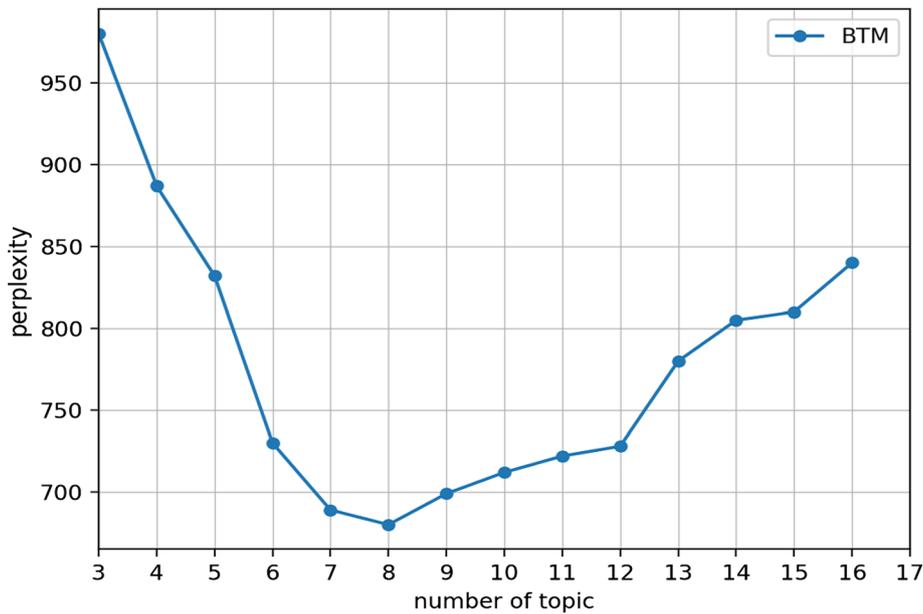
Fig. 2 The perplexity curve of the topic model under different number of topics

We summarize all the topics and salient terms in Table 1. It shows that there are eight aspects of concern reflected in customer reviews, namely clean, decoration, service, quiet, feeling, amenity, subway and location.

4.2 Sentiment analysis for review

In this section, we present two steps of opinion sentiment analysis. First, we apply the state-of-the-art method BERT (Devlin et al. 2019) to embed sentences into a tensor space. Second, we represent each sentence in an accommodation review as a two-dimensional tensor and use BiLSTM (Mousa and Schuller 2017) for sentiment polarity classification. The process is shown in Fig. 4.
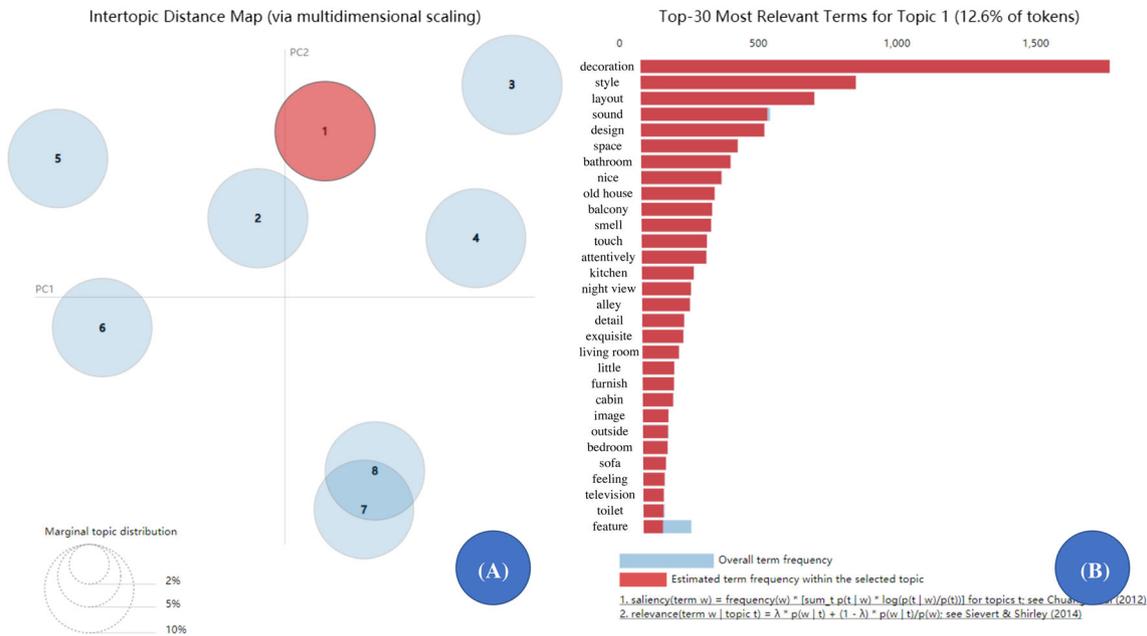
Fig. 3 Topic model visualization and evaluation. **a** The overall distribution of topic models. **b** Words most relevant to the topic

Table 1 Opinions extracted from comments

Topics	Most relevant terms
topic 1(clean)	Clean, restrooms, comfortable, neat, satisfied...
Topic 2 (decoration)	Decoration, style, design sense, warm, amazing...
Topic 3 (service)	Warm, attentive, thankful, considerate, super nice...
Topic 4 (quiet)	Quiet, sound insulation, comfortable, good environment, not noisy...
Topic 5 (feeling)	Great experience, comfortable, warm, satisfaction, value for money ...
Topic 6 (amenity)	Full facilities, design, kitchen, sofa, balcony...
Topic 7 (subway)	Subway, location, distance, easy to travel, very close...
Topic 8 (location)	Location, convenience, attractions, neighborhood, downtown...

Tensor embedding Word2Vec (Mikolov et al. 2013) is the most commonly used word vector tool in NLP, but it is context-independent. In our pipeline, we employ a context-dependent model BERT. BERT provides state-of-the-art performance and improves many NLP tasks (Dai and Le 2015) and subsequent tasks. So we choose it as our pre-trained model. We use BERT to obtain static representations of sentences. We divide each review corpus into sentences and input each sentence to the model for encoding. The output is an $L \times 768$ matrix, where L is the length of the sentence.

Polarity classification In this step, we use BiLSTM as our prediction model. First, we collect documents from the corpus of accommodation reviews proposed by Songbo Tan. We acquire 4, 000 reviews including 2, 000 positive reviews and 2, 000 negative reviews. Then, we expand the dataset. We crawl a lot of reviews with user scores from Ctrip[2]. Next, we filter reviews with scores below 2.0 (Full score of 5 of 5.0) from the corpus and consider them as negative reviews. We consider the reviews with scores of 5.0 as positive reviews. Finally, we add 2, 000 positive reviews and 2, 000 negative reviews to expand the dataset. Our dataset totally includes 8, 000 reviews.

Inspired by Minh et al. (2018), we combine the advantages of sentiment dictionary and deep learning to propose a dual-channel fusion model of semantic information and sentiment information based on Bert word embedding (Bert-BiLSTM-EIC) for comment sentiment analysis. As shown in Fig. 4, downstream model includes two input channels. The first is the semantic information channel, which is used to extract semantic information, and the second is the emotional information channel, which is used to extract the emotional information features of sentences. The feature fusion layer merges the features extracted from the two
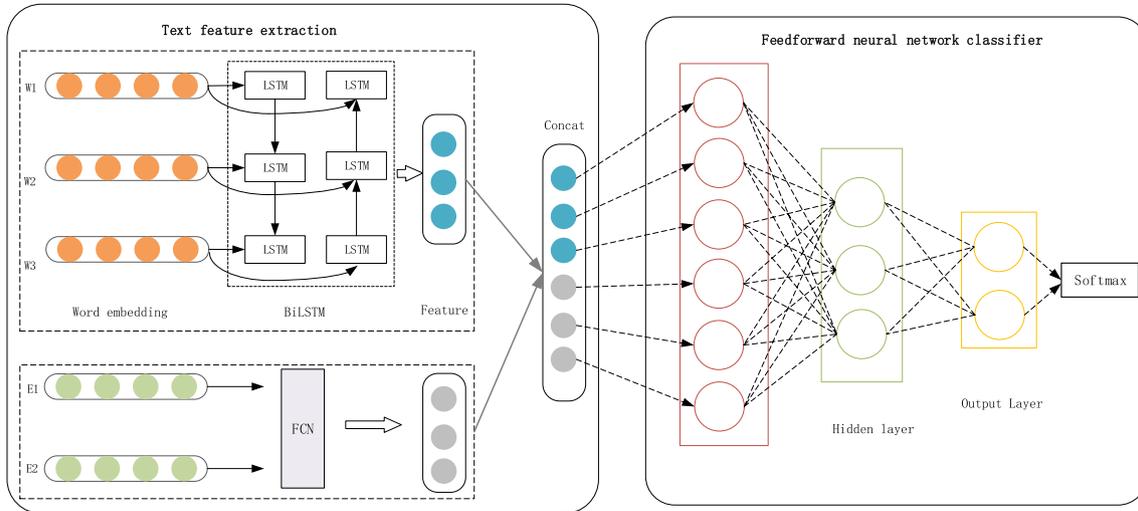
Fig. 4 Neural network overview. The input is a sequence of sentences extracted from a review. The output is the probability

channels, and then passes through the FCN layer and the Softmax layer to obtain the probabilities belonging to each category.

Model evaluation The training dataset includes 70% reviews. The testing dataset is composed of the remaining 30% reviews. We use python 3.6 and PyTorch with a GPU of 1660ti. We finally achieve **94**.42% accuracy for our model, which shows that the model is effective for subsequent sentiment prediction. We also compared with other classic methods, and the experimental results are shown in Table 2.

In fact, some reviews did not show strong emotional inclinations. Therefore, we predict the review as positive or negative only when the probability of the polarity category is greater than 0.7. Otherwise, we consider the opinion to be neutral.

4.3 Sentiment vector generation for recommendation

We combine both LDA model and sentiment classification model for aspect-based sentiment analysis to construct sentiment vectors. We describe the definition and generation process of the sentiment vector below.

Definition We define the sentiment vector space with $|K|$-dimensions, where $H = \{h_1, h_2, ..., h_N\}$ are accommodations. Given a collection of reviews associated with H, the sentiment vector space is designed as follows. We define the number of topics as K, and $T = \{t_1, t_2, ..., t_K\}$ are topics extracted by LDA model in review comments. We regard these topics as different aspects of accommodations and define that each topic corresponds to a dimension of accommodation sentiment vector to represent accommodation sensibly. For each accommodation $h_n \in H$, we denote the vector of accommodation h_n as $\vec{V}^{h_n} = \{\vec{s_1}, \vec{s_2}, ..., \vec{s_k}\}$, where value s_k is the sentiment of each topic in review comments.

Generation We extract accommodation h_n's review comments $r_n = \{r_{n1}, r_{n2}, ..., r_{nm}\}$ from text corpus, where M is the number of comments and r_{ni} is a review which belongs to h_n. For each comment $r_{ni} \in r_n$, we need to assign sentiment to topics respectively. A single review includes multiple aspects, which belong to different topics. So we split a comment r_{nm} into sentences. For each sentence S_r, we generate the topic-sentiment pair as $<t^{s_r}, S>$, where t^{s_r} is the topic and S is the sentiment of sentence S_r. The sentiment polarity of S_r is predicted by our model and S is calculated by Eq. 2. The sentiment of each topic in comment r_m is calculated by aggregation of such topic-sentiment pair.

$$S = \text{compute (Sr)} = \begin{cases} 1, & \text{polarity (Sr)} = \text{positive,} \\ 0, & \text{polarity (Sr)} = \text{neutral} \\ -1, & \text{polarity (Sr)} = \text{negative.} \end{cases} \qquad (2)$$

LDA is not effective in inferring the topic of the sentence since the sentences are usually short. We define $P(t_k|r_m)$ as the probability of a review r_m related to topic t_k. Similar to the topic prediction process in Jo and Oh (2011), we infer the topic of a given sentence S_r by the following equation:

Table 2 Experimental results of different methods

Methods	P/%	R/%	F_1/%
Bert-CNN	89.76	88.92	89.34
Bert-LSTM	91.76	90.92	91.34
Bert-BiLSTM	92.62	92.24	92.43
Bert-BiLSTM-EIC(Ours)	**94.42**	**93.38**	**93.90**

$$P(S_r|t_k) = \prod_{W_i \in s_r} P(W_i|t_k) \cdot P(t_k|r_m) \cdot P(r_m) \tag{3}$$

where $P(S_r|t_k)$ is the probability that the sentence S_r related to topic t_k. $\prod_{W_i \in s_r} P(W_i|t_k)$ is the probability that the words W_i of S_r related to topic t_k, which is given by topic-word matrix from the LDA model. $P(t_k|r_m)$ is the probability that topic t_k related to review r_m, which is also calculated by the LDA model.

We calculate probability $P(W_i|t_k)$ of each topic $t_k \in \{t_1, t_2, ..., t_K\}$, and choose the one which has highest probability of $P(S_r|t_k)$ as the topic of sentence t^{S_r}:

$$t^{S_r} = argmax\{P(S_r|t_k)|t_k \in \{t_1, t_2, ..., t_K\}\} \tag{4}$$

Algorithm 1 Introduces the complete generation process.

Algorithm 1 Sentiment Vector Generation Algorithm

Input: accommodation h_n, sentiment analysis model, topic model
Output: h_n's sentiment vector $V^{h_n} = \{s_{n1}, s_{n2}, ..., s_{nk}\}$

1: $r_n = D(h_n)$
2: **for all** r_{nm} in r_n **do**
3:      split comment r_{nm} into sentence list $Sen = \{S_1, S_2, ...S_r\}$
4:      **for all** S_r in Sen **do**
5:          $t^{S_r} = argmax\{P(S_r|t_k)|t_k \in \{t_1, t_2, ..., t_K\}\}$
6:          $S = compute(S_r)$
7:          generate topic-sentiment pair $< t^{s_r}, S >$
8:      **end for**
9:      Aggregate and average the topic-sentiment pairs of all sentences to get the sentiment vector of r_{nm}
10:     $V_i^{h_n} = \{s_i^{n1}, s_i^{n2}, ..., s_i^{nk}\}$
11: **end for**
12: $V^{h_n} = \sum_{i=1}^{m} V_i^{h_n}$
13: **return** h_n's sentiment vector $V^{h_n} = \{s_{n1}, s_{n2}, ..., s_{nk}\}$

Recommendation Through the above algorithm, we know the sentiment of each accommodation in various aspects. This is more authentic than a simple rating on a website for consumers. Different consumers have their own personal preferences for accommodation. Our algorithm finally recommends accommodation to users that meet their preferences through weighted calculations. Suppose one user's preference vector for accommodation is $u_i = \{r_{i1}, r_{i2}, ..., r_{ik}\}$, k is the number of topics. Finally, we use Eq. 5 to calculate the score for each accommodation:

$$score = \sum_{i=1}^{k} u_i^T V^{h_n} \tag{5}$$

5 Visual design

Analyzing customer feedback data interactively through visualization system is urgently needed, as it not only helps potential customers make informed decisions when booking products, but also helps accommodation managers track and manage customer feedback. In addition, since end users such as potential customers or accommodation managers may not have much background in information technology, the

system should be designed to be as simple and intuitive as possible. By iteratively discussing with experts to improve the system prototype, we designed a good set of visualization techniques and user-friendly interactions. This section describes the design goals of our visualization system and introduces the system.

5.1 Design requirements

In order to understand the problem domain and determine potential use for the design, we work closely with many accommodation managers. Through a series of interviews, we refined design requirements, demonstrated prototypes and gathered feedback to improve our system iteratively. The analytical tasks are described below.

T1 What is the geospatial distribution of accommodations? How to filter the accommodations? How to select accommodations in a specific area for analysis?

T2 What is the distribution of customers' travel types? Do they prefer to travel on weekdays or holidays?

T3 What aspects are included in the reviews? What are their sentiment values in various aspects? How to effectively display and compare the differences among accommodations.

T4 How to find the advantages and disadvantages? How to succinctly present a description of all aspects of the review (including both positive and negative views)?

T5 How do sentiment evolve in time series? Are there any underlying driving factors when the sentiment values appear abnormal?

These requirements help us propose the appropriate design principles and make informed decisions on our visual design. As shown in Fig. 5, our system interface consists of a series of views that meet the design requirements.

5.2 Map overview

Map view is a common method in multidimensional geographic information visualization research. We use a map to show the geographic distribution of accommodations in this study.

Users can select different administrative districts to explore and analyze. The map shows the geographic boundaries of the corresponding districts. At the same time, the selection of the analysis area can be made using our tool. The purple polygon is the area selected for analysis as shown in Fig. 5c. The map consists of three views, including a heatmap, point mapping view and sunburst view.

Fig. 5 System overview. **a** Navigation window provides functions of selection and filtering. **b** Customer feedback shows the raw review. **c** Map view shows the geographical distribution and provides interactions. **D** Visual comparison view shows ranking of related listings. **e** Theme river shows the evolution of sentiment over time sequences. **f** Treemap diagram shows the review opinions for a particular accommodation

Fig. 6 **a** Heatmap of the accommodations. **b** Point mapping view of the accommodations

Heatmap We obtain the geographic location information of all relevant accommodations and display it on the map in the form of a heatmap (Fig. 6a). This overview view allows the user to know which areas are popular. If the end user is a homeowner or accommodation manager, the view gives him an approximate idea of how many potential competitors exist in the neighborhood.

Point mapping view Accommodations are plotted on the map as circles (Fig. 6b). We divide the average price of accommodations by zones, assign different colors and map the colors to these circles. The mapping of prices and colors reflects the popularity of a certain area, and provides an intuitive reference for accommodation prices or the choice of potential customers.

Sunburst view This view is an overview of accommodation customers. Accommodation managers would like to know the types of their customers. We conduct a statistical analysis of the accommodation, and combine the two dimensions of travel time and travel type to design this view (Fig. 5c). The first ring of the view is the percentage of weekdays and holidays, and the second ring reflects the percentage of the type of travel. The customer-side analysis helps the accommodation improve its services (rental time adjustment, housing structure).

5.3 Visual comparison view

We use the LineUp view Gratzl et al. (2013) to show the features and differences among accommodations. The LineUp is an interactive technique designed to create, visualize and explore rankings of items based on a set of heterogeneous attributes. Users can analyze the differences among accommodations with the LineUp view (T3). The left side of Fig. 7 is the selected area, which is selected on the map view by turning on the AreaSelect function. The corresponding LineUp view is on the right side. The "ID" is a unique identifier of the accommodation, and each accommodation contains eight aspects of sentiment indicators. The indicators are represented by different colored bars, which are described in Sect. 4.1. By clicking on an accommodation in the view list, the LineUp view highlights the accommodation selected by the user and display the sentiment value. At the same time, it triggers an update of the corresponding treemap diagram (described in Sect. 5.4).

Moreover, LineUp allows users to customize the sorting by customizing the property priority. This makes it easy to let users know how accommodations are ranked in different areas. Whether they are
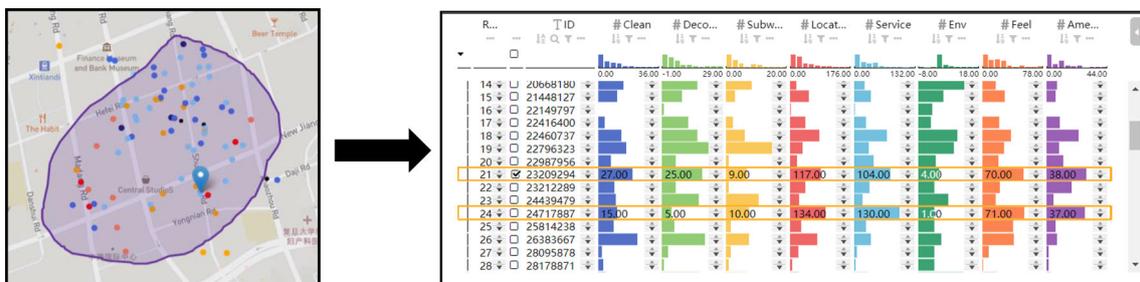
Fig. 7 LineUp view shows sentiment indicators of selected accommodations, and each accommodation (left) has their indicators in the LineUp view (right)

potential customers or accommodation managers, LineUp provides them with useful suggestions. As shown in Fig. 7, we select two accommodations and display their indicators. The first accommodation has a sentiment value of 27.00 for "cleanliness" and the second accommodation has a sentiment value of 15.00 for "cleanliness," so we believe that the first accommodation has higher feedback on "cleanliness" than the second accommodation.

5.4 Review opinion view

We design the treemap diagram (Fig. 8) to visualize the sentiment features of the feedback data (T4). In our design, each 1-level node of the tree diagram represents an aspect (obtained from Sect. 4.1), and we map the number of comments related to the aspect to the node size. This design makes it easier for us to know what is the focus of most consumers for this accommodation. In addition, each node contains some sub-nodes (two-level nodes), and there will be at most three sub-nodes at the same time, which represent positive (green), negative (red) and neutral (gray) comments. In the color design of the child nodes, we refer to the knowledge of color psychology. Red represents danger, just like most of the signs (traffic lights, standstill) and buttons (fire button) that represent warnings are red. Green represents the meaning of passing and positive. Gray, it lies between black and white and represents neutrality in most cases. Finally, the aggregate sentiment value is mapped to the size of the child nodes (two-level nodes).

For the interaction, when the mouse moves over each rectangle, the border of the corresponding rectangle is bolded to indicate that the topic is being viewed. When the corresponding rectangle is pressed, the selected topic is highlighted, and the feedback list shows the views of the corresponding topic. As shown in Fig. 8, when the user presses the negative feedback rectangle of the service topic, the feedback list shows the negative views related to the "service" topic. The positive feedback related to the "decoration" topic is shown on the right side.

In addition, we optimize the treemap diagram. We rank each topic according to the positive sentiment value. This design helps recognize the strengths and weaknesses of accommodation well from the feedback (T4). The aspects of accommodation with the best feedback should be focused, while the aspects with the worst feedback should be improved. Experts also believe that our sorting improvement is more effective than the original tree diagram.
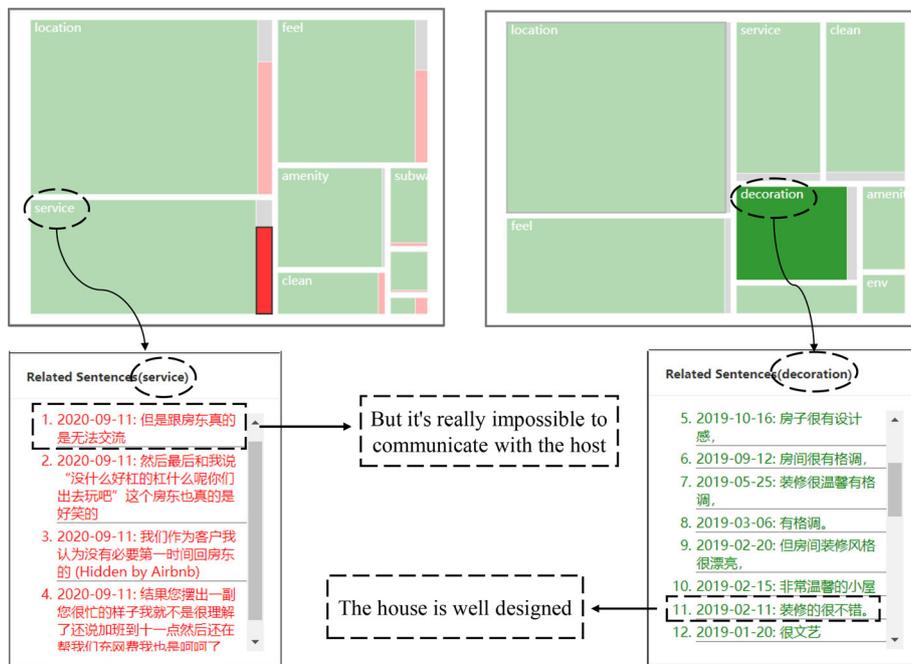
Fig. 8 Treemap diagram and opinions list shows various aspects of feedback sentiment

5.5 Sentiment evolution view

The OpinionManager system is also provides a sentiment theme river (Fig. 5e) to show the evolution of theme sentiment on a time series. The flow can usually be used to represent the change of time, the change of the flow can reflect a certain trend, and the change trend of the flow can be easily observed by people. In our design, each aspect (theme) is coded into a stream and assigned a color, where the color coding matches the color of the bars in the LineUp view (Sect. 5.3). As far as we know, the change of time series sentiment value also reflects the travel of the visitors (T2). For example, seasonal changes in sentiment values indicate that accommodations are often seasonal as well (reviews can only be made if rent is available). Besides, the sentiment evolution view can easily capture abnormal changes. There are usually some driving factors behind the large changes in sentiment values, and we can easily find these potential reasons through the feedback of the comments. Our system can observe and analyze abnormal sentiment fluctuations and analyze the possible factors behind them, which is of great help to the improvement of follow-up services and the formulation of policies (T5).

5.6 User interactions and system implementation

Our system provides the following interactions to help with data analysis:

Navigation with overview Users can select the administrative area to be analyzed in the drop-down box. The corresponding geographical boundaries would be displayed on the map. The system also provides some basic filtering options, such as filtering by average price and room type. Finally, the user can enable the area selection function in the navigation bar, which allows drawing any polygon on the map to determine the area to be analyzed. After turning off the area selection function, users can move and zoom the map view. The Delete Area button would clear all selected areas on the map.

Highlight and brush Many visual elements in our system are associated with each view. To allow the end user to get consistent information from different aspects, the relevant visual components are highlighted if they are selected. When a rectangle in the tree diagram is clicked, the rectangle is highlighted and the corresponding content is displayed in the feedback list. In addition, a sliding function has been added to view the situation during a specific time period.

System implementation According to the needs of business scenarios and system users, the system adopts B/S architecture for development. The front-end interface of the system uses the latest HTML5, D3.js, Echarts and other technologies. This provides users with an interactive Web interface that is responsive, concise and easy to operate. Meanwhile, Spring MVC is used for back-end development to decouple between layers, reduce dependencies and ensure stable system operation.

6 Case study

To show the effectiveness of the system, we invite experts and business managers to conduct case studies. More content can be found in our video as supplementary material[3].

6.1 Assist in customer decision making

Online reviews can help potential consumers make decisions. We invite travel enthusiasts to use our system, who have different requirements for accommodation. Suppose Tourist 1 now would like to choose an accommodation in Huangpu District. First, he has high requirements for the location of the accommodation, which is close to the subway station and surrounded by many shopping malls. Second, he cares about the environment around the accommodation and hopes that the environment is quiet. OpinionManager supports customized sorting, which allows us to set priorities for different topics and gives the final ranking results. The priority here can actually be understood as the user's preference requirements. The higher the priority, the higher the corresponding topic emotional weight. Our algorithm will assign different weights according to the priority. Therefore, the system will give the final TopK ranking results according to the user's personalized requirements to assist consumers in making good decisions.
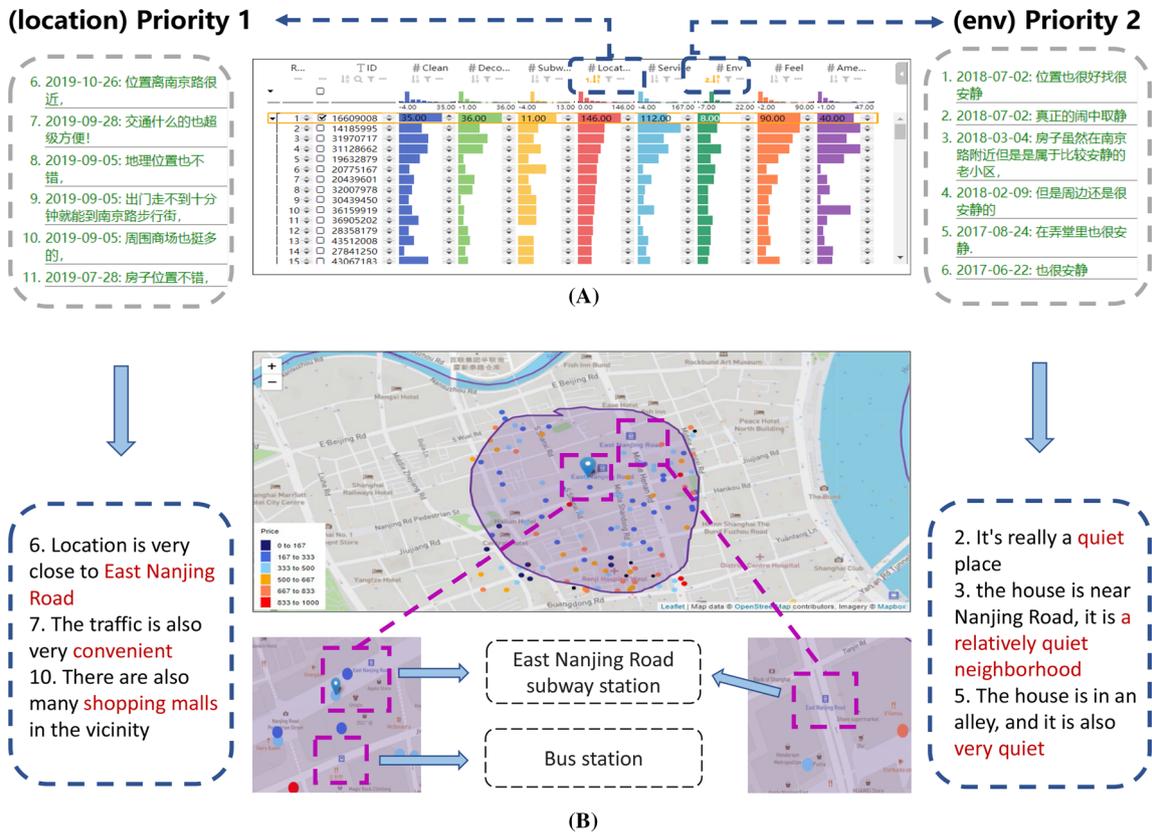
Fig. 9 OpinionManager assist in the decision making of potential consumers. **a** Users set the priority of each aspect according to their personal needs. **b** Locations of the recommended accommodations

As shown in Fig. 9a, he set the priority for different aspects of the accommodation according to his needs and obtained the final ranking result. To prove the validity of the recommendations, we check the top recommended accommodation locations. As shown in Fig. 9b, we find that the top-ranked accommodation is located near Nanjing East Road. It is an excellent location with two subway stations and bus stops in the vicinity. In addition, we access the online link [4] to the accommodation via the unique ID. We find that the accommodation is located in a quiet elevator apartment in the middle of a downtown area. The feedback of these two aspects is also consistent with the actual as shown in Fig. 9b.

6.2 Anomaly in time-series sentiment

Sentiment anomalies on time series are of great concern to both accommodation managers and consumers. We discuss two cases. The first one is the sentiment fluctuation caused by an unexpected disaster. For example, Fig. 10a shows that the sentiment of each theme suddenly all dropped to 0 in the red dashed box (we examine many accommodations in the same time period), which is generally rare. After discussions with experts, we know that in the year 2020, all over the world was affected by COVID-19 and the travel industry was slow to develop, making both P2P accommodation and the traditional accommodation industry suffer heavy losses. We selected an accommodation with a high degree of popularity for further analysis. Accommodation popularity can be measured by the number of reviews. Figure 11 shows the thematic treemap of the content that consumers care about in accommodation before and after the epidemic. It can be seen that before the outbreak, most of the comments on this listing were about location and services, and the topic of hygiene was only ranked 6th; after the outbreak, the proportion of topics changed from the original 6th to the 6th. 1. This shows that after the outbreak, the hygiene theme seems to have attracted more

---

Par72 https://zh.airbnb.com/rooms/16609008.

attention from consumers, which may have something to do with Airbnb's relevant policies or hygiene requirements during the epidemic.

The second is a more granular level of sentiment anomaly detection. As shown in Fig. 10b, we find that the sentiment of the cleanliness and environment theme for a given accommodation dropped rapidly from October to November 2019 (please check the specific sentiment value of each theme in the tooltip), which we think is abnormal. We then immediately search for relevant feedback comments within this time period. Figure 10b shows the relevant feedback. From the feedback on both topics, we find that during this period of time, the road construction around the accommodation resulted in poor cleanliness and noise. Our system captures this particular abnormal sentiment change. Through the analysis of related visual components, the factors behind the incident are finally discovered. This provides useful insights for accommodation management and timely improvement of related services. The above two cases demonstrate the effectiveness of our system on sentiment anomaly detection.

6.3 Spatiotemporal analysis of topic and sentiment

The spatiotemporal analysis of topic and sentiment of the accommodation is meaningful to the manager. Assuming that you need to choose a region for rent, so how to highlight the advantages of housing and build your own brand is very important. In our case, we conduct a spatiotemporal comparative analysis of the accommodations around East China Normal University (school) and the accommodations around Disney land (scenic area). As shown in Fig. 12, the heatmap distribution (macro level) shows that the competition around Disney is more intense. We analyze the popular listings in the two regions (ranked by the number of reviews). From the treemap diagram, it can be observed that the consumers of the accommodations around East China Normal University mainly focus on the location, environment and feelings, while the consumers of the accommodations around Disney mainly focus on the location and services (such as pick-up service, etc.). We think these are caused by the different purpose of users' accommodation.

As shown in Fig. 13, we see that the accommodations around East China Normal University have higher sentiment in April, July, August and September. We think that April may be the annual re-examination of graduate students who need to rent houses, while July, August and September may be related to the peak season tourism in Changfeng Park and the opening of schools every year. At the same time, Disney land's sentiment river shows cyclical changes. The sentiment value is high in July, August and January each year. We believe that winter and summer vacations are peak seasons for tourism, and the increase in the number of tourists makes house rentals increase. Through the above case analysis, we learn that our system can compare and analyze consumer concerns in different regions, which can help accommodations in different regions to highlight their advantages better. At the same time, periodic sentiment changes can also help managers adjust their rental time.

7 User study

7.1 Participants and procedure

We invite 16 volunteers (half of males and females) with normal or corrected-to-normal vision to participate in our experiment. In particular, three of them are managers with rich experience in accommodation management and thirteen are travel enthusiasts who like to experience different accommodations. OpinionManager system was built on a Java web server (hardware required with 3.40 GHz Intel Core i7 CPU and 32GB memory) for structuring the dataset, visualizing the interface and supporting the interaction. Participants took tasks on a computer with the Chrome browser (no special hardware required but screen display required with 1920 × 1080 pixel resolution).

The study conducts with three sessions. First, we give all participants a short tutorial about the problem domain and the usage of our system. Second, participants used our system to complete the relevant analysis tasks listed in Sect. 5.1. Third, participants took a questionnaire survey. The survey contained five questions, as shown in Table 3. Participants answer with seven-point Likert-scale type, which 7 means "strongly agree" with a question, 1 means "strongly disagree" with it, and 4 indicates "neutral." Moreover, the participants had to comment what parts of the tool they liked or disliked.
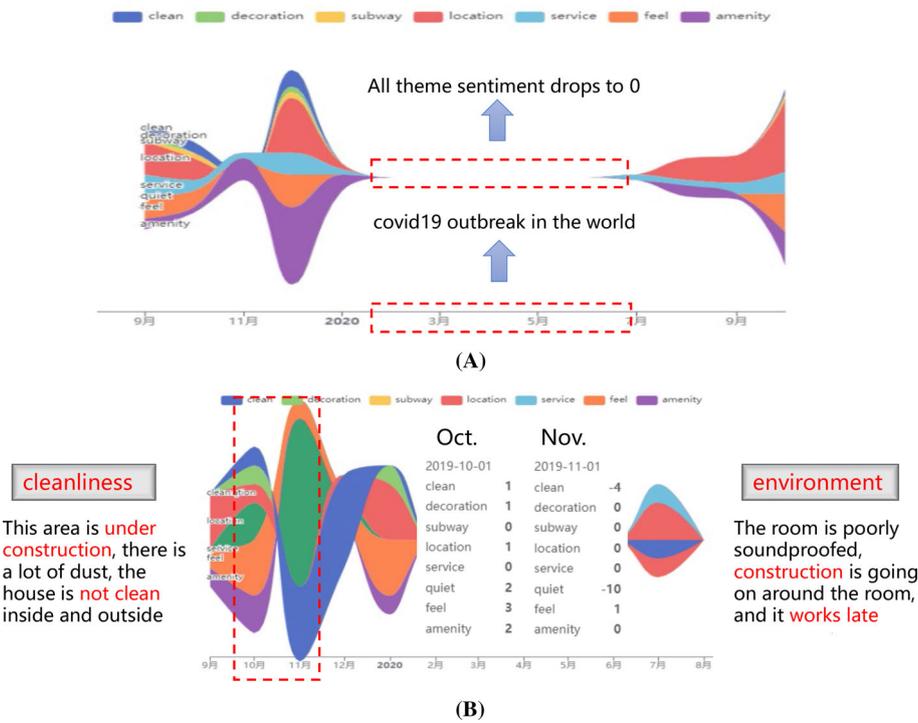
Fig. 10 OpinionManager detects sentiment abnormalities. **a** System captures sentiment disappearance at a special time. **b** System detects abrupt changes in sentiment and finds out the driven reason
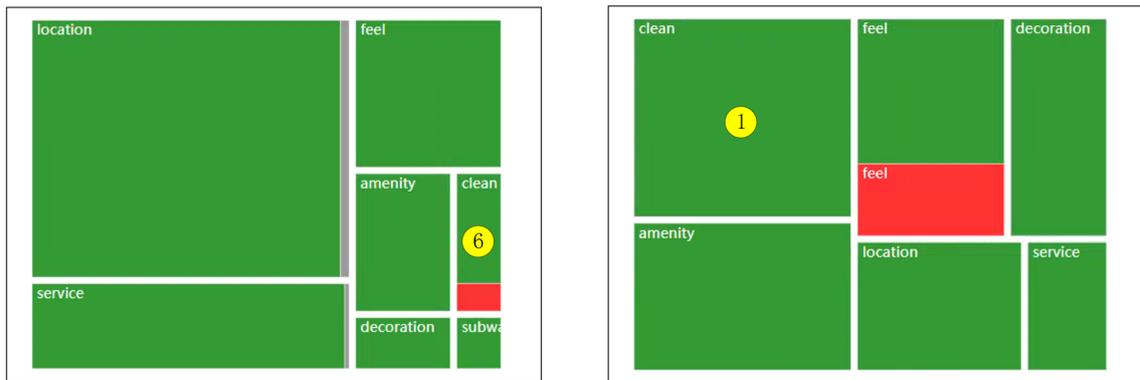
Fig. 11 Comparison of the proportion of topics that consumers care about before and after the epidemic

7.2 Survey result

We quantitatively evaluated the performance of the system OpinionManager. Figure 14 summarizes the results from our experiments. It shows the average ratings for almost all questions are 5.0 or above, which indicates that OpinionManager system meets the expectations of most participants.

For FQ2 and FQ3 (competition analysis), the average scores are 6.6 and 6.3, which are the top two scores among all questions. This shows that our system provides very satisfactory features for users. For example, P1 stated that "*Sometimes it is difficult for me to choose a suitable accommodation based on reviews or overall ratings, because I have to compare various aspects of different accommodations.*" Before using the system, participants find it difficult to know how the accommodation is rated in all aspects. Conversely, after using our system, they definitely find it very helpful, as they could quickly learn how the accommodation is rated in various aspects, as well as the sentiment value. They can also quantitatively compare among different accommodations to facilitate decision making. However, some participants gave low scores, and
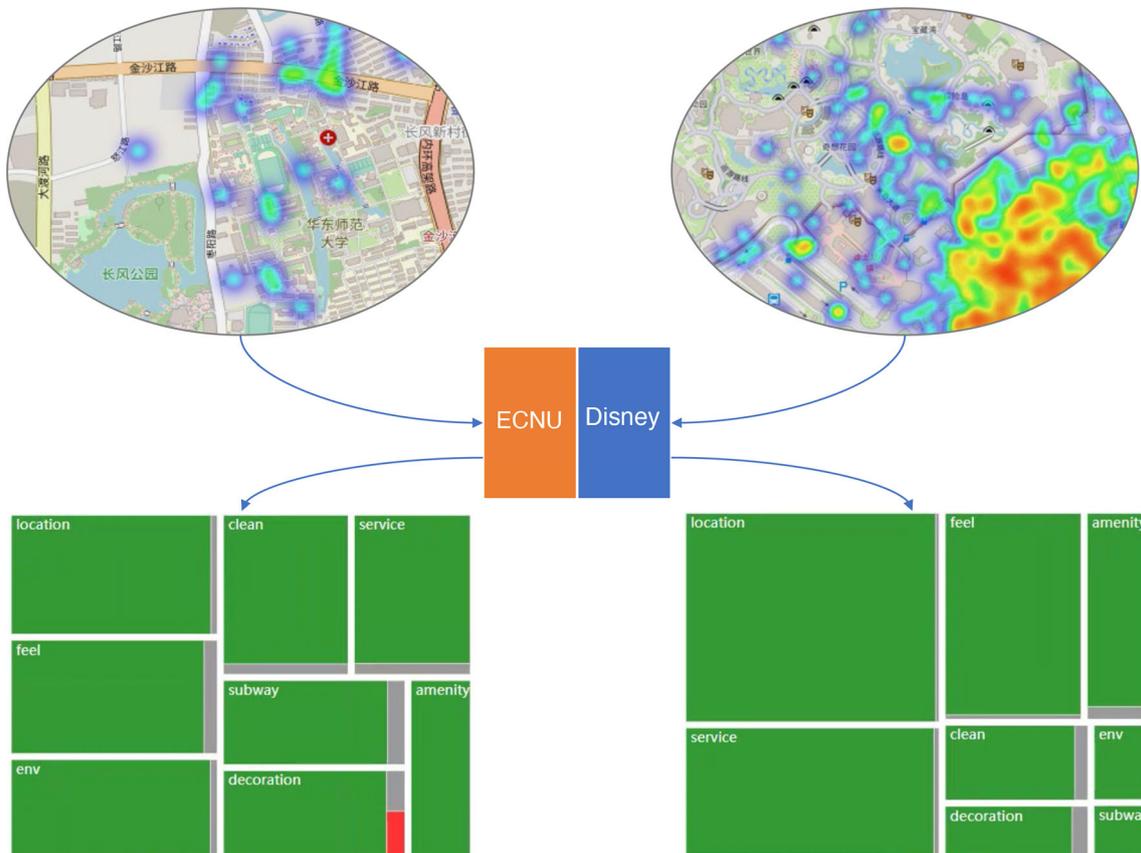
Fig. 12 Comparison of topic sentiment of accommodations in East China Normal University and Disney land. By comparing the proportions of topics in different regions, we acquire the main concerns of consumers in different regions

Fig. 13 Time-series sentiment changes of accommodations in in East China Normal University and Disney land. By comparing the temporal sentiment of different regions, we acquire the active period of consumers in different regions

Table 3 Feedback survey for our System

	Questions
FQ1	Users can well discover the geographical distribution of the listings.
FQ2	Users can easily identify the opinions contained in the review.
FQ3	Users can easily compare the strengths and weaknesses of the listings.
FQ4	Users can identify why sentiment appears anomalous in the time series.
FQ5	The system is easy to learn and use.

Table 4 User evaluation questionnaire for recommendation effectiveness

FQ6	At least one satisfactory accommodation is included in the recommendations.
FQ7	Top-K recommendations include multiple satisfactory accommodations.
FQ8	Top-K recommendation results show that the more you like the accommodation, the higher the ranking is.

after the test was completed, we learned from them that our algorithm could not extract the views well in the case of a low number of comments in a particular accommodation, which lead to the lack of visual elements in the design of the presentation.

For FQ1 (spatial exploration), the system also receive a high score. Participants can easily observe the geographical distribution of accommodations. With the filtering of accommodation types, they understand how many competitors are in the surrounding area. After the test was completed, we interviewed 4 volunteers who gave low scores on this question to improve our system. For example, P2 stated that "*the graphics on the map are too single. In order to better understand potential competitors, different types of glyphs should be drawn on the map, such as using triangles or squares to represent different room types.*" He thinks this design will be more effective.

For FQ4 (identify the reason of abnormal sentiment change), the average rating (μ) shows that the system does not satisfy users well. Participants indicate that discovering the drivers of sentiment change is not convenient enough. Because multiple factors exist and cannot be captured from the comments. For example, P3 stated that "*the accommodation happened to be under renovation at a certain time and is not released for rent*" . If our system encounters this situation, it performs poorly.

In the FQ5 (Usability), it is including two parts, easy to learn and easy to use. The average ratings of these two parts are 6.2. Most participants report that our system is easy to learn and use, since the interaction in our visual design is user-friendly and almost based on general visual techniques.

In addition, we conducted another questionnaire on volunteers who participated in the use of the system, which was an evaluation of the effectiveness of the system's recommendation algorithm. We let the participants choose an area in different areas of Shanghai to choose the accommodation. Each participant uses our system to set priorities according to their own requirements, and then according to the Top-K results recommended by the system, check the accommodation with the corresponding ID on the Aibrnb website and complete our question Table 4 to verify the validity of our recommendation. We let users carry out the Top-5 and Top-10 recommendation experiments respectively. According to the feedback from users, it can be seen from Fig. 15 the scores of the three questions are all around 6 points, which shows that the recommendation effect of the system is satisfactory. Users can basically find satisfactory accommodation through our ranking results.

8 Discussion

The case study and user study demonstrate the usability and effectiveness of our recommendation algorithm and the visualization system. However, in the process of case studies, there still exists limitations for OpinionManager system usage.

For the perspective of algorithm, the best topic model for consumer reviews may change. The number of topics should change as the review data changes, and an adaptive model should be designed. Deep learning
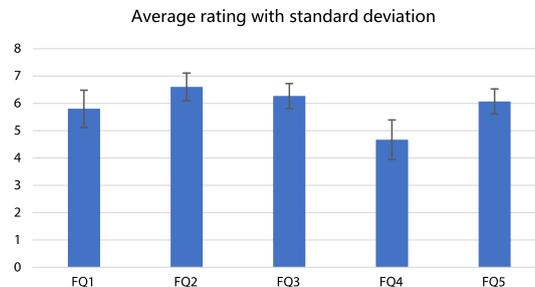
Fig. 14 Results of feedback survey

Fig. 15 User evaluation of recommendation result effectiveness

model training and sentiment calculation takes a lot of time, which is acceptable for offline operations because they can be stored in the Mysql database. But if there are scenarios that require real-time calculation, then the algorithm needs to be optimized in terms of time performance.

For the perspective of visual analysis, the design of the sentiment theme river is too simple, and requires human observation of abnormal changes. There is no automatic detection and marking mechanism. In addition, the number of indicators in the ranking view is currently fixed, and it should be adaptively changed as the number of model topics changes in the future.

9 Conclusion and future work

In this paper, we propose OpinionManager, an interactive visual analysis system to recommended p2p accommodations. This study has three important contributions. First, we propose an approach that combines topic modeling and deep learning for review aspect sentiment analysis. Second, our pipeline is equally applicable to other data, such as industry data and service product data. This provides support for improving services or aiding decision making. Third, our system employs a variety of well-designed visualization and interaction techniques.

In future work, we will further optimize the visual design and enhance the scalability of the system, and consider the real factors in the recommendation algorithm for weighting. Moreover, we aim to deploy OpinionManager to platforms to make it available to relevant end users.

References

Agarwal B, Mittal N, Bansal P, Garg S (2015) Sentiment analysis using common-sense and context information. Comput Intel Neurosci

Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. J Mach Learn Res 3:993–1022

Dai AM, Le QV (2015) Semi-supervised sequence learning. Adv Neural Inf Process Syst 28

Ding K, Choo WC, Ng KY, Ng SI (2020) Employing structural topic modelling to explore perceived service quality attributes in airbnb accommodation. Int J Hospital Manag 91:102676

Godakandage M, Thelijjagoda S (2020) Aspect based sentiment oriented hotel recommendation model exploiting user preference learning. In: 2020 IEEE 15th international conference on industrial and information systems (ICIIS), IEEE, pp 409–414

Gratzl S, Lex A, Gehlenborg N, Pfister H, Streit M (2013) Lineup: visual analysis of multi-attribute rankings. IEEE Trans Visual Comput Graph 19(12):2277–2286

Hoffman M, Bach FR, Blei DM (2010) Online learning for latent dirichlet allocation. In: Advances in neural information processing systems, Citeseer, pp 856–864

Hu N, Zhang T, Gao B, Bose I (2019) What do hotel customers complain about? text analysis using structural topic model. Tourism Manag 72:417–426

Jo Y, Oh AH (2011) Aspect and sentiment unification model for online review analysis. In: Proceedings of the fourth ACM international conference on web search and data mining, pp 815–824

Kenton JDMWC, Toutanova LK (2019) Bert: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL-HLT, pp 4171–4186

Kim SG, Kang J (2018) Analyzing the discriminative attributes of products using text mining focused on cosmetic reviews. Inf Process Manag 54(6):938–957

Krumm J, Davies N, Narayanaswami C (2008) User-generated content. IEEE Pervasive Comput 7(4):10–11

Lerttripinyo T, Jatukannyaprateep P, Prompoon N, Pattanothai C (2015) Accommodation recommendation system from user reviews based on feature-based weighted non-negative matrix factorization method. In: 2015 12th international joint conference on computer science and software engineering (JCSSE), IEEE, pp 108–113

Li D, Yin H, Wang C, Song S, Li K, Li C (2021) Opinionmanager: Visual exploration of online reviews in p2p accommodation. In: The 14th international symposium on visual information communication and interaction, Association for Computing Machinery, New York, NY, USA, VINCI 2021. https://doi.org/10.1145/3481549.3481551, https://doi.org/10.1145/3481549.3481551

Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. Adv Neural Inf Process Syst 26

Minh DL, Sadeghi-Niaraki A, Huy HD, Min K, Moon H (2018) Deep learning approach for short-term stock trends prediction based on two-stream gated recurrent unit network. IEEE Access 6:55392–55404

Mousa A, Schuller B (2017) Contextual bidirectional long short-term memory recurrent neural network language models: a generative approach to sentiment analysis. In: Proceedings of the 15th conference of the European chapter of the association for computational linguistics, Vol 1, Long Papers, pp 1023–1032

Oelke D, Hao M, Rohrdantz C, Keim DA, Dayal U, Haug LE, Janetzko H (2009) Visual opinion analysis of customer feedback data. In: 2009 IEEE symposium on visual analytics science and technology, IEEE, pp 187–194

Padma P, Ahn J (2020) Guest satisfaction and dissatisfaction in luxury hotels: an application of big data. Int J Hospital Manag 84:102318

Peng L, Cui G, Zhuang M, Li C (2014) What do seller manipulations of online product reviews mean to consumers?

Prameswari P, Surjandari I, Laoh E, et al. (2017) Mining online reviews in indonesia's priority tourist destinations using sentiment analysis and text summarization approach. In: 2017 IEEE 8th international conference on awareness science and technology (iCAST), IEEE, pp 121–126

Puschmann T, Alt R (2016) Sharing economy. Business Inf Syst Eng 58(1):93–99

Shi HX, Li XJ (2011) A sentiment analysis model for hotel reviews based on supervised learning. In: 2011 international conference on machine learning and cybernetics, IEEE, Vol 3, pp 950–954

Sievert C, Shirley K (2014) Ldavis: A method for visualizing and interpreting topics. In: Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, pp 63–70

Sievert C, Shirley K (2014) Ldavis: a method for visualizing and interpreting topics. In: Proceedings of the workshop on interactive language learning, visualization, and interfaces, pp 63–70

Wu Y, Wei F, Liu S, Au N, Cui W, Zhou H, Qu H (2010) Opinionseer: interactive visualization of hotel customer feedback. IEEE Trans Visual Comput Graph 16(6):1109–1118