International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 7308-7315

https://doi.org/10.70517/ijhsa463626

Research on Service Quality Improvement of Takeaway Platform Based on Natural Language Processing

Weijun Tang[1] and Shaohong Gu[1,*]
[1] School of Business, Jiangnan University, Wuxi, Jiangsu, 214122, China
Corresponding authors: (e-mail: 18840619851@163.com).

Abstract Takeaway platform reviews are users' direct feedback on the platform service quality, mining and analyzing the emotional connotation of review data has important significance for the improvement of takeaway platform service quality. This paper adopts Meituan catering takeaway platform as the data source for the study, and utilizes text de-emphasis and filtering phrase methods for data preprocessing of the comment corpus text. Then a bidirectional LSTM takeout review data sentiment classification network is proposed to construct a bidirectional LSTM sentiment classification model. On the basis of this model, a self-attention mechanism is added to fully mine and learn the relevant laws of the takeout platform comment data to improve the accuracy and reliability of the model's takeout comment sentiment classification results. Incorporating the sample screening strategy of uncertainty and diversity to enrich the sample variety and improve the accuracy of the model's classification and prediction of the sentiment of the sample corpus text in the training set, the sentiment classification model based on the self-attention mechanism is constructed. Compared with similar model algorithms, the accuracy and precision of the model in predicting and categorizing the sentiment of the review corpus text of takeaway platforms is as high as 93.10% and 94.51%, which is suitable for the needs of the sentiment processing of the review data corpus text of takeaway platforms.

Index Terms takeaway platform review data, self-attention mechanism, bi-directional LSTM sentiment classification model, text pre-processing

I. Introduction

As a new business model in the era of “Internet +”, takeaway platforms are popular among consumers for providing full-time, cross-category online takeaway services. China's takeaway industry is expected to develop into a trillion-dollar market in the next one to three years, and there is still huge room for development in China's takeaway industry market. However, with the expansion of the takeaway market scale and the intensification of competition, the performance of the service quality of takeaway platforms, however, is not as satisfactory as it should be [1]. For example, food safety, delivery safety, as well as the mismatch between online information and offline services, and the user's crisis of trust in the platform, etc., have triggered the dissatisfaction of platform users [2]-[4]. These negative information not only affects consumers' perception and evaluation of the logistics service quality of the takeaway platform, but also directly or indirectly affects customer satisfaction and customer loyalty [5]-[7]. It can be seen that the service quality problem has become one of the reasons limiting the development of takeaway platforms [8].

In the face of the future trillion-dollar market, if takeaway platform enterprises want to gain an advantage in the fierce competitive environment, they need to conduct a scientific and reasonable evaluation of the service quality, find out the weak links in the service quality, and improve the service quality in a targeted way [9]-[11]. Only by emphasizing the service quality of the takeaway platform can we fundamentally eliminate the dissatisfaction of the platform users, maintain the relationship with the platform users, and thus establish a solid foundation for the long-term development of the takeaway platform [12], [13].

This paper takes the review corpus of Meituan catering takeout platform as the research data source, and completes the data preprocessing of the review corpus text by text de-weighting and filtering phrases after using Octopus Collector to capture and acquire the data source. Secondly, the basic principle and calculation method of the bi-directional LSTM sentiment classification model for takeaway review data are elaborated in detail, and then the bi-directional LSTM sentiment classification model based on the self-attention mechanism is proposed as the mining and classification method for takeaway review sentiment. The operation method of active learning sample screening strategy is analyzed again, and the bi-directional LSTM sentiment classification model based on self-attention mechanism is integrated to build a sentiment analysis model under the sample screening strategy

based on uncertainty and diversity. Finally, the model is used to mine and analyze the sentiment connotation of takeaway review data on Meituan platform, and the effectiveness of the designed model is examined by comparing the algorithms of similar models.

II. Preprocessing of Text Data for Takeaway Platform Reviews

II. A.Acquisition of Takeaway Review Text

The data selected in this paper comes from the takeout reviews of a certain region on Meituan. When analyzing the user review information, it is necessary to extract effective information from this massive data in order to guide the takeout platform and merchants to improve the deficiencies, and to enhance the user's satisfaction with the platform and the merchants. The webpage source code of the online review text of Meituan Takeout is in HTML format, so Python can be used to parse and extract the text data. In addition to writing Python code to realize data collection, the third-party collection tool Octopus can also be used to realize the crawling of takeout review data. The source of the dataset in this paper is mainly through the Octopus collector for data crawling in order to obtain the dataset of takeaway review text, the software is stable and efficient, crawling data speed, after crawling the data there will be the option of removing duplicated data, so that it can achieve the purpose of initially reducing the duplicity of data.

In order to avoid too little review data from a single merchant, the labeled raw dataset used in this paper is accumulated from the takeout review data of multiple stores. A total of about 15,000 review data are crawled, including 7,000 positive reviews and 8,000 negative reviews. The crawled fields include takeout reviews, order time, order user ID, product ID, review star rating and so on.

II. B.Text de-emphasis

Text de-duplication, i.e. removing duplicates in the text. The reason for duplicates in the evaluation may be that some merchants want to increase the rate of favorable reviews and store ratings of food in the catering industry, so as to avoid the reduction of traffic exposure and thus affecting the conversion rate of customers' consumption in the store, and therefore provide some tempting conditions for diners, such as commenting on coupons and other attractive conditions, which may cause takeaway users to copy the comments of other people or use the online positive feedback templates, resulting in the emergence of a large number of duplicates, and it may be due to the fact that some of the It may also be due to the fact that some takeaway users directly write simple and short words such as “delicious” and “unpalatable” in order to save time, and the probability of duplicates is often higher in this way. If we don't remove a large number of duplicate texts, it is easy to cause storage pressure and unnecessary time-consuming for subsequent text analysis.

There are many text de-duplication methods, such as KSentence, Simhash, Minhash, Kshingle, etc. KSentence is generally to remove some of the punctuation and special characters, to retain commas, periods, colons, line breaks and other statement-separating punctuation.Simhash, Minhash and Kshingle is generally to remove punctuation, spaces, special characters, etc. In engineering applications, the most commonly used text de-emphasis is Simhash algorithm, edit distance de-emphasis algorithm and comparison deletion method. But here Simhash algorithm is not very applicable, this is due to many customers do not pay too much energy and time to comment, so the text caused by the similarity is unavoidable, in the Simhash algorithm to determine the similarity of the text is the standard Hammingdistance is less than or equal to 3, when the Hammingdistance is less than or equal to 3, it is considered that Two texts are similar, need to remove one of the duplicated text, but they even similar there may be some information with analytical value, so as to remove completely duplicated takeaway review text using comparative deletion method. After the text de-duplication, a total of 3355 text duplicates are removed in this paper, and the remaining data is 12545.

II. C.Filtering phrases

In the process of commenting, for the comment data, the smaller the number of words, the less information can be extracted. If the text data of takeaway comment is longer, it means that the user's emotional expression of the takeaway goods is comprehensive, and the more features can be extracted, but if the text data of the takeaway comment is shorter, it expresses fewer meanings and is unlikely to contain too much useful information. However, if the takeaway review text data is short, it is less meaningful and unlikely to contain too much useful information. Single review descriptions, such as “OK”, “average”, “good”, etc., are not very meaningful for sentiment analysis. Therefore, first of all, the length of the comment text is counted, the threshold is set, and the comments whose takeaway comment length is below the threshold are deleted, and in this paper, the comment text whose comment length is less than 3 is deleted. After the operation of filtering phrases, a new data is finally obtained, and the

amount of new commodity comment data is 9,785, and data cleaning reduces a large amount of invalid data, and improves the running speed for the subsequent model building.

III. Emotion Classification Model Based on Self-Attention Mechanism

III. A. Bidirectional LSTM Sentiment Classification Models

Although LSTM can effectively solve the problem of long-term dependence, the information in the following part of the takeaway review data is not fully utilized, and the bidirectional LSTM model can well solve this problem. The basic principle of this model is to connect a takeaway review data with a forward LSTM network and a backward LSTM network, so as to extract the upper part of the takeaway review data in the forward network and the lower part in the backward network, and ultimately achieve a significant improvement in the accuracy of the model.

The hidden state H_t of the bi-directional LSTM model at moment t consists of forward $\vec{h}_t$ and backward $\overleftarrow{h}_t$, which are computed in Eqs. (1)-(3):

$$\vec{h}_t = \overrightarrow{LSTM}(h_{t-1}, w_t, c_{t-1}), t \in [1, T] \tag{1}$$

$$\overleftarrow{h}_t = \overleftarrow{LSTM}(h_{t+1}, w_t, c_{t+1}), t \in [1, T] \tag{2}$$

$$H_t = [\vec{h}_t, \overleftarrow{h}_t] \tag{3}$$

At this point the output H_t of the bi-directional LSTM will be taken as the feature vector of the text.

III. B. Bidirectional LSTM sentiment classification model based on self-attention mechanism

The bidirectional LSTM sentiment classification model based on the self-attention mechanism is shown in Fig. 1.

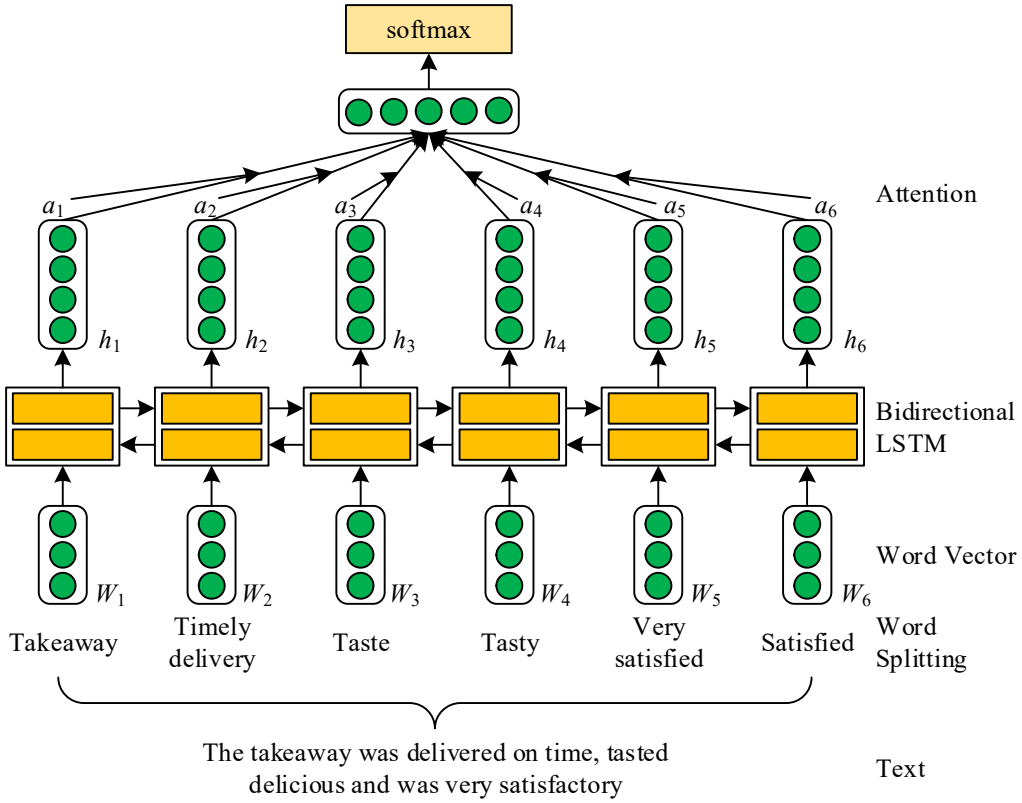

Figure 1: LSTM sentiment classification model based on self-attention mechanism

The self-attention mechanism works according to the working characteristics of the human brain, and its important feature is to allocate more attention to relatively important content and less attention to content of lower importance. Due to the brain-like characteristics of the self-attention mechanism, it has achieved more extensive applications in many fields, such as in image annotation, text level classification, etc. In the self-attention mechanism, its calculation method is shown in Eqs. (4)-(6):

$$u_t = \tan(w_w H_t + b_w) \tag{4}$$

$$a_t = softmax(u_t^T, u_w) \tag{5}$$

$$v = \sum_t a_t H_t \tag{6}$$

where u_t is the hidden unit of H_t, u_w is the context vector, a_t is the attention vector, and v is the output vector obtained after the operation of the self-attention mechanism. The variable u_w adopts a strategy of random initialization and continuous ongoing learning in the model.

The output v of the self-attention mechanism is input to the softmax layer to predict the results, and the prediction results are calculated as shown in equation (7):

$$Y = softmax(w_c v + b) \tag{7}$$

If Y' is the real distribution of takeout review data, and the loss function selects the cross-entropy function, then the loss function calculation method is shown in Equation (8):

$$loss = -\sum_i y_i' log(y_i) \tag{8}$$

The model parameters were subsequently trained using stochastic gradient descent.

III. C. Sample selection strategies that incorporate uncertainty and diversity

As only considering the uncertainty of the samples and ignoring the diversity of the screening samples, it leads to the selection of a single type of samples, which makes the diversity of samples in the training set decrease and affects the accuracy of the prediction. Therefore, the sample selection strategy is improved by considering the uncertainty of the screening samples and the diversity of a batch of samples. Uncertainty is kept unchanged from the above definition, while distance is considered for the definition of diversity. Through the method of cluster analysis, the samples with high similarity are clustered into one class, and then one sample is selected from each class, which can effectively reduce the similarity between samples.

In this paper, we further develop a model for sentiment analysis under the uncertainty- and diversity-based sample screening strategy. The method framework is similar to the method of screening samples based on uncertainty, but in the process of sample screening, not only consider the uncertainty of the samples, but also consider the diversity of a batch of samples, and set each iteration to screen K samples to be labeled manually, and the specific operation methods are as follows:

(1) Initially randomly select a batch of data for labeling, the establishment of the basic sample screening model and prediction model.

(2) Use the sample screening model to predict all the samples in the unlabeled sample pool, and calculate their uncertainty, denoted as uncertainty.

(3) Select the top n samples with the highest uncertainty $(n > K)$. Cluster these n samples and set the number of categories to K.

(4) Calculate the distance of these n samples to their respective cluster centers, denoted similarity.

(5) Combine uncertainty and similarity, since uncertainty takes values in the range of $[0, 0.5]$, transform similarity to this range as well, and transform the indicator to between $[0,1]$ through the sigmoid function, and then divide it by 2. Finally, combine these two indicators, and construct the indicator as in Equation (9):

$$US = \frac{1}{2(1+e^{-similarity})} - uncertain\,y \tag{9}$$

When the US value is smaller, the uncertainty and diversity of the samples are higher, containing more information, suitable for manual judgment.

(6) Calculate the US value of each of the n samples, and select the sample with the smallest US value in each of the K classes, resulting in a total of K samples, i.e., the final filtered samples.

IV. Analysis and evaluation of model applications

IV. A. Sentiment Analysis of Review Data on Takeaway Platforms

IV. A. 1) Word2Vec word vectorization

Word2Vec training is performed on the acquired takeout review data source to obtain word vectors with the number of words used by the model being 3333 as well as the word length of 100 at a time, among which the top 10 words with the highest word frequency are (w1) good, (w2) flavor, (w3) delicious, (w4) delivery, (w5) eat, (w6) deliver, (w7) nice, (w8) hour, (w9) point and (w10) said, and the word vectors for the 10 words are shown in Table 1.

Table 1: The word vectors of the top 10 words in terms of frequency

Word	Word embedding	Word frequency
(w1)	[-0.023598 0.012369 0.085264 0.032632 -0.089041 …]	1981
(w2)	[0.0598631 0.021458 0.012396 0.025698 -0.096325 …]	1913
(w3)	[0.0478952 -0.063698 0.014725 0.014053 0.086321 …]	1761
(w4)	[0.0853214 0.012458 0.0366523 0.012603 0.074154 …]	1569
(w5)	[-0.0479893 0.023659 0.0112323 0.014962 0.096325 …]	1520
(w6)	[0.0147236 -0.025896 0.047456 0.012358 0.086325 …]	1436
(w7)	[-0.0963586 0.044523 0.0144236 0.026935 0.065234 …]	1225
(w8)	[-0.0589662 0.011236 -0.069325 0.036523 0.059632 …]	1058
(w9)	[0.0412587 -0.047852 -0.025652 0.045963 0.048524 …]	953
(w10)	[0.0178239 0.096325 0.013489 0.059632 0.069325 …]	819

Word2Vec model training can also get the word index dictionary vocab_dict, in which the index of each word is obtained by sorting the word frequency from the largest to the smallest, for example, the index of “good” with the highest word frequency is 1, which can be used to match the words in each text of the dataset with the word vectors trained by Word2Vec model to construct the word vector matrix of the text, i.e., each text is transformed into a word index vector to be input into the model. Thus, the word index can be used to match the words in each text of the dataset with the word vectors trained by the Word2Vec model to construct the word vector matrix of the text, i.e., each text is transformed into a word index vector to be input into the model. The length and frequency distribution of the review data of Meituan Takeout platform are shown in Fig. 2, for the words that are not included in the word index dictionary, they are replaced by 0, and because the length of the review texts in the dataset varies after word splitting, and most of the texts are statistically less than 20, we set the fixed length of the word index vectors of the texts to 20, and fill in 0 at the end of the word index vectors of the texts with less than 20, and truncate them at the end of the word index vectors with more than 20, we set the word index vectors of the texts to 20. For the length of less than 20, we use the way of filling 0 at the end, and for the length of more than 20, we use the way of truncation at the end, so that we can get the text data of the same length.

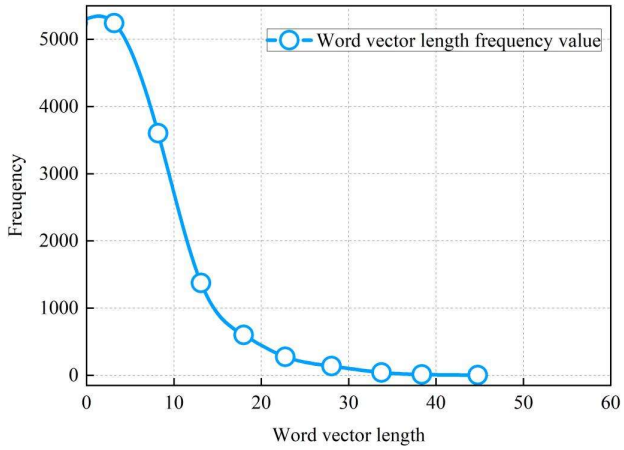

Figure 2: The length and frequency of comment data on Meituan Waimai platform

IV. A. 2) Sentiment analysis based on word frequency

The pre-processed review data is input into the sentiment classification model based on the self-attention mechanism to start the sentiment classification analysis of the review data of Meituan food delivery platform. Figure 3 shows the frequency of the 10 words with the highest word frequency in the comment data set of Meituan Takeout, and it can be found that the frequency of the words (w3) delicious, (w7) good, (w2) flavor, (w4) delivery, and (w6) delivery are all in the range of 5,000 times and above, which means that they belong to the keywords in the comment text, and the comments from the keywords can be obtained from these keywords, and they are mostly concentrated on the feedback on the taste of the dishes and the delivery process. From these keywords, we can get that customers' comments mostly focus on giving feedback on the taste of the dishes and the delivery process.

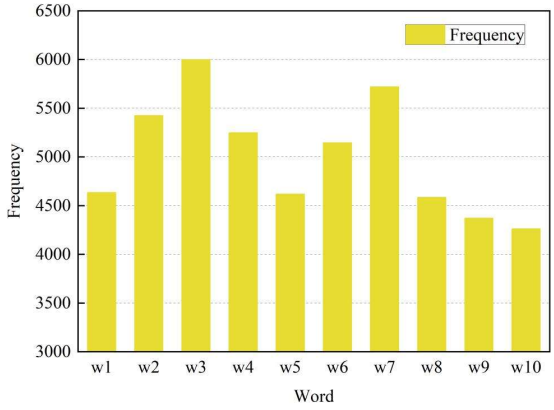

Figure 3: The frequency of comments appearing on the Meituan platform

IV. A. 3) Sentiment Analysis Based on TF-IDF Feature Extraction

Each comment represents a vector with the dimension of dictionary length, and the vector takes the value of 0 or 1, where 1 means that the corresponding word is included in the comment and 0 means that the corresponding word is not included in the comment, and then the extraction of text features is performed using TF-IDF. The obtained text features are input into the proposed model to analyze the emotional connotation of the text.

Table 2 shows the TF and IDF values of the words in the text. The first column of the comment ID is the ID in the comment statement where the participle is located. Text is the result of the comment text after the participle is performed, which are (a1) delicious, (a2) soon, (a3) amount, (a4) taste, (a5) foot, and (a6) jittery. N is the number of times all the words appeared in the comment ID where they are located. In the comment where the ID is 1, the All words appear only once, so the word frequencies are all 20%.IDF gives smaller values to more common words, such as (a1) delicious, (a4) flavor in the table, which correspond to smaller IDF values, all less than 2.500. And it gives larger values to less common words, such as (a6) shake, (a5) foot, (a3) amount which correspond to larger IDF values, all greater than 3.000.N is the number of times all words appear in the comment IDs where they are found. N is the number of times all words appear in the comment IDs where they are found.

Table 2: Examples of the values of the words TF and IDF

Comment ID	Text	N	TF	IDF
3	(a1)	1	0.326	2.046
3	(a2)	1	0.326	3.056
4	(a3)	1	0.326	3.337
9	(a4)	1	0.326	2.064
6	(a5)	1	0.326	4.31
12	(a6)	1	0.293	7.655
13	(a1)	1	0.293	2.048
13	(a3)	1	0.293	3.329

Table 3: Document analysis of TF-IDF value examples

Comment ID	3	12	13
(a1)	0.553	0	0.489
(a2)	0.755	0	0
(a3)	0.811	0	0.704
(a4)	0.557	0	0
(a5)	1.006	0	0
(a6)	0	1.424	0

Table 3 shows the TF-IDF values of the words in the text, the TF-IDF value is used to measure the importance of a word to the article, the larger its value, the higher the importance of the word to the article. The training of the comment data is processed for a whole comment statement, words appear in different comments with different

frequencies, corresponding to its TF-IDF in different comments also have different values, 0 represents that the word does not appear in the current comment, (a1) delicious, (a4) flavor corresponding to the TF-IDF is smaller, indicating that these words are more common in the document, giving it a smaller word weights and are less important to the document.

IV. B. Assessment of overall performance

The change of accuracy and loss of the model proposed in this paper on the training set is shown in Fig. 4, with the increase of the number of iterations, the accuracy of the model gradually increases, and stabilizes at 100 iterations, and finally stabilizes at about 81.00%. Meanwhile, the model loss gradually decreases and stays around 0.2 at 400 iterations.

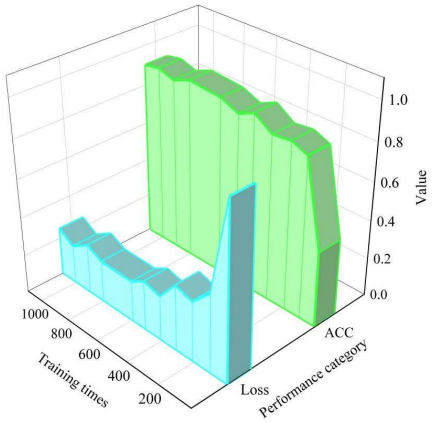

Figure 4: The accuracy rate and loss variation of the training set

The experimental results of comparing with a total of four similar models, namely (M1) TextCNN, (M2) Tex-tRNN, (M3) TextRNN_Attention, and (M4) Transformer, are shown in Table 4.The accuracy of this paper's model on this dataset is 93.10%, which is 3.65% higher than (M1) TextCNN, 0.4% higher than (M2) Tex-tRNN by 0.4%, than (M3) TextRNN_Attention by 1.15%, and than (M4) Transformer by 0.98%. And it generally outperforms other models in all three evaluation indexes, with performance data of 94.51%, 96.00% and 95.10% in precision and recall, and F1 value, respectively. In terms of training time, the model in this paper takes only 12 seconds, which is much better than other similar model algorithms, indicating that this model is suitable for fast classification of the emotion of the takeaway.

Table 4: Comparison of model experiments

Model	Accuracy Rate(%)	Precision Rate(%)	Recall Rate(%)	F1 Value(%)	Training time (s)
(M1)	89.45	92.86	92.75	92.8	49
(M2)	92.7	94.11	96.62	95.35	35
(M3)	91.95	91.54	98.74	95	22
(M4)	92.12	94.15	95.12	94.83	40
Textual	93.1	94.54	96.69	95.6	12

V. Conclusion

(1) In this paper, the self-attention mechanism is added to the constructed bi-directional LSTM sentiment classification model for takeaway review data to build a sentiment classification model based on the self-attention mechanism. And it combines the sample screening strategy of uncertainty and diversity as a natural language processing method for takeaway platform review data. In the experiments, the accuracy rate of the designed model stabilizes at 81.00% in 100 iterations, and the loss value starts to converge in 200 iterations and finally stays at 0.2. In the overall performance evaluation, the model of this paper shows far better performance than similar model algorithms in terms of accuracy rate, precision rate, recall rate, F1 value, and training time, which are 93.10%, 94.51%, 96.0%, 94.51%, and 96.0%, respectively, 94.51%, 96.00%, 95.10% and 12s.

(2) The Word2Vec model is used to preprocess the data source samples, extract the TD-IDF features of the corpus text, and input them into the sentiment classification model based on the self-attention mechanism to carry out the sentiment analysis of the review data of the takeaway platform. Among them, the frequency of words such as delicious, good, flavor, delivery, send, etc. are 5000 times and above, which are keywords in the review text. Accordingly, this paper suggests that the takeaway service platform should focus on the optimization of merchant dishes and rider delivery management in the construction strategy of service quality.

References

[1] Liu, C., & Chen, J. (2021). Consuming takeaway food: Convenience, waste and Chinese young people's urban lifestyle. Journal of Consumer Culture, 21(4), 848-866.

[2] Song, C., Guo, C., Hunt, K., & Zhuang, J. (2020). An analysis of public opinions regarding take-away food safety: A 2015–2018 case study on Sina Weibo. Foods, 9(4), 511.

[3] Flanagan, M. A., & Soon-Sinclair, J. M. (2025). Consumers' perceptions of regulatory food hygiene inspections of restaurants and takeaways. British Food Journal, 127(3), 897-913.

[4] Chen, J., & Liu, C. (2022). Takeaway food consumption and its geographies of responsibility in urban Guangzhou. Food, Culture & Society, 25(3), 354-370.

[5] Sikder, A. S., & Anae, N. (2023). Transforming Takeaway Management: Enhancing Efficiency and Customer Experience through an Online Ordering System for Sydney's Takeaway Shops.: Online Takeaway Shops in Sydney. International Journal of Imminent Science & Technology., 1(1), 64-78.

[6] Wang, C. (2020, December). Customer Satisfaction Evaluation of Food Delivery Platforms-Taking Meituan as an Example. In 2020 International Conference on Big Data Economy and Information Management (BDEIM) (pp. 124-127). IEEE.

[7] Chan, H. L., Cheung, T. T., Choi, T. M., & Sheu, J. B. (2023). Sustainable successes in third-party food delivery operations in the digital platform era. Annals of operations research, 1-37.

[8] Li, C., Mirosa, M., & Bremer, P. (2020). Review of online food delivery platforms and their impacts on sustainability. Sustainability, 12(14), 5528.

[9] Kumolu-Johnson, B. (2024). Improving service quality in the fast-food service industry. Journal of Service Science and Management, 17(1), 55-74.

[10] Alghamdi, S. Y., Kaur, S., Qureshi, K. M., Almuflih, A. S., Almakayeel, N., Alsulamy, S., & Qureshi, M. R. N. (2023). Antecedents for online food delivery platform leading to continuance usage intention via e-word-of-mouth review adoption. PloS one, 18(8), e0290247.

[11] Torres, E. N. (2014). Deconstructing service quality and customer satisfaction: Challenges and directions for future research. Journal of Hospitality Marketing & Management, 23(6), 652-677.

[12] Furunes, T., & Mkono, M. (2019). Service-delivery success and failure under the sharing economy. International Journal of Contemporary Hospitality Management, 31(8), 3352-3370.

[13] Mai, X. T., & Nguyen, T. (2025). Understanding users' trust transfer mechanism in food delivery apps. Journal of Hospitality and Tourism Insights, 8(4), 1582-1601.