Research Article

Research on Design Strategy of B&B: Based on Text Mining and Machine Learning Method

Hongzhou Pan\textsuperscript{1} and Hongmin Shen\textsuperscript{2}

\textsuperscript{1}College of Humanities, Xiamen Huaxia University, Xiamen 361024, China
\textsuperscript{2}Xiamen University Tan Kah Kee College, Zhangzhou 363105, China

Correspondence should be addressed to Hongmin Shen; hmshen@xujc.com

Received 30 June 2022; Revised 25 July 2022; Accepted 2 August 2022; Published 5 September 2022

A bed and breakfast (typically shortened to B&B) is a small lodging establishment that offers overnight accommodation and breakfast. At present, the design strategy of B&B is generally based on the personal views of B&B operators or interior designers, rather than the actual market data. Thus, the determination of design strategy is always divorced from the actual sales data of B&B. In this study, the optimal design strategy of B&B will be figured out based on the analysis of the comments data and operation data of B&B with methods of text mining and machine learning. The B&B design strategy based on text mining and machine learning would be more consistent with the consumer needs of users and business needs of B&B operators. It is a future evolution method of design to formulate design strategies for B&B based on big data and artificial intelligence.

1. Introduction

A B&B is often private family home and typically has between four and eleven rooms, with six being the average. In addition, a B&B usually has the hosts living in the house. At present, the design strategy of B&B is generally based on the subjective conception of B&B operators or interior designers, instead of objective market data. Thus, the determination of design strategy is always divorced from the actual market data. In this study, the optimal design strategy of B&B will be figured out based on the analysis of the comments data and operation data of B&B with methods of text mining and machine learning. The B&B design strategy based on text mining and machine learning would be more consistent with the consumer needs of users and business needs of B&B operators. It is a future evolution method of design to formulate design strategies for B&B based on big data and artificial intelligence.

First, analyze the users’ comments data by text mining technology starting from users’ interests and obtain the design strategy that best meets the needs of B&B users.

Second, analyze sales data of Beijing Airbnb using the regression model based on the machine learning algorithm, establish the relationship between the design strategy of the B&B and the profitability of the B&B starting from hosts’ interests, and finally determine the optimization design strategy of the B&B design.

So, the B&B design strategy based on text mining and machine learning is in accordance with the consumer needs of users and business needs of hosts.

2. Related Works

Nowadays, research studies on B&B mainly include price prediction, pricing determinants, user satisfaction and dissatisfaction, operation, spatial distribution, tourist perception evaluation, and emotional design in the design of B&B and few of them involve the design strategies of B&B based on machine learning.
Kalehbasti et al. created a model for predicting the price of an Airbnb listing using property specifications, owner information, and customer reviews for the listing [1]. Hong and Yoo explored the spatially heterogeneous relationship between price and pricing variables using an innovative spatial approach, multiscale geographically weighted regression [2]. Voltes-Dorta et al. presented a study about the drivers of Airbnb prices in Bristol using ordinary least squares and geographically weighted regression methods [3].

Chattopadhyay and Mitra used a dataset of Airbnb accommodation listings for Toronto; the study established a relationship between room pricing and various listing variables and identified a reduced number of listing attributes that influence the room price significantly [4].

Ding et al. examined key attributes affecting Airbnb users' satisfaction and dissatisfaction through the analysis of online reviews [5].

Chung and Sarnikar wrote a paper to understand the use of listing descriptions by Airbnb hosts and the impact of such descriptions on sales performance [6].

Sun et al. studied the factors affecting the spatial distribution of Airbnb in Nanjing of China by using the zero-expansion negative binomial regression [7].

Hu et al. used the network evaluation data collected by Ctrip as a sample; based on the theory of tourism image perception, the content analysis method is used to extract the high frequency feature words of Shanghai B&B tourist image perception and to explore the characteristics, perceived image and emotional image of Shanghai B&B tourists [8].

Shen and He made a series of analysis on the physiological and emotional needs of passengers during their accommodation by analysis of emotion, word frequency, and content of network evaluation, demonstrating the importance of emotional design in the design of B&B [9].

3. Research on Design Strategy of B&B Based on Text Mining

Comments dataset of Beijing Airbnb will be analyzed with the method of text mining in this section, and the key factors that need to be considered in the design strategy of B&B of Beijing will be found out.

3.1. Data Sources. Beijing Airbnb comments dataset is available for download from website of Inside Airbnb with address https://insideairbnb.com/beijing/. The name of the comments dataset is “reviews.csv.” This dataset has 22796 rows and 6 columns and its sample contents are given in Table 1.

3.2. Analysis Process of B&B Design Strategy Based on Text Mining. User comments data in Table 1 are analyzed in this study with word segmentation and word frequency statistics with the aid of jieba library which is a natural language processing library based on TF-IDF and HMM models. TF-IDF is the product of the frequency of high words in a particular document and the frequency of that word in reverse files in the entire document set. TF-IDF is inclined to filter out common words and retain important words [10]. TF-IDF formula is

$$TF - IDF = TF \times IDF,$$

where TF is the term frequency, which means the frequency of entry appearing in the text, and IDF is the inverse document frequency.

The HMM model can be determined by hidden state initial probability distribution $\Pi$, state transition probability matrix $A$, and observation state probability matrix $B$. $\Pi$ and $A$ determine the sequence of states, and $B$ determines the sequence of observations. Therefore, the HMM model can be represented by triples as follows:

$$\lambda = (A, B, \Pi).$$

The detailed analysis process is as follows:

1. Conduct words segmentation for the field named “comments” in the review dataset (Table 1) with the aid of jieba library based on TF-IDF and HMM models
2. Make word frequency statistics of the words in users' comments according to the result of words segmentation with the aid of jieba library
3. Select top 60 design keywords related to the design strategy of B&B by the interior designer team according to word frequency ranking
4. Create word cloud diagram using WordCloud library for the design keywords selected by the interior designer team according to the word frequency statistics
5. Make the word frequency histogram using Matplotlib library for the design keywords selected by the interior designer team
6. Calculate the word frequency percent of the design keywords for the keywords selected by the interior designer team according to the number of word frequency statistics and total number of comments and make the histogram of the word frequency percent of the design keywords using the Matplotlib library

3.3. Analysis Results of B&B Design Strategy Based on Text Mining. On the ground of the analysis of B&B design strategy based on text mining and 21,575 Chinese comments from reviews.csv of Beijing Airbnb comments dataset in Table 1, the author wrote the python code with the aid of jieba, WordCloud, Matplotlib, RE, Pandas, and collections library. As the running results of the code, the cloud map of the design keywords, the histogram of the word frequency statistics of the design keywords, and the word frequency percent of the design keywords are obtained, as shown in Figures 1-3.
3.4. Research on Design Strategy of B&B of Beijing Based on Text Mining. Combined with the B&B design strategy analysis based on text mining, the author studies the design strategy, and the results are as follows:

(1) First, it can be seen from the histogram of the word frequency statistics of the design keywords (Figure 2) that “room” ranks first, and the word “room” accounts for 40% in the histogram of the word frequency percent of design keywords (Figure 3). Thus, “room” is the first design element as well as the core factor in the functional space requirement of the design strategy of B&B of Beijing. The other functional requirements element words in turn are, respectively, “courtyard,” “children,” “kitchen,” “toilet,” “sitting room,” “terrace,” and “the restaurant,” reflecting that the consumer demand for B&B facility design is not only confined to the basic function of sleep and rest but also include family life functions as well as the children's space requirements. “Courtyard” and “terrace” used in recreational function space are high-ranking in word frequency. Especially, the word of “courtyard” is ranked second in Figure 2, accounting for 12% in Figure 3, only second to “room,” which also reflects customer’s increased demand for leisure space in B&B design with the improvement of life quality.

(2) Second, in high frequency words of customer’s comprehensive perception experience of B&B space, words ranked by frequency are “clean,” “comfortable,” “tidy,” “quiet,” “very big,” and “Soundproof,” in which the word of “clean” in the histogram of design keywords percent accounts for 36%, ranking first among design key words related to comprehensive perception experience. This indicates that customers have the highest requirements for the cleanliness and tidiness of the B&B rooms, followed by the comfort and sound insulation conditions of the B&B rooms, and the size of the B&B rooms.

(3) Third, the subjective evaluation of customer’s aesthetic demand for B&B appears in the high frequency words, such as “decoration,” “sweet,” “style,” and “beautiful,” which shows customer’s subjective feelings and aesthetic preferences of B&B design style. Among the words related to aesthetic evaluation of B&B, the word “sweet” ranks first, accounting for 7% in the histogram of word frequency percentage, which also represents customers’ subjective needs for the design style of B&B.

4. Research on B&B Design Strategy Based on Machine Learning

This part first analyzes the Airbnb sales dataset of Beijing by establishing linear regression and Lasso regression models between design element features and B&B values (B&B value = the price of B&B × the sales volume of B&B) and judges which design elements should be considered in the design strategy of Beijing B&B by the variable weight of the model. Then, it studies the optimal optimization design strategy in B&B design according to the design element features.

4.1. Data Sources and Preprocessing. Airbnb sales dataset of Beijing is downloaded from a website named Inside Airbnb at https://insideairbnb.com/beijing/. The dataset name is listings.csv. This dataset has 4483 rows and 74 columns. After eliminating null rows and selecting four columns of data, the dataset has 4449 rows and 4 columns. The dataset is given in Table 2.

4.2. Feature Extraction and Selection. Through the segmentation, feature word extraction, manual analysis, selection, and merging of the string data in the column named “amenities” of the dataset in Table 2, the authors obtain 66 feature words related to B&B design, as given in Table 3. According to the above design element features, the method of feature encoding is adopted to establish a dataset of design element features of B&B containing 4449 rows and
67 columns as machine learn training data, which contains 4449 rows and 67 columns, among which the first three rows and nine columns are given in Table 4.

The data field named “total price” in the above table is the output vector, the B&B value. These data are calculated by multiplying the two columns of data named “price” and “number_of_reviews_ltm” in Table 2. Since there are no sale volume data of B&B in the dataset, the authors use the column named “number_of_reviews_ltm” instead, which is the number of comments per month.
In order to obtain the linear regression coefficient \( \theta \), the authors need to define a loss function, an optimization method to minimize the loss function, and a method to verify the algorithm. Different loss functions, different optimization methods of loss functions, and different verification methods form different linear regression algorithms. This paper uses the following two linear regression models in Scikit-learn library:

(1) Linear regression: the most commonly used linear regression model has the following loss function:

\[
J(\theta) = \frac{1}{2m}(X \theta - Y)^T (X \theta - Y)
\]  

(4)

(2) Lasso regression: For high-dimensional feature data, especially the linear relationship is sparse, Lasso regression is effective. If want to find out the main features in a bunch of features, Lasso regression is the first choice. Because of the high dimension of training data of design strategy, the authors chose Lasso regression as one of the training models. Lasso regression can reduce the coefficients of some features, and even make some coefficients with small absolute values directly become 0, which enhances the generalization ability of the model.

The loss function of the model is as follows:

\[
J(\theta) = \frac{1}{2m}(X \theta - Y)^T (X \theta - Y) + \alpha \| \theta \|_1
\]

(5)

### 4.3. Establishment and Evaluation of the B&B Design Strategy Model Based on Machine Learning

#### 4.3.1. Modeling. This model is implemented based on the linear regression model of python machine learning library Scikit-learn. The purpose of linear regression is to obtain the linear relationship between output vector \( Y \) (field named “total price” in training dataset) and input feature \( X \) (66 design elements in training dataset) and to find the linear regression coefficient \( \theta \). The function is

\[
Y = \theta \times X.
\]

(3)

In the function, the dimension of \( Y \) is \( m \times 1 \), the dimension of \( X \) is \( m \times n \), and the dimension of \( \theta \) is \( n \times 1 \). \( m \) represents the number of samples, and \( n \) represents the dimension of sample features.

In order to obtain the linear regression coefficient \( \theta \), the authors need to define a loss function, an optimization method to minimize the loss function, and a method to verify the algorithm. Different loss functions, different optimization methods of loss functions, and different verification methods form different linear regression algorithms. This paper uses the following two linear regression models in Scikit-learn library:

(1) Linear regression: the most commonly used linear regression model has the following loss function:

\[
J(\theta) = \frac{1}{2m}(X \theta - Y)^T (X \theta - Y)
\]

(4)

(2) Lasso regression: For high-dimensional feature data, especially the linear relationship is sparse, Lasso regression is effective. If want to find out the main features in a bunch of features, Lasso regression is the first choice. Because of the high dimension of training data of design strategy, the authors chose Lasso regression as one of the training models. Lasso regression can reduce the coefficients of some features, and even make some coefficients with small absolute values directly become 0, which enhances the generalization ability of the model.

The loss function of the model is as follows:

\[
J(\theta) = \frac{1}{2m}(X \theta - Y)^T (X \theta - Y) + \alpha \| \theta \|_1
\]

(5)

#### 4.3.2. Training and Evaluation of Models. Input features data (design elements) and output vector data (B&B value field named “total price”) of Table 4 are fitted into the linear regression model of Scikit-learn for training, and then, a linear regression model is obtained. The correlation coefficient \( \theta \) of various design elements of the model is shown in Figure 4.

Taking 4449 rows of training data in Table 4 as the input data of the linear regression model equation, the predicted data are calculated and compared with the actual data, and the model evaluation chart is obtained as shown in Figure 5. It can be seen that the predicted value of the model (blue line in the figure) has a high degree of fitting with the actual value (pink line in the figure), and the accuracy of the model can meet the requirements of design strategy research.

The input features data (design elements) and output vector data (B&B value field named “total price”) of Table 4 are fitted into the Lasso regression model of Scikit-learn for training, and then, a Lasso regression model is obtained. The correlation coefficient \( \theta \) of various design elements of the model is shown in Figure 6. It can be seen that this model directly changes some coefficients with smaller absolute values to 0, enhancing the generalization ability of the regression model.
The comparison line chart between the predicted value and the actual value of the Lasso regression model is shown in Figure 7. It can be seen that the predicted value (blue line in the figure) has a high degree of fitting with the actual value (pink line in the figure) of the model, which is basically consistent with the fitting degree of the linear regression model above. These two models jointly verify that the machine learning model in this study meets the requirements of B&B design strategy.

4.4. Research on B&B Design Strategy Based on the Machine Learning Model. According to the ranking of feature importance of linear regression and Lasso regression models in Figures 4 and 6, the contribution of each feature variable to the model can be judged, that is, which design elements are more important for B&B design can be determined. Among the total 66 design elements, the top 30 design elements with the largest weights are given in Table 5.

According to the design elements’ weights of the linear regression models and Lasso regression models, the Nightingale rose diagram of the top 30 design elements is shown in Figure 8.

This study makes further analysis and research from the perspective of design strategy and proposes the design strategy of B&B based on two regression models.

(1) In Figure 8, entertainment facilities have the highest design weight, especially piano and game console: PS4 contribute the most to the value of Beijing B&B, followed by the Bluetooth sound system, board games, and pool table. It can be seen that strengthening the design of entertainment space and entertainment facilities in the design of B&B can...
significantly improve the profitability of B&B and realize the differentiation in the design of B&B.

(2) The design elements of leisure facilities contribute greatly to the value of Beijing B&B. The leisure design elements ranked by weight are as follows: private patio or balcony, shared garden or backyard, waterfront, TV, dining table, stove, BBQ grill, and coffee maker. Among the top 30 design elements, there are 8 leisure elements, accounting for 27%. Among them, private patio or balcony and shared garden or backyard have the greatest weight, which indicate that patio and courtyard are the key...
Figure 8: Nightingale rose diagram of top 30 design elements.

Functional spaces in the design of B&B space. That is to say, in the design strategy of Beijing B&B, patios and balcony space and outdoor courtyard space should be set up, and corresponding leisure facilities should be equipped, such as BBQ grill and coffee maker.

3. The design elements of comfort facilities are central heating, bathroom, bedroom comforts, free street parking, dishwasher, bathtub, bed linens, and indoor fireplace which in turn indicates that in the design strategy of Beijing B&B, the interior of the room requires a warm environment, furnished with central heating or indoor fireplace, and comfortable bedding for the bedroom. Bathtub should be considered in the design of toilet, dishwasher should be in the design of kitchen, and the area adjacent to free street parking should be selected as far as possible in the site selection of Beijing B&B.

4. The design elements of children’s facilities are baby safety gates, baby bath, table corn guards, and high chair in sequence, which show that the design of children’s space has become the key design factor of B&B. The design strategy of Beijing B&B should not only meet the basic spatial functions of children’s needs but also provide daily living facilities for children.

5. The design elements of safety facilities are smart lock and lockbox, which indicates that smart lock and lockbox that are conducive to enhancing the value of B&B and the safety of residents are indispensable in the design of B&B

5. Discussion

In the research of B&B design strategy based on text mining, according to the 21,575 rows of user reviews data of Airbnb of Beijing, word segmentation and word frequency statistics are carried out on the user review data of Airbnb of Beijing. Then, the word cloud map, word frequency statistical histogram, and word frequency statistical ratio histogram of design keywords are obtained based on jieba library which is a natural language processing library based on TF-IDF and HMM models. Later, on the basis of these three graphs, the design strategy of Beijing B&B is conceived.

In the research of B&B design strategy based on the machine learning model, according to 4449 rows of operation data of Airbnb of Beijing, the regression model between the design elements and the values of B&B listings is established in virtue of two machine learning models: linear regression and Lasso regression. The machine learning model is trained using the training dataset created by authors to obtain two regression models. According to design element weight of the regression models, the design strategies are formulated, including coefficients weight histogram in Figure 4, coefficients weight histogram in Figure 6, and Nightingale rose diagram of the top 30 design elements in Figure 8. Finally, according to the relevant graphics of the design elements, the optimal design strategy of B&B of Beijing based on the machine learning model is conceived.

This B&B design method based on text mining and machine learning method has the following advantages:

Advantage 1: Different from traditional design strategies, which are formulated by the personal opinions of designers, the design strategy of this study fully refers to volume, variety, and velocity’s business data and users’ living experience data and is an optimal design strategy based on big data.

Advantage 2: The design strategy of this study is based on artificial intelligence technology of the computer. Artificial intelligence technology can efficiently and accurately summarize the design strategies implied in the big data of B&B operation and users’ living experience. Traditional methods of interior appear to be inefficient and inaccurate as it is impossible to obtain data results solely by comprehensive analysis of human resources faced with such a huge amount of business data.

Advantage 3: The design method in this study can improve designers’ work efficiency and optimizes their work focus. Since the most critical design elements are mined from a large amount of data through text mining and machine learning, the designers only need to formulate the design strategy of B&B in accordance with key design elements instead of analyzing the huge business data. In this way, designers can be freed from data work and focus more on interior designing work.
Advantage 4: This design method of this study helps to unify the opinions of B&B hosts and designers. Due to the information asymmetry between hosts and designers, Party A (B&B host) and Party B (interior designer) often disagree about the design strategy in the traditional interior design scheme formulation process. The design strategy of this study is calculated by artificial intelligence based on large and unified data of B&B operation, which is helpful for hosts and designers to unify their designing opinions.

In the era of big data and artificial intelligence, it is a future trend of design methods and a future evolution method of design to formulate design strategies for B&B based on big data and artificial intelligence.

6. Conclusion

In this study, the design strategy of B&B is studied with two data mining methods of text mining and machine learning based on two datasets of the consumers and operators, with both two professional directions of art and technology.

According to the above two methods, the authors have studied the optimization design strategy of Beijing B&B and summarized as follows.

First, optimize the spatial function layout. Increase the space allocation such as entertainment space and parent-child activity space and try to set up patio and balcony space. In basic space functions, outdoor courtyard space and other leisure space should be set up.

Second, improve the physical environment. In the interior design of B&B, the application of interface sound insulation materials can control the generation and spread of noise sources, which improves the sound environment, provides good sound insulation conditions, and protects the customer privacy. Geothermal, central air conditioning, humidification or dehumidifier, and other equipment are applied in the design to create a good temperature and humidity environment and improve human comfort.

Third, upgrade facilities. Add entertainment and leisure facilities, children’s daily living facilities, safety facilities, and intelligent facilities.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was funded by the project “Curriculum reform of living space design guided by building a green living environment” (HXJG2022213) of 2022 Undergraduate Education and Teaching Reform Research Projects of Xiamen Huaxia University.

References