LDA Topic Mining of Light Food Customer Reviews on the Meituan Platform

Miaojia Huang[1], Songqiao Wen[2], Manhua Jiang[2](✉), and Yuliang Yao[2](✉)

[1] College of Management, Shenzhen University, Shenzhen, People's Republic of China
2016040301@email.szu.edu.cn

[2] College of Mathematics and Statistics, Shenzhen University, Shenzhen, People's Republic of China

Abstract. Light food refers to healthy and nutritious food that has the characteristics of low calorie, low fat, and high fiber. Light food has been favored by the public, especially by the young generation in recent years. Moreover, affected by the COVID-19 epidemic, consumers' awareness of a healthy diet has been improved to a certain extent. As both take-out and in-place orders for light food are growing rapidly, there are massive customer reviews left on the Meituan platform. However, massive, multi-dimensional unstructured data has not yet been fully explored. This research aims to explore the customers' focal points and sentiment polarity of the overall comments and to investigate whether there exist differences of these two aspects before and after the COVID-19. A total of 6968 light food customer reviews on the Meituan platform were crawled and finally used for data analysis. This research first conducted the fine-grained sentiment analysis and classification of the light food customer reviews via the SnowNLP technique. In addition, LDA topic modeling was used to analyze positive and negative topics of customer reviews. The experimental results were visualized and the research showed that the SnowNLP technique and LDA topic modeling achieve high performance in extracting the customers' sentiments and focal points, which provides theoretical and data support for light food businesses to improve customer service. This research contributes to the existing research on LDA modeling and light food customer review analysis. Several practical and feasible suggestions are further provided for managers in the light food industry.

Keywords: Light food · Review analysis · LDA topic modeling · Sentiment analysis · Customer review

1 Introduction

With the consumption upgrading and the continuous improvement of the national awareness of healthy eating, the term "light food", which features low fat, low heat, low sugar, high fiber, and high protein, has been popular with more and more consumers. According to the "China Light Food Takeout Consumption Report" released by the Meituan Takeout, the number of searches of four keywords on the Meituan APP, namely light food, fat-reducing meal, weight-loss meal, and healthy meal, increased by 235.8%, 200.6%,

Y. Tan et al. (Eds.): DMBD 2021, CCIS 1454, pp. 108–121, 2021.
https://doi.org/10.1007/978-981-16-7502-7_13

186.4%, and 116.0%, respectively, compared with that of 2018. As of September 2019, orders for the Meituan takeaway light food increased by 98% and numbers of light food stores increased by 58% year on year [1]. As the consumption of "light food" is increasing, a large number of emerging light food restaurants are also taking advantage of this trend.

As light food gets popular across the whole nation, there are many new entrants flooding the market and many losers exiting the market. The data released by Trend makers [2] shows that a total of 1,251 light food businesses had finished the logout registration by May 2019. Although this market has the features of low industry barriers, serious homogenization and fierce competitions, a small number of light food brands still have succeeded in standing out in the market and formed their characteristics with high repurchase rate, and even became daily the rigid needs among some college students and urban white-collars. Therefore, how to understand the hidden preferences of the light food consumers and help light food merchants cultivate user stickiness is a problem that every entrant in the light food industry needs to think about.

With the rapid growth of light food orders, there is a huge number of multi-dimensional customer reviews left on the Meituan platform, conveying customers' inner feelings and emotions under certain consumption situations. For potential customers, this user-generated content (UGC) plays a key role in reducing information overload and facilitating them to make consumption decisions. For light food businesses, this UGC acts as electronic word of mouth (eWOM) on the Meituan platform. However, this UGC has not yet been fully extracted and utilized and there is a lack of research on text analysis of the customer review in the light food industry. This is a pity because positive reviews left by satisfied customers were found to have significant positive impacts on product sales [3] and long-term performance [4]. Furthermore, customer negative reviews might cost a lot to businesses since these dissatisfied customers might spread negative eWOM or even switch to competitors [5]. In such cases, regaining the lost consumers via improvement is profitable to businesses instead of attracting potential consumers [6]. Therefore, investigating customers' negative reviews would help managers better understand their service design and service failure while making improvements accordingly.

Moreover, the COVID-19 brings about some changes to the light food industry. On one hand, the consumption of light food seems to increase as the public gradually enhances their health awareness in the post-epidemic era. On the other hand, as one of the industries that are hardest hit by the COVID-19 epidemic, managers in the catering industry are highly concerned about the new business operation tactics after the COVID-19 outbreak. Therefore, examining whether there is a difference in light food customers' focus points and emotional polarity based on the text mining results from their reviews and then giving better insights to light food marketers have become important and urgent.

Therefore, to address the above research gaps, this study seeks to explore the following two research questions:

RQ1: (1) What are the topics of the light food customer reviews on the Meituan platform?
RQ2: (2) Is there any change in the focus and emotion of consumers based on the comments left before and after the outbreak of the COVID-19 epidemic?

This research aims to get a more in-depth capture of the emotional experience of light food consumers and extract the main topics based on massive and scattered customer reviews, which is conducive to the merchants to accurately improve their products, services and consequently improve customer experiences. The novelty of this study is twofold. Firstly, the large dataset across 2017 to 2021 enables us to explore the influencing factors of customers' satisfaction and dissatisfaction before and after the COVID-19 epidemic via different text analysis tools, namely, frequency analysis, word cloud, and sentiment analysis. Secondly, previous studies identified customer preferences and complaints via survey data, which might fail to capture the hidden dimensions for improvement [7]. By analyzing the changing process of the focal points and emotional polarity of the comments of the light food customers before and after the epidemic, this paper provides the basis for the relevant merchants to improve the user experience and helps these small and medium-sized enterprises to better serve customers and cope with the business crisis in the post-epidemic era.

The remainder of this research is arranged as follows: Sect. 2 presents the related literature. Section 3 demonstrates the data-driven approach (i.e., LDA modeling) and the experiments using the light food consumer reviews in Meituan.com. Section 4 concludes the data analysis results and discusses some practical implications for managers in the light food industry. Section 5 summarizes the study and proposes some future directions of research.

2 Literature Review

2.1 Sentiment Analysis

Sentiment analysis methods that are widely adopted in previous research include the deep learning method based on neural network, the method based on sentiment dictionary and rules, the machine learning method based on feature extraction, and the multi-strategy hybrid method. For example, Ruz et al. [8] used the Bayesian network classifier to conduct sentiment analysis on two Spanish data sets (Chile earthquake in 2010 and Catalonia independence referendum in 2017). The results show that the Bayesian network classifier is superior to other machine learning methods (support vector machine and random forest). Gopalakrishnan et al. [9] proposed a simplified LSTM model with six different parameters to achieve sentiment analysis on the Twitter dataset of the debates of the Republican Party in America. Results reveal that different parameter settings and model layer settings would have an impact on the experimental results. Wang and Wu [10] calculated the emotional tendency and the degree of attention of the tourists' reviews on the tourist attractions in Xiamen by constructing the positive and negative emotion dictionary and the tourism image attribute glossaries. Based on the hierarchical quantile regression model, Wang and Gao [11] analyzed the emotions of the online travel reviews, investigated the main factors affecting the scenic spot evaluation, and further explored the potential correlation among these factors.

2.2 Latent Dirichlet Allocation (LDA) Topic Modeling

Customer reviews often convey a wealth of information, but as Aggarwal and Zhai [12] noted, dealing with large volumes of unstructured textual data requires an automated approach to facilitate effective analysis. Text mining techniques can draw some useful business insights from online reviews by computing features and their corresponding weights (e.g., word frequency) [13]. LDA topic modeling is the most widely used and effective method of topic extraction. For example, Yang et al., [14] completed fine-grained sentiment analysis and classification of tourist reviews through sentiment dictionary and SnowNLP technique. They further extracted the positive and negative topics in tourist comments via LDA topic modeling to better understand the underlying reasons for their sentiment polarity. Chen et al. [15] employed the LDA topic model to identify the topics of the user reviews of the health and medical wearable devices. They further conducted sentiment analysis on the user reviews under the corresponding topic to explore the relationship between the extracted topics and the users' satisfaction.

To sum up, there are few studies on the comment analysis of the light food area and limited studies employing both the SnowNLP and LDA modeling for detailed analysis of the users' sentiment and focal points. To fill the above-mentioned research gaps, this research proposes a method combining SnowNLP and LDA modeling, where the SnowNLP technique can effectively extract the customers' emotions and the LDA modeling can help extract the positive and negative topics of the comments. This will help to identify the customers' focus during the consumption of light food and provide some practical implications for the operators of light food to better improve the customer service.

3 Methodology

3.1 Naive Bayes Method

Regarding sentiment analysis, this study adopts the SnowNLP module in Python, which classifies emotions of the review text (e.g., positive and negative) based on the Naive Bayes method. The Bayesian algorithm was first proposed by Bayesian and then improved and extended by Laplace. Pang et al. [16] first applied the Naive Bayes method to sentiment analysis. The implementation of the algorithm is as follows. First, the algorithm sets the segmentation rules and stopping thesaurus. Second, the positive and negative corpora are segmented and some useless words are removed according to the stop word list and the frequency of each word segmentation is then calculated. Third, according to the frequency, the conditional probability of a word appearing in positive or negative comments as well as in the overall text can be calculated. Finally, the comment text to be classified is segmented, and the posterior probability is calculated according to the words in the text. Based on the aforementioned steps, the text can be classified into positive or negative.

The reasons for choosing this model are as follows: (1) Compared with other sentiment analysis methods, this model has stronger mobility. By replacing the corpus of related fields, the model can perform sentiment analysis on any domain. (2) There is no need to decompose the sentence components or analyze the grammatical semantics in using this model. Instead, it calculates the posterior probability after separating the words in the corpus to judge the emotion of comments.

3.2 LDA Topic Mining

LDA, as one of the most popular methods for topic modeling, is first introduced by Blei et al. [17]. The researchers assume that the topic extraction process is as follows: First, the model randomly generates the topic distribution of a specific text and randomly generates a topic according to the topic distribution at each position in the text. Second, it randomly generates a word according to the word distribution of the topic. Third, the above process is repeated continuously until words are generated at each position of the text.

LDA topic modeling is a probabilistic model based on Bayesian estimation based on this assumption. Its characteristic is to learn a given text set with Dirichlet distribution as a prior distribution of multi-nominal distribution and to solve all parameters of the model through the estimation of a posteriori probability distribution to get the topic distribution of the text.

When adopting the LDA method, the number of topics, the hyper-parameter, and Dirichlet distribution need to be determined first. Figure 1 presents the plate notation of the LDA model, in which the solid nodes represent observed variables, hollow nodes represent hidden variables, directed edges represent probabilistic dependency relations, the rectangle represents repeated operations, and the numbers inside the rectangle represent the number of repeats. The four steps of the LDA algorithm for text generation is demonstrated as the figure shows:

(1) Randomly generates a topic distribution θ_i of the given text i from the Dirichlet distribution α;
(2) Randomly generates the topic $Z_{i,j}$ for the ith word of the given text i from the topic polynomial distribution θ_i;
(3) Randomly generates the corresponding word distribution $\varphi_{i,j}$ of each topic $Z_{i,j}$ from the Dirichlet distribution β;
(4) Repeated the above three steps and finally generates the word $\omega_{i,j}$ from the word polynomial distribution $\varphi_{i,j}$.

The reasons for choosing this model in this study are as follows: (1) It only needs to train the sample data without manually annotating the data. (2) It has a stronger generalization ability that can avoid the over-fitting issue to a large extent; (3) The model has high explanatory performance because the model can find some words to describe each extracted topic.

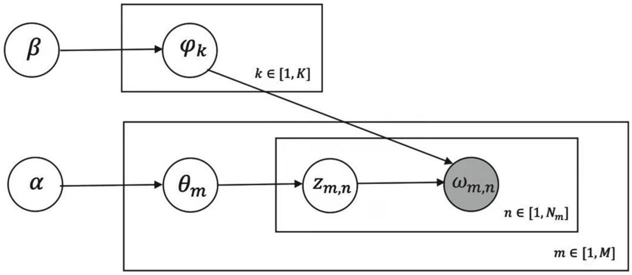

Fig. 1. Process of LDA modeling

4 Experiments

4.1 Dataset

The data used in this research was crawled by a data collector called Houyi data collector (http://www.houyicaiji.com). We focus on the light food store information and consumer reviews of the top 11 cities on the Meituan official website, including Beijing, Shanghai, Guangzhou, Shenzhen, Tianjin, Xi'an, Chongqing, Hangzhou, Nanjing, Wuhan, and Chengdu. Each data sample contains seven features: store name, rating, group purchase discount, menu, the total number of reviews, review text, and review time. A total of 30041 sample data was collected and 6968 samples were left after data cleaning. The data cleaning process includes the following two steps: (1) Considering the sample timeliness and the content validity, we deleted the comments before January 1st, 2017, and the light food stores that have no comments; (2) We deleted the samples that are blank, repeated and the ones that have less than four characters.

4.2 Data Description

According to the "Hot Cities" ranking on the Meituan official website, the main arena of domestic light food consumption is Beijing, Shanghai, Guangzhou, Shenzhen, Tianjin, Xi 'an, Chongqing, Hangzhou, Nanjing, Wuhan, and Chengdu. Therefore, research focuses on light food store information and consumer reviews in these 11 cities. The number of light food stores, as well as the total number of consumer reviews in each city is shown in Fig. 2. As can be seen from the figure, Beijing, Shanghai, and Guangzhou have the largest number of light food businesses. The light food businesses in these cities can be explained by the following reasons. These cities have a high level of economic development and advanced high-tech industries and service industries, which attract a large number of young people to these cities. These younger generations are living a fast-paced and stressful life and they have begun to embrace a new, healthy lifestyle.

Since 2016, some entrepreneurs who have a keen sense of business have started to set up their light food businesses in these cities. As of April 1st, 2021, Beijing had a total of 194 light food businesses, followed by Shanghai (137), and Guangzhou (119). The number of stores in Tianjin outperforms the other 7 new first-tier cities indicates that the light food industry has gradually developed from first-tier cities to the surrounding new first-tier cities (e.g., the correlation between Beijing and Tianjin) and that the market potential of the light food industry is huge in these new first-tier cities.

In terms of the number of reviews, Guangzhou has nearly 1,700 valid samples, which is more than twice as many as those of other cities (e.g., Beijing and Shanghai). Further analysis of the review text showed that the number of reviews at all SUBWAY (http://subway.com.cn/) branches in Guangzhou accounted for 31 percent of the total reviews. As the first-mover in the domestic light food industry, SUBWAY has its unique business strategy and it enjoys a series of features that successfully win its customer loyalty, such as fast serving speed, free matching of various dishes, transparent production process, and so on.

4.3 Experiment Settings

Firstly, sentiment analysis of the review text using the SnowNLP module was conducted after data cleaning. Secondly, topic extraction of the positive and negative comments obtained by sentiment analysis was conducted using LDA modeling. The optimal parameters were found through multiple training models before the final fitting model was determined.

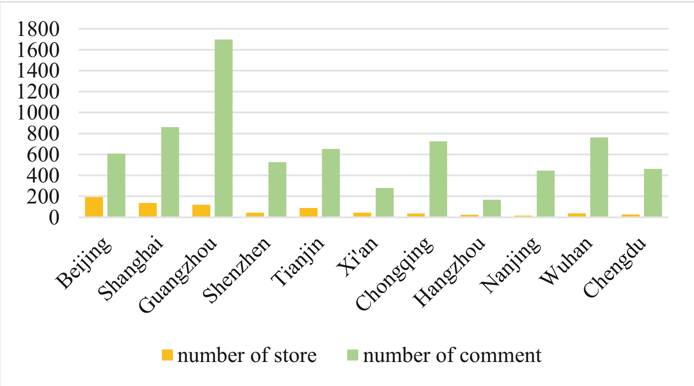

Fig. 2. Histogram of the number of light food stores and the number of customer reviews in each city

To improve the model prediction accuracy, three improvements were made regarding the SnowNLP module: (1) We replaced the word segmentation function in SnowNLP source code with Jieba word segmentation code because the precise model of Jieba word segmentation can achieve a more accurate sentence division, which is more suitable for text analysis. (2) We imported stop words and a dictionary, combined adverbs and verbs, filtered out a large number of meaningless words and symbols, and improve the word segmentation ability of SnowNLP on light food review analysis. (3) Replaced the default shopping corpus of SnowNLP with the corpus related to the catering industry to further improve the accuracy of sentiment analysis. 1000 comments were tested in the end and the accuracy of sentiment prediction has reached 88.4%.

To get the best number of topics, LDA models with a different number of topics are built step by step from 2 to 10. For each model, the first 50 high-frequency words and word frequency of different topics are obtained, and the number of all non-repetitive high-frequency words is represented by a vector to get the word frequency vector of each topic. The cosine similarity of these word frequency vectors is calculated, and the average value is obtained. Cosine similarity represents the similarity between vectors. The smaller the value is, the lower the similarity between vectors is, and the more different the content expressed between topics is.

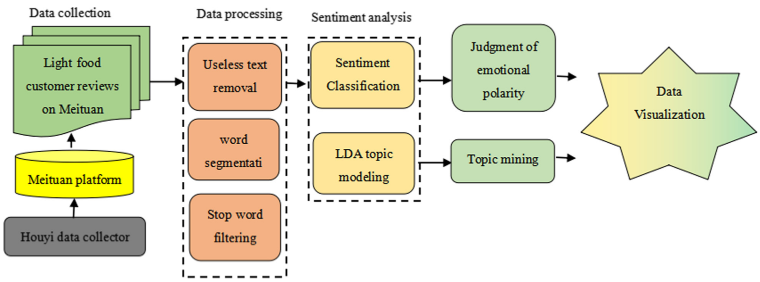

Fig. 3. Experiment setting

5 Results and Discussion

5.1 Topic Analysis of Overall Reviews

Keyword Analysis. The keywords of light food customer reviews on the Meituan were extracted and synonymously merged based on word representation. The top 20 keywords were selected for statistics, and the overall word frequency graph was shown in Fig. 4.

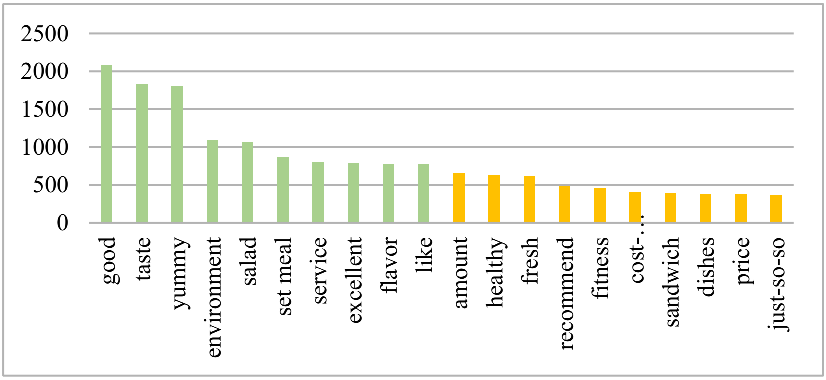

Fig. 4. Histogram of word frequency of overall reviews.

As can be seen from Fig. 4, the customer reviews on the light food area of the Meituan platform are generally positive as the keyword of "good" ranks the first and "yummy" ranks the third, and the customers' focal points mainly include taste, types of set meals, which again highlights that the emerging light food lovers value the taste of light food most. The keywords of "service", "environment" and "cost performance" revealed that these customers also pay great attention to the perceived service quality and dining environment.

Specifically, the overall focal points of light food consumer reviews on the Meituan platform can be categorized as follows: (1) Regarding the selection of light food, the main keywords are "salad", "set meal", "sandwich", "dishes", "recommendation", "cost performance", etc., indicating that consumers value the set meal or dishes that are available for choice during the light food consumption. As for the specific types of dishes, "salad" and "sandwich" are the most popular ones. Moreover, the customers are more likely to try out light food based on other consumers' recommendations and they are also inclined to actively recommend the light food options that they favor. (2) In terms of light food products, the main keywords include "taste", "health", "freshness", etc., indicating that consumers focus not only on the taste but also on the freshness of the raw materials. (3) In terms of light food service, the main keywords are "environment", "service", "feeling" and "boss", revealing that light food lovers also pay great attention to the dining environment and service. For example, some customers care about whether the restaurant is clean, whether the boss is approachable, as well as the service quality of the service staff.

Topic Analysis. Topic analysis of both the positive and negative comments was conducted via LDA topic modeling. Three positive topics and three negative ones as well as their corresponding keywords were eventually extracted and summarized. As illustrated in Table 1, three positive topics consist (1) healthy diet and fitness; (2) taste and nutrition; (3) cost performance and customer service, whereas the negative topics include: (1) Group purchase discount and taste; (2) customer service and dishes selection; (3) dining environment and food security.

On one hand, the three extracted positive topics show that the customers tended to give satisfied reviews when they felt the benefits of healthy eating and the function of fitness of the light food consumption. They also focus on the taste and nutrition of the food, the cost performance, and the service when giving positive reviews. Therefore, to win better electronic word of mouth, managers should establish long-term cooperative relations with high-quality suppliers, build traceable food materials management systems, and set up specialized food freezers to keep the raw materials fresh and healthy. Moreover, the constant focus on light food lovers is the taste of dishes, which requires the light food merchants to improve their cooking skills, pay special attention to the scientific collocation of the ingredients and dishes and create an impressive light food brand that features healthy and balanced. Regarding the fitness function of light food, managers in the light food industry can also partner with organizations like yoga, meditation, body shaping, or fitness clubs or community to offer regular light food delivery services.

On the other hand, the negative topics indicate that customers also complain about the low actual perception of group purchase discount, the limited types of food and flavor, the unprofessional customer service, and the dining environment and food with

Table 1. Topic extraction based on overall reviews

Keywords	Positive topics	Keywords	Negative topics
Good, environment, like, favor, yummy, healthy, clean, fitness, weight loss	Healthy diet and fitness	Attitude, taste, not yummy, goods, group purchase, service staff, little amount, the Meituan, service attitude	Group purchase discount and taste
Taste, yummy, taste, good, like, healthy, fresh, salad, sandwich, nutritious	Taste and nutrition	Service, just-so-so, service attitude, waiter, set meal, bad, food, bad, dishes	Customer service and dishes selection
Good, environment, favor, taste, set meals, service, yummy, salad, cost performance	Cost performance and customer service	Not tasty, little amount, taste, service, just-so–so, environment, group purchase, service staff, food hygiene	Dining environment and food security

poor sanitary level. Results also reveal that "group purchase discount" and "cost performance" are the focus of consumers' attention. Thus, businesses can design more group purchase discount schemes and implant some convenient tools that consumers can share on various social platforms and quickly realize crowd-ordering. In this way, consumers can enjoy discounts whereas the managers can use group ordering tools to achieve a unified food distribution and after-sales management system. Furthermore, light food, as a new star in the catering industry, customer perception of service quality is a very important part of the whole consumption process, considerate service can greatly attract and retain customers. Employers in the light food industry can train their employees regularly to improve their professional quality. As for the environment, managers in the light food industry can create a dining environment in harmony with light food culture and improve service quality. For example, they can create unique storefront layout and decoration style according to the themes of green, ecology, health, nature, and so on. In this way, customers can get a unified dining experience of light food products and a dining atmosphere. They can also provide customers with a transparent production process of light food, reducing customers' concerns about food hygiene issues.

5.2 Topic Analysis of Reviews During COVID-19 Epidemic

To track the changes of the online review focal points before and after the COVID-19 epidemic, we first analyzed how the number of comments changed over time from the fourth quarter of 2017 to the first quarter of 2021.

As illustrated in Fig. 5 the cumulative number of consumers' comments on light food in the last three years before the epidemic (at the end of 2019) was the same as the number of comments in the year after the epidemic. Moreover, a significant drop in the number of comments can be easily noticed in the first quarter of 2020 because of the outbreak of COVID-19 whereas, in the post-pandemic era, consumers' enthusiasm for light food consumption has significantly improved. The following reasons might account for this phenomenon. Firstly, according to the data released by the National Bureau of Statistics, catering income declined by 43.1% year on year from January 2020 to February 2020, leaving heavy blows and huge challenges for most light food businesses. Secondly, most restaurants suspend operations or even close down because of the unprecedented challenges brought by the COVID-19 and some food services in the light food restaurants are unavailable (e.g., eat-in service) since the epidemic-combating policies are very strict among first-tier cities. During the epidemic period, many consumers voluntarily reduced the frequency of eating out and ordering takeaway food based on safety considerations. They tended to cook at home or eat in school or company canteens.

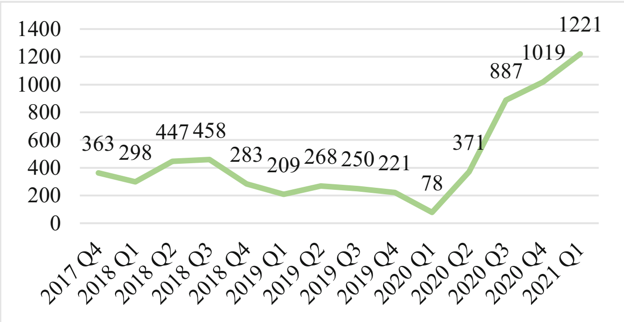

Fig. 5. Comparison of the number of reviews before and after the epidemic.

After diving the overall comments into two groups, where one is the pre-epidemic phase with 3514 reviews left from January 2017 to December 2019 whereas the other is the post-epidemic phase with 3679 comments covering the period from January 2020 to March 2021. we compared and analyzed the content in these two stages respectively and the results are visualized in the two Word Cloud maps (Fig. 6) and the keywords are listed in Table 1.

Fig. 6. Hot word cloud of reviews before (left) and after (right) the COVID-19.

As Table 2 presented, customer reviews of light food tend to appear positive, regardless of the COVID-19 outbreak. After experiencing the epidemic, people's favorable attitude towards light food consumption is improving (the word frequencies of the keyword of "like" are 823 and 1030 respectively).and the consumers seem to still put "health", "taste" and "environment" in the first place. Moreover, after the epidemic, consumers pay more attention to "cost performance", "set meals" when leaving messages in the light food comment area.

Table 2. The changes of the top 11 keywords before and after the epidemic (excluding the keyword of "good").

No.	Keywords before COVID-19 (word frequency)	Keywords after COVID-19 (word frequency)	Rank (changes)
1	Cost performance (70)	Cost performance (380)	↑49
2	Set meal (129)	Set meal (741)	↑18
3	Taste (217)	Taste (646)	↑5
4	Service (327)	Service (512)	↑2
5	Amount (172)	Amount (395)	↑2
6	Like (823)	Like (1030)	—
7	Yummy (818)	Yummy (1007)	—
8	Environment (445)	Environment (672)	—
9	Fresh (317)	Fresh (299)	↓4
10	Salad (614)	Salad (455)	↓4
11	Health (336)	Health (293)	↓7

6 Conclusion, Limitation and Future Research

6.1 Conclusion

Based on a dataset of 6968 light food customer reviews on the Meituan platform, this research first conducted the sentiment analysis and classification of the light food customer comments via the SnowNLP technique. In addition, LDA topic modeling was used to analyze positive and negative topics of customer reviews. The experimental results were visualized and the research showed that the SnowNLP technique and LDA modeling achieve high performance in extracting the customers' sentiments and focal points, which provides theoretical and data support for light food businesses to improve customer service.

6.2 Limitation and Future Research

There are some limitations to this research. Firstly, this research only focuses on the specific type of food (i.e., light food) in a specific platform (i.e., the Meituan) and it might have some limitations on the generalization of the experiment results and implications. Secondly, we focus on how to use data analysis and visualization methods to mine the useful information in the comments in the light food area, whereas we do not compare the existing algorithm with other text analysis algorithms. More algorithm comparison and evaluation will be added to in future research to better prove the high performance and robustness of the results. Thirdly, this research only focuses on the word frequency and sentiment analysis of the customer online reviews. Future research might consider employing linear regression models to examine the associations of various text-based review features on customers rating scores, light food businesses ranking, overall financial performance, and so on.

Acknowledgement. This study is supported by Shenzhen University Postgraduate Innovation and Development Research Project (Special project for the centennial of the Founding of the Party) in the year of 2021(**No.315-0000470616**).

References

1. China Light Food Takeout Consumption Report. http://www.ce.cn/cysc/sp/info/201910/18/t20191018_33377815.shtml
2. Why on earth has the one hundred billion light food market dropped from the top? https://www.tmtpost.com/3972685.html. Assessed May 30
3. Chevalier, J.A., Mayzlin, D.: The effect of word of mouth on sales: online book reviews. J. Mark. Res. **43**(3), 345–354 (2006)
4. Alrawadieh, Z., Law, R.: Determinants of hotel guests' satisfaction from the perspective of online hotel reviewers. Int. J. Cult., Tourism Hospital. Res. **13**(1), 84–97 (2019)
5. Chittiprolu, V., Samala, N., Bellamkonda, R.S.: Heritage hotels and customer experience: a text mining analysis of online reviews. Int. J. Cult. Tourism Hospital. Res. **15**(2), 131–156 (2021)
6. Kumar, V., Bhagwat, Y., Zhang, X.: Regaining 'lost' customers: the predictive power of first-lifetime behavior, the reason for defection, and the nature of the win-back offer". J. Mark. **79**(4), 34–55 (2015)
7. Sun, K.A., Kim, D.Y.: Does customer satisfaction increase firm performance? An application of the American Customer Satisfaction Index (ACSI). Int. J. Hosp. Manag. **35**, 68–77 (2013)
8. Ruz, G.A., Henríquez, P.A., Mascareño, A.: Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers. Futur. Gener. Comput. Syst. **106**, 92–104 (2020)
9. Gopalakrishnan, K., Salem, F.M.: Sentiment analysis using simplified long short-term memory recurrent neural networks (2020)
10. Wang, S., Wu, S.: Tourist attention and emotion analysis based on online comment text of scenic spots. J. Guizhou Univ. Nat. Sci. **34**(6), 69–73 (2017)
11. Wang, Q., Gao, W.: Sentiment statistical analysis of online travel reviews based on the hierarchical quantile regression model. China Price **12**, 82–85 (2019)
12. Aggarwal, C.C.: Mining text data. In: Data Mining, pp. 429–455. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-14142-8_13

13. Miner, G., Elder IV, J., Fast, A., Hill, T., Nisbet, R., Delen, D.: Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications. Academic Press (2012)
14. Yang, X., Yang, D., Su, H., Song, Z., Yang, X., Luo, Z.: Topic mining of scenic spot reviews based on sentiment analysis. Libr. Inf. Guide **5**(08), 59–65 (2020)
15. Chen, S., Zhang, Y., Li, B.: Sentiment analysis of health and medical wearable device comments based on LDA model. Chin. J. Med. Libr. Inf. **29**(12), 41–47 (2020)
16. Pang, B., Lee, L., Vaithyanathan. S.: Thumbs up? sentiment classification using machine learning techniques. In: Language Processing (EMNLP), pp. 79–86. Philadelphia (2002)
17. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)