Fuzzy Cloud Evaluation of Service Quality Based on DP-FastText

HUANZHUO YE [1], YUAN LI [2]

[1] Zhongnan University of Economics and Law
Add: 182# Nanhu Avenue, East Lake High-tech Development Zone, Wuhan, Hubei, 430073, P.R.CHINA
[2] NetEase (Hangzhou) Network Co., Ltd.
Add: Netease Building, No.399 WangShang Road, Binjiang District, Hangzhou, Zhejiang, 310052, P.R.CHINA

Abstract: This study proposes a service quality evaluation model framework which integrates automatic data acquisition, intelligent data processing and real-time data analysis with online comment data as data sources by introducing natural language processing technology based on management methods to break the traditional idea of over-reliance on human resources for service quality evaluation. The framework is mainly divided into text data preparation, fine-grained sentiment analysis and fuzzy cloud evaluation models. Data preparation module is responsible for preparing the initial data, and the fine-grained sentiment analysis module is responsible for pre-training a fine-grained sentiment classification model. The fuzzy cloud evaluation module uses the data obtained from the first two modules to evaluate service quality. By applying the model into catering industry, the feasibility of the model is proved and individuality, efficiency, dynamicity and intelligence of the model give it more advantage in the practice of service quality evaluation.

Key-Words: Service Quality Evaluation, Fine-grained Sentiment Analysis, Fuzzy Cloud Model, FastText Model
Received: February 28, 2021. Revised: July 5, 2021. Accepted: July 18, 2021. Published: August 2, 2021.

1 Introduction

Service industry is an important driving force of China's economic growth. Despite the continuous development of China's service industry and the improvement of the hardware level of service enterprises, there are still some gaps between the software services in China's service industry and the ones with the advanced international level. The proportion of service consumer complaints reached 52.67% in 2017, increased by 16.72 percentage points [1], and since 2015, service complaints have been showing an upward trend year by year. Therefore, improving service quality makes great differences in reducing service complaints and enhancing the market competitiveness of enterprises.

The foundation of improving service quality lies in establishing a set of scientific and effective service quality evaluation model to accurately identify the existing problems of service quality. Service quality evaluation can be divided into data collecting, data processing and data modeling stages. Data collecting and processing are time-consuming, subjective, lagging and limited. Some researches obtain structured data through questionnaires or surveys, which are of poor efficiency and limited scope. Other researches obtain unstructured text data from websites, and in this way, data collecting efficiency can be greatly improved and more information can be tapped, however, the process to transfer text data into useful evaluation information is quite time-consuming and subjective.

This research aims to contribute to breaking the traditional idea of over-reliance on human resources for service quality evaluation and proposing a service quality evaluation model framework which integrates automatic data acquisition, intelligent data processing and real-time data analysis with online comment data as data source. More specifically, the objectives of this research are to:

1) Improve the efficiency and intelligence of service quality evaluation

2) Seek a dynamic method to achieve real-time monitoring and quickly detecting problems of service quality

3) Transfer origin data into required data more efficiently, freely and objectively.

To achieve the objectives of this research, customer satisfaction is introduced as an index to evaluate service quality, which comes from the idea that customer satisfaction is a level of people's feeling state and customer satisfaction depends on the experience after customer's perception of product or service being compared with customer's expectations before it being accepted [2]. To obtain customer satisfaction more efficiently and intelligently, sentiment analysis is applied to transferring customer reviews into index data of customer satisfaction, as sentiment analysis is an intelligent and automatic technique, and customers' sentiments towards services revealed in reviews can directly reflect whether one is satisfied or not.

The rest of the paper is organized as follows: The next section presents a literature review of service quality evaluation, followed by different stages of service quality research and the conclusion of current research status of service quality evaluation. In the architecture and methodology section, the automatic, intelligent and dynamic solution of service quality and the detailed module of each model in the solution are present. Then, the proposed model is applied to catering field to prove the effectiveness and superiority of the model in the following section. Finally, we conclude the innovation of the proposed model and the implication of the research.

2 Literature Review

The researches of service quality can be roughly divided into three stages [3]: the first stage is the initial stage of service quality, mainly focused on the definition and theoretical construction of service quality; the second stage is the theory testing era, using empirical methods to analyze the factors affecting service quality and construct evaluation models of service quality on the basis of theoretical and conceptual studies; the third stage is the theory applied and interdisciplinary research era, with the objectives of researches in this stage mainly concentrating on solving practical problems, improving efficiency and obtaining useful advices for market competition.

2.1 Initial Stage

Service quality is subjective perceived quality, not objective [4], therefore, researches on service quality hold different views on the definition and meaning of service quality and the most representative achievement of the initial stage is the definition of the concept of service quality. Levitt [5] is the first person to define service quality theoretically and he believes that service quality is that the service result can meet the set dimension. Sasser believes that the quality of service includes three levels: equipment, materials and personnel, and in addition to service results, it should also include the process of service [6]. Drawing on the thought of the division of product perceived performance [7]. Gronroos [8] puts forward the concept of customer perceived service quality, which considers that customer perceived service quality is composed of two variables: perceived service performance and expected service quality, and divides service quality into functional quality and technical quality. The concept of customer perceived service quality proposed by Gronroos and its difference structure lay a theoretical foundation for further research in the field of service quality.

In summary, this stage of researches is mainly focused on the definition of a single concept and the further division of service quality. There are few researches on the measurement of service quality and the relationship between service quality and its influencing factors.

2.2 Theory Testing Stage

With the maturity of the basic theory of service quality, more and more researches concentrate on the factors of service quality and use empirical methods to prove it. The most representative results of the theory testing stage are service gap model, SERVQUAL model and SERVPEFF model. Based on the research of Gronroos, the Parasuraman, Zeithaml & Berry (PZB) research portfolio in the United States proposes a service quality gap model, which divides the core gap between service perceived performance and service expectation into understanding gap of managers, quality standard gap, service delivery gap and service communication gap. Further researches are carried out based on gap model, and SERVQUAL model is put forward [9]. The original SERVQUAL model divides the factors related to service quality into 10 dimensions. Later, the research reduces the 10 dimensions to 5 dimensions: tangible, reliability, responsiveness, assurance and empathy [10]. SERVQUAL model has laid the theoretical foundation for service quality evaluation, and most of the later researches are based on the theoretical improvement or application researches of SERVQUAL model. [11] revised the model from the aspects of questions and mood, and the number of samples, reliability and validity of the revised model are better than those of the original model. [12] put forward the viewpoint of separating the theory of service quality from customer satisfaction, believing that the key to separating the two theories is whether to use "expectation". [12] holds the view that customer satisfaction is the result of comparing perceived service with expected service, while service quality is the result of comparing ideal service with appropriate service. SERVPERF is another classical evaluation model of service quality. The measurement and questionnaire are inherited from the SERVQUAL model, which only defines service quality as customer perceived service performance rather than difference comparison [13]. Later, a comparative empirical research is made to prove the superiority of SERVPERF model [14].

This stage of researches is mainly around PZB's 5 dimensions, 22 indicators, but with the continuous development of the service industry, this stage of researches in the field of adaptability is declining. Different characteristics of different fields lead to more authoritative models such as SERVQUAL being questioned, and scholars in the academic community have not reached a consensus, so relevant researches in the field of service quality need more detailed evaluation in different fields. Besides, the approaches of empirical researches are primarily questionnaires or surveys, which are time-consuming and small-scale.

2.3 Theory Applied Stage

Researches at this stage have more extensive research directions, more diverse research methods and more diversified research data, and empirical researches on service quality in various service industries have attracted more and more attention. The characteristics of researches at this stage are its practicability and guidance.

Fuzzy set method is always used in service quality evaluation. Benítez et al. [15] put forward a dynamic evaluation method of hotel service quality, which uses fuzzy number to resolve the ambiguity of concepts related to subjective judgements and develop a service performance index by TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution) method. Xiao Wei et al. [16] put forward the FCE (Fuzzy Comprehensive Evaluation) model. It is helpful to reveal the problems of service quality and put forward more economical methods to improve service quality by using fuzzy comprehensive evaluation to transform the target qualitatively to quantitatively, mining the causal relationship between the evaluation indexes and establishing a fuzzy cognitive map. Aydin [17] combines statistical analysis, fuzzy trapezoidal numbers and TOPSIS together to evaluate service quality of rail transit systems.

Apart from fuzzy set methods, traditional methods like IPA (Importance-performance analysis), TOPSIS, Content Analysis and new methods like neural network, sentiment analysis are also introduced in service quality evaluation. You Liping et al. [18] take online reviews as research data and adopt TOPSIS to evaluate service quality by using sentiment analysis technology. Wang Huiling et al. [19] use "ROST Content Mining" tool to analyze the comments of travelers, construct the social network and semantic network between feature words, extract the evaluation indexes, and then construct the evaluation model with content analysis method. Chen Lifei et al. [20] put forward an innovative framework that integrates the advantages of IPA, the zone of tolerance concept, and Kano's model.

In this stage, the research methods, data sources and angles of service quality evaluation are more abundant. Researchers use different methods to build appropriate models for different situations. With the continuous development of online reviews, the most significant difference between this stage and previous researches is the change of data sources. More and more research will replace the previous questionnaire survey with online reviews as data sources. In addition, with the development of big data and natural language processing technology, besides the commonly used content analysis method, the method of text semantic analysis has also been applied by many researchers, which proves the feasibility of interdisciplinary research.

2.4 Literature Summary

At present, the research in the field of service quality evaluation mostly focuses on the improvement of theory and the application-oriented research of service quality evaluation index system and few researches involve the technical route and framework scheme of service quality evaluation. Based on the current research status and problems in the practice of service quality evaluation, this research aims to construct the technical framework of service quality evaluation.

3 Architecture and Theory

In the research, the technical framework of service quality evaluation is designed with the goal of "efficient, intelligent and scientific". Three basic principles of modularization, standardization and intellectualization should be followed in the design process.

Modularization refers to the division of service quality assessment tasks into several "high cohesion, low coupling" modules, with clear functional division between them. Service quality evaluation is a complicated process with many links and long data chains. Modularization is a process that combines closely related parts to form a module with clear responsibilities. Data transfer between modules is carried out through reserved data interfaces. The process of service quality evaluation is clearer, the coupling degree between modules is reduced, and the data flow direction is simpler and clearer.

Standardization is also called streamlining, which integrates the process of service quality evaluation into a repeatable standard process which has standard input and output information. Standardization is the basis of automation and systematization of service quality evaluation, as well as an important guarantee of information transmission between modules. Standardization principle standardizes service quality evaluation process into a simple and repeatable system process, which makes service quality evaluation more efficient.

Intelligence means that service quality evaluation can be triggered automatically and analyzed intelligently without manual intervention. Intelligence is the symbol of the new stage of service quality evaluation research, as traditional service quality researches have a low degree of intelligence and too much manual intervention. With the continuous development of artificial intelligence technology and its introduction into the field of service quality evaluation, the intelligence of data processing can be improved, and then the service quality evaluation can be more efficient and intelligent.

3.1 Architecture

Based on the principles of modularization, standardization and intellectualization, and aiming at standardizing the critical links of "data acquisition data processing and data analysis and modeling", a systematic technical framework (Figure 1) for service quality evaluation is proposed, which integrates automatic data acquisition, intelligent data processing and real-time data analysis.

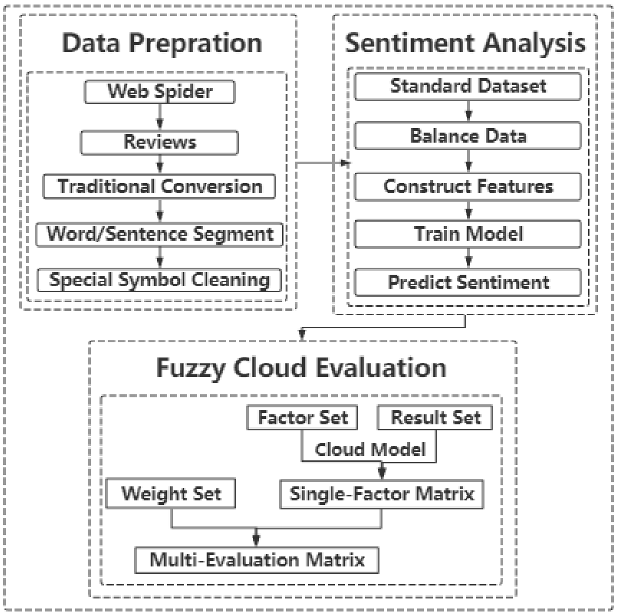

Figure 1. Overall Architecture of the Research.

From the point of view of data life cycle, the process of service quality evaluation is mainly divided into three critical links: data preparation, data processing and analysis modeling.

Data acquisition is the data generation stage in the data life cycle, and the collected data is the data source of service quality evaluation. Data is the core of service quality evaluation, and the quality of data has a great impact on service quality evaluation. In order to obtain high-quality service quality evaluation data, data acquisition link needs to solve two key problems: what kind of data to acquire and how to acquire data quickly and efficiently. “What kind of data to acquire” refers to the specific data characteristics such as the type and scope of data to be acquired, and different data types and ranges are different for different problems in different periods. Quick and efficient acquisition of data requires the use of automated tools and technologies to replace manual data acquisition to achieve faster data acquisition and renewal.

Data processing is the processing stage in the data life cycle. In this stage, the original data undergoes a series of standardized processing and transformation to form standard structured data. Data processing is an important data conversion center in service quality evaluation. Traditional data processing methods of service quality evaluation mostly use content analysis method and other low-efficiency methods which need manual assistance. With the increasing amount of data, these low-efficiency data processing methods relying on manual assistance can no longer meet the growing needs of service quality evaluation. The results of data processing are greatly influenced by subjective factors and then objectivity is difficult to guarantee. Only by getting rid of manual dependence, can the data processing process be more objective and efficient. Therefore, the core of improving the efficiency and objectivity of data processing lies in intelligent data processing.

Analytical modeling is the stage of data value generation link in data life cycle. The goal of this stage is to convert standard data into evaluation result of service quality evaluation. The task of analysis and modeling is to organize and apply the data reasonably and then get the service level of the enterprise. The application of data in the process of analysis and modeling are embodied in the selection of models and the combination of models. Reasonable selection and combination of application models can establish a reasonable service quality evaluation model and obtain the actual service level of enterprises.

3.2 Data Preparation

Data preparation module is to provide the standard data needed by the service quality model. This module is mainly divided into two tasks: web text capture and text preprocessing.

The task of web text capture is to use web crawler technology to obtain external web page information, and then to obtain the required web comment information through web page parser. WebMagic framework based on Java language is used in the research of web text grabbing technology. WebMagic framework is mainly composed of 4 modules: Schedule, Downloader, PageProcessor and Pipline (Figure 2).

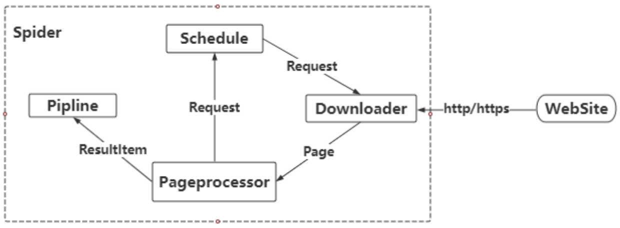

Figure 2. Architecture of WebMagic.

Schedule module is the task scheduler, which is responsible for scheduling and processing URLs. Schedule module can not only queue the target URL in order to schedule, but also parse the new URL into the queue. The design of Schedule module makes WebMagic crawler technology more flexible and changeable. Downloader module is the task downloader, which is responsible for sending web page requests and downloading the returned HTML pages or JSON objects. The data downloaded by Downloader module is the original request data and the basis of subsequent data analysis. The PageProcessor module is a data parsing component based on JSOUP, which is responsible for parsing the data downloaded by Downloader. Pipline module is a result processor, primarily responsible for data calculation and data persistence.

The original data obtained from websites are not standard, so it is necessary to preprocess them before transferring it into structured tags. Firstly, some of the reviews are traditional Chinese, which affects the result of sentiment analysis, so converting traditional Chinese to simplified Chinese is necessary. Secondly, special symbols should be handled because they may affect word segmentation and feature selection. Finally, word segmentation is the key of the preprocessing stage, for word is the basic unit of sentiment analysis and feature selection. LTP segmentation is used to segment reviews into words in this research.

3.3 Fine-Grained Sentiment Analysis

Sentiment analysis task is an important field in natural language processing. At present, there are many sentiment analysis studies at the text level in the research of emotional analysis tasks. The emotional analysis at the text level is based on the hypothesis that all emotions expressed in a text are identical. However, such hypothesis is usually not valid for online reviews, such as "the environment is good, and the service is poor", which obviously does not conform to the emotions at the text level. The hypothesis of the analysis is that text-level sentiment analysis is too coarse in mining customer satisfaction, which will result in sentiment information loss and not suitable for this research. In order to discover the customer's emotional tendency in one aspect, the fine-grained attribute-based sentiment analysis method is adopted in the research.

The goal of fine-grained sentiment analysis task is to mine the emotional tendency of comment text in terms of specified attributes. There are 3 elements in the task: comment text, attributes and emotional tendency. Therefore, fine-grained sentiment analysis task is defined in mathematical language as follows:

Given comment text $c_k = (w_{k1}, w_{k2}, \ldots, w_{kn})$, attribute set $\mathrm{a} = (a_1, a_2, \ldots, a_t)$, and sentiment set $\mathrm{s} = (s_1, s_2, \ldots, s_q)$, $f(\cdot)$ is a mapping from c_k and a to s, s_{ki} is the result of c_k in each element of a and $s_{ki} = f(c_k, a_i)$. The task is to figure out $f(\cdot)$ and s_{ki}.

In this research, fine-grained sentiment analysis task is regarded as a multi-classification task based on attributes. The essence of the task is to train a classifier in every aspect of the evaluation object attribute set. The set of all classifiers is the mapping model in fine-grained sentiment analysis task. Fine-grained sentiment analysis task is divided into two sub-tasks: evaluation object recognition and attribute-based sentiment judgement. In order to merge the two tasks into one task, the tag of "not mentioned" is introduced when constructing classifiers in addition to positive, negative and neutral tags. Therefore, fine-grained sentiment analysis task can be transformed into training multiple classifiers of four categories.

Each review contains limited evaluation attributes. The lower the probability of attributes being mentioned, the higher the probability of attributes being judged as "not mentioned" and the more obvious the problem of data imbalance. In fine-grained sentiment analysis tasks, the actual meaning of data with emotional labels is greater than that of "not mentioned" label, so it is necessary to balance data.

3.3.1 Minor Class on Seed Sentence Algorithm

There are 3 common problems in origin data. Firstly, most of the origin reviews refer to limited attributes, so many attributes are marked as “not mentioned” label. Secondly, some attributes often appear in reviews, while others hardly appear, which results in more error of the model. To solve the problem of unbalanced dataset, minor class on seed sentence (MCSS) algorithm is proposed, which comes from the ideal of combination features. For example, the review A does not mention an attribute, and review B expresses the sentiment inclination to the attribute, if the two reviews are combined, the sentiment inclination on the target attribute is consistent with the inclination of review B to the attribute.

There are 2 important parts of the MCSS algorithm. One aspect is to decide how much data needed to constructed, the other is the construction of seed sentences. Training with different data construction rate a, a is equal to the number which the DP-FastText performance better result. The main process of constructing seed sentences by artificial method is as follows: first, we find the comment text labeled -1, 0, 1 of the target attribute from the data set, then select k comment texts to filter manually, and find the sentences only associated with the target attribute and add them to the seed sentences of the attribute. Because manual selection is slow, the number of seed sentences is generally less. In order to improve the quality of the constructed data, manual seed selection should obey the following 5 principles:

1）Explicitness which means selected seed sentences must be able to clearly judge sentiment tendencies, without ambiguity.

2) Representativeness which means selected seed sentences are commonly used to describe the target attributes.

3) Diversity which means the expression of seed sentences is as diverse as possible.

4) Monism which means seed sentences had better only involve target attribute and reduce the interference of irrelevant words.

5) Generality which means that the specific words in seed sentences are replaced by generic words.

The concrete algorithm does as follow (Table 1), while s referring to seed sentence set, c referring to review set and se referring to sentiment label, a referring to the data constructed rate.

Table 1 MCSS Algorithm

input: $\mathbf{s} = \{\boldsymbol{s_1}, \dots, \boldsymbol{s_m}\}$, $\mathrm{c} = \{\boldsymbol{c_1}, \dots, \boldsymbol{c_n}\}$, $\mathbf{se} = \{\boldsymbol{se_1}, \dots, \boldsymbol{se_n}\}$, a

1. calculate N_0, N_1, N_2, N_3 where $se_i = -2$, $se_i = -1$, $se_i = 0$ and $se_i = 1$
2. construct subset c_{un} where $se_i = -2$
3. construct subsets s_1, s_2, s_3 where $se_i = -1$, $se_i = 0$ and $se_i = 1$
4. define $\mathrm{t} = \{t_1, t_2, t_3\}$, $t_i = a * \frac{(N_0 - N_i)}{len(s)}$
5. for t_i in t
    if $t_i > 0$
        define nc_i to save new reviews
        construct rc_{un} by t_i piece of reviews randomly selected from c_{un}
        for rc in rc_{un}
            for s in s_i
                add $nc = rc + s$ into nc_i
6. nc $= c \cup nc_1 \cup nc_2 \cup nc_3$

7. output nc

3.3.2 DP-FastText Model

DP-FastText model is a fine-grained sentiment analysis model based on FastText model and integrated with semantic information. Like FastText model [21], DP-FastText model is a three-layer neural network model, which consists of input layer, hidden layer and output layer (Figure 3).

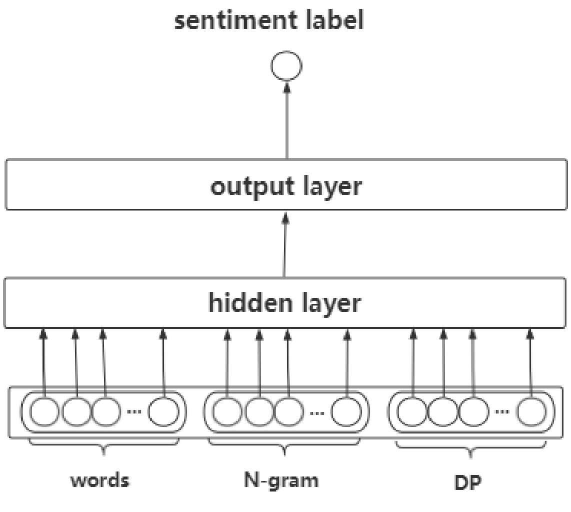

Figure 3. DP-FastText Model.

The input layer of the model is the feature representation layer of the comment text. The input layer is responsible for extracting the features of the comment text and expressing the whole comment text with text features. The text features in the study are composed of three parts: basic word features, N-Gram features and DP features.

Words are the basic features of Chinese texts. In this study, the reviews are transformed into a set of words by means of text segmentation, custom dictionary and stop word filtering and the set of words is the basic feature of reviews.

N-Gram features can regulate the influence of word order on the final classification results. The order of words has a great influence on the sentiment tendency of the whole sentence. If the set of words is used as the input level features and the order of words is ignored, some semantic information will be lost and the classification effect will be deviated. For example, only the expression of "not like" can express negative feelings, while "not" and "like" as two separate characteristics do not necessarily express negative feelings. N-Gram is a language model-based algorithm, which uses sliding window technology to slide text in order to form a word sequence of length N [22]. In order to solve the influence of word order on sentiment analysis, N-Gram features with better training effect are mixed into DP-FastText model to make full use of word order information.

DP features represent the smallest sentiment unit. N-Gram features can represent the common features of adjacent words, but some words which are not adjacent can also reveal sentiment tendencies, as there are specific dependencies to represent modified relation. In order to express the phrase features of such closely related words that meet certain conditions, LTP platform of Harbin Institute of Technology is used in the study to analyze dependency syntax and extract qualified phrases to form DP features.

LTP platform integrates abundant global features and clustering features based on the dependency parsing using neural network model [23], and the speed and effect of analysis can be guaranteed. Dependency parsing is used to extract features on the basis that the form of Chinese text is constrained by grammar, sentence pattern and word. It is helpful to predefine rules of dependency among grammar, sentence pattern and part of speech to capture the key information in the text. In the fine-grained sentiment analysis task, the key information captured by dependency parsing technology is the sentiment evaluation unit. The sentiment evaluation unit is an N-tuple composed of sentiment words, degree words, attributes of evaluation objects and negative words. By observing the grammatical structure and part-of-speech information of many texts, the following 2 rules for extracting sentiment evaluation units are predefined.

Rule 1: Core words are adjectives. If the dependent word is a noun or verb and the subject-predicate relationship between the core word and the dependent word is satisfied, then the core word is added to the sentiment evaluation unit. If there are nouns or verbs in the first two windows of the word satisfying centering relation, these words and the target words are combined in order and then added to the sentiment evaluation unit. Otherwise, the dependent words are directly added to the sentiment evaluation unit. If the dependent word is an adverb, the core word satisfies the adverbial-middle relationship with the dependent word, then the adverb is added to the sentiment evaluation unit.

Rule 2: Core words are verbs. If there are nouns or verbs satisfying the subject-predicate relationship and the verb-object relationship at the same core word and there are nouns or verbs in the first two windows of the dependent word satisfying centering relations, these words and target words are merged in order and then added to the sentiment evaluation unit, otherwise they will be added directly. If the dependent words are adverbs and the core words satisfy the adverbial-neutral relationship with the dependent words and are negative words, the adverbs are added to the sentiment evaluation unit.

When N-Gram features and DP features are added to the input layer, the amount of data in the input layer increases exponentially and a large amount of data has an impact on the performance of the model and the training effect. In order to reduce the amount of data and the impact of noise data, the feature threshold is set to filter low-frequency N-Gram and DP features.

The hidden layer of the model carries out the operation of adding and averaging the features. The features represented by one-hot are simply weighted averaging and transformed into the output of the hidden layer. Finally, the output layer of the model calculates the conditional probability of each leaf node in the Huffman tree, and chooses the category with the highest conditional probability as the final label output.

3.3.3 Fuzzy Cloud Evaluation Module

Data preparation module and fine-grained sentiment analysis module are both data preparation stages in the process of service quality evaluation. They are all for acquiring data and converting data into the standard from needed by the fuzzy cloud evaluation module. Fuzzy cloud evaluation module is the real module to evaluate service quality. Fuzzy cloud evaluation module uses the fuzzy comprehensive evaluation model based on cloud model to evaluate service quality.

Fuzzy cloud evaluation model is based on the fuzzy comprehensive evaluation model, and uses the cloud model as membership function to calculate the weight set and single factor evaluation matrix of the model. The module mainly consists of establishing evaluation factor set and comment set, determining membership function, constructing single factor evaluation matrix, determining weight set, fuzzy synthesis and normalization. Factor set and comment set are constructed by fine-grained emotional analysis based on online comments, which is composed of factor set and comment set. Factor set $\mathrm{U} = \{u_1, u_2, \dots, u_n\}$ denotes a set of attributes related to service quality evaluation. The comment set $\mathrm{V} = \{v_1, v_2, \dots, v_m\}$ denotes the evaluation level of the service. In the fuzzy cloud evaluation model, the initial set of weights is set by analytic hierarchy process. By comparing the importance of two factors, the eigenvalues are obtained, and the work out the initial weight vector $A_0 = (a_{01}, a_{02}, \dots, a_{0n})$.

Finally, extract the digital features of each factor (Ex, En, He) using reverse cloud generator, and get the final evaluation weight A (formula (1)).

$$A = \begin{bmatrix} Ex_{a_1} & En_{a_1} & He_{a_1} \\ Ex_{a_2} & En_{a_2} & He_{a_2} \\ \vdots & \vdots & \vdots \\ Ex_{a_n} & En_{a_n} & He_{a_n} \end{bmatrix}^T \tag{1}$$

From formula (1), each weight coefficient has a certain degree of fuzziness and randomness, evaluation factors can be scored as Ex_{a_i}, the score of different people lies in $[Ex_{a_i} - 3 \times En_{a_i}, Ex_{a_i} + 3 \times En_{a_i}]$ and He_{a_i} further reflects the random of subjective evaluation [25].

The construction of single factor evaluation matrix is similar to the weight set. Instead of synthesizing the membership degree of each factor at each evaluation level, it uses the digital characteristics of cloud model (Ex, En, He) to fuzz the evaluation level under each factor into a qualitative concept. The single factor evaluation matrix is shown in formula (2):

$$R = \begin{bmatrix} Ex_1 & En_1 & He_1 \\ Ex_2 & En_2 & He_2 \\ \vdots & \vdots & \vdots \\ Ex_n & En_n & He_n \end{bmatrix} \tag{2}$$

After confirming A and R, appropriate composition operator should be chosen, and applied to combine A and R to get the fuzzy comprehensive evaluation model. After selecting the operator and applying the cloud operation rules, the evaluation model B can be expressed in the form shown in Formula (3)

$$B = \begin{bmatrix} Ex_{a_1} \times Ex_1 + \cdots + Ex_{a_n} \times Ex_n \\ \sqrt{\left(|Ex_{a_1} Ex_1| \sqrt{\left(\frac{En_{a_1}}{Ex_{a_1}}\right)^2 + \left(\frac{En_1}{Ex_1}\right)^2} \right)^2 + \cdots + \sqrt{\left(|Ex_{a_n} Ex_n| \sqrt{\left(\frac{En_{a_n}}{Ex_{a_n}}\right)^2 + \left(\frac{En_n}{Ex_n}\right)^2} \right)^2}} \\ \sqrt{\left(|Ex_{a_1} Ex_1| \sqrt{\left(\frac{He_{a_1}}{Ex_{a_1}}\right)^2 + \left(\frac{He_1}{Ex_1}\right)^2} \right)^2 + \cdots + \sqrt{\left(|Ex_{a_n} Ex_n| \sqrt{\left(\frac{He_{a_n}}{Ex_{a_n}}\right)^2 + \left(\frac{He_n}{Ex_n}\right)^2} \right)^2}} \end{bmatrix} \tag{3}$$

4 Empirical Researches

The data of empirical study in this paper is composed of standard dataset of reviews in restaurant from AI Challenger 2018 Competition and 28,358 reviews of 5 restaurants in Wuhan from Meituan website. The standard dataset of AI Challenge is used to train the fine-grained sentiment analysis model, and the data from website are used to verify the effectiveness of the model proposed in this paper.

The whole process of the empirical study can be divided into the following steps. Firstly, the captured reviews from website are transformed into standard input data through the text preparation module of service quality evaluation, and then the DP-FastText fine-grained sentiment analysis model, which has been pretrained with the standard dataset from AI Challenger 2018, is used to convert the data into sentiment label output on each single evaluation attribute. Then the original importance degree of evaluation attributes obtained from the survey is transformed into the weight set by using analytic hierarchy process. Finally, the fuzzy cloud evaluation model is introduced to combine the weight set of evaluation attributes and the sentiment labels of evaluation attributes into the final service quality evaluation results.

In order to prove the validity of the model proposed in the paper, the research compares the results of the empirical research with those of the network rating of the Meituan and Dianping. It is the similarity between the empirical research results and the network rating results that matters to the validity of the model in the paper, and the empirical results are comprehensively analyzed from five perspectives: evaluation effectiveness, evaluation efficiency, personalization, dynamic and intellectualization.

4.1 Weight Set Results

In empirical research, according to the indicators of restaurants from the company of Meituan in AI Challenger 2018 and the index division proposed by Wall and Berry [26], 5 coarse categories and 17 fine-grained categories of restaurant indicators are given on this research. To make the experiment clearer, these indicators are encoded by C1~C15(Table 2).

Table 2 Indicator Table

Code	Meaning
C1	Service Wait Time
C2	Service Waiters Attitude
C3	Service Parking Convenience
C4	Service Serving Speed
C5	Price Level
C6	Price Cost Effective
C7	Price Discount
C8	Environment Decoration
C9	Environment Noise
C10	Environment Space
C11	Environment Cleanness
C12	Dish Portion
C13	Dish Taste
C14	Dish Look
C15	Dish Recommendation

In order to identify the importance of these indicators of restaurant, 10 users composed of 5 experts and 5 common users are invited to attend the survey of comparison of the importance of pairwise. Then the results of 10 users were modeled by analytic hierarchy process (AHP) and the software of Tianjin University School of Management was used to carry out the AHP experiment. The result of one user is taken as the example to analyze the process of weight set construction.

Firstly, a judgement matrix is constructed according to the comparison of every two indicators' importance of restaurant filled in by users. Since the evaluation attributes in the study are divided into two layers, it is necessary to construct the judgement matrix hierarchically. First, the judgement matrix of the evaluation result layer (A) and the four coarse indicators of service (B1), price (B2), environment (B3) and dishes (B4) is constructed (Table 3). According to the result of judgment matrix A-B, the maximum eigenvalue of the matrix is 4.115, and the consistency test index CR=0.042<0.1. Therefore, the consistency test of the judgment matrix is passed, which can be used to identify the importance of evaluation attributes.

Table 3 A-B Judgement Matrix

A	B1	B2	B3	B4
B1	1	3	1/3	1/5
B2	1/3	1	1/5	1/5
B3	3	5	1	1
B4	5	5	1	1

Then, the judgement matrix of each coarse indicators of layer B is divided into more fine-grained indicators. The maximum eigenvalues of each judgement matric from B layer to the next layer (C) are 4.082, 3.039, 4.073 and 4.120, and CRs of the four decision matrices are 0.030, 0.033, 0.027 and 0.044, respectively, which are all less than 0.1. Therefore, the results of the four decision matrices all pass the consistency test and can be used for the next process.

Combining the single ranking results of A-B and B-C layers, we can get the final ranking results of one user. In the experiment, 10 users are involved in identifying the weight of indicators, and the results of 8 users pass the consistency test, so we combine the 8 users' result together by cloud model. The result of original weight set is shown in formula (4). The columns represent expectation, entropy and hyper entropy from left to right and lines represent indicators from C1 to C15.

$$A=\begin{bmatrix} 0.029 & 0.0219 & 0.009 \\ 0.077 & 0.0495 & 0.0061 \\ 0.0102 & 0.0038 & 0.0017 \\ 0.0392 & 0.0192 & 0.0039 \\ 0.0316 & 0.0198 & 0.0089 \\ 0.1007 & 0.0689 & 0.0379 \\ 0.0139 & 0.0076 & 0.0036 \\ 0.0184 & 0.0092 & 0.0046 \\ 0.0242 & 0.0175 & 0.0055 \\ 0.0368 & 0.0201 & 0.0113 \\ 0.1306 & 0.0816 & 0.0421 \\ 0.078 & 0.0219 & 0.0025 \\ 0.287 & 0.0548 & 0.0185 \\ 0.065 & 0.0517 & 0.0167 \\ 0.0582 & 0.0314 & 0.0161 \end{bmatrix} \quad (4)$$

4.2 Single-factor Evaluation Results

After the weight set is constructed, 28358 reviews from 5 restaurants are processed, and then the sentiment labels of reviews in each evaluation attribute by using the pre-trained DP-FastText model. The sentiment labels predicted by the DP-FastText model in 5 restaurants can be expressed by a triple T and T is defined as follows (formula (5)). In the formula: N_{neg} denotes the total number of reviews that refer to the negative sentiment for the indicator, N_{neu} denotes the total number of reviews that refer to the neutral sentiment for the indicator, and N_{pos} denotes the total number of reviews that refer to the positive sentiment for the indicator. The complete results of the transformation of reviews are shown in Appendix A, and Figure **4**~Figure 6 show some of the important indicators' results.

$$T=[N_{neg},N_{neu},N_{pos}] \quad (5)$$

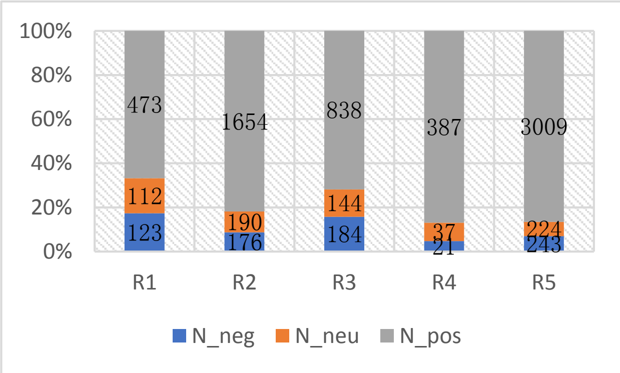

Figure 4. Service Waiters Attitude Sentiment Distribution.

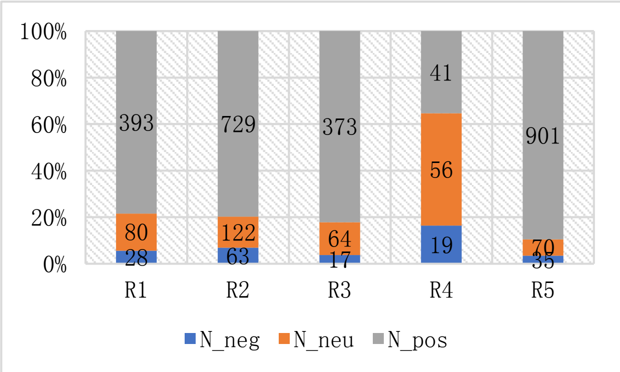

Figure 5. Environment Cleanness Sentiment Distribution.

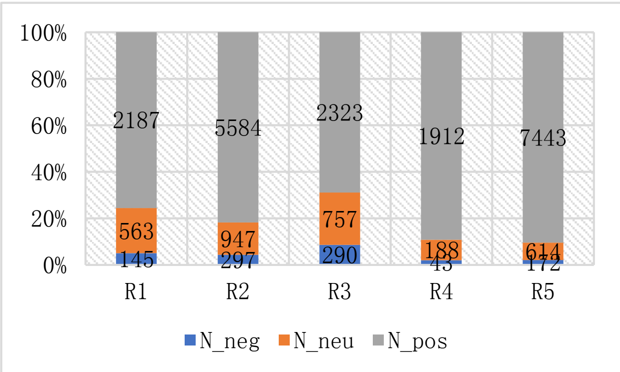

Figure 6. Dish Taste Sentiment Distribution.

According to the results of sentiment distribution, given that the score of each negative label is 1, the score of neutral label is 3, and the score of positive label is 5, the results of each restaurant's evaluation attributes based on the DP-FastText model are input into the cloud model, and three digital characteristic matrices of the cloud model of each restaurant on the 15 evaluation attributes of C1~C15 can be obtained. According to the evaluation result (Appendix B), 5 restaurants are compared on each factor. The service quality evaluation results of 5 restaurants are ranked according to the priority of Ex>En>He, and the ranking results are obtained as shown in Table 4.

Table 4 Single Factor Ranking Results

Indicator	Ranking Result
C1	R1>R2>R3>R5>R4
C2	R4>R5>R2>R3>R1
C3	R3>R1>R5>R2>R4
C4	R4>R2>R1>R3>R5
C5	R4>R1>R2>R5>R3
C6	R4>R5>R2>R1>R3
C7	R2>R5>R4>R1>R3
C8	R5>R3>R1>R2>R4
C9	R5>R1>R3>R2>R4
C10	R5>R2>R3>R1>R4
C11	R5>R3>R2>R1>R4
C12	R5>R4>R2>R3>R1
C13	R5>R4>R2>R1>R3
C14	R4>R5>R2>R3>R1
C15	R5>R4>R1>R2>R3

Comparing the service quality of R1~R5 restaurants shown in Table 4 with the evaluation of R1~R5 restaurants in the website of www.dianping.com, the conclusion can be drawn that the method proposed in the study is reasonable and effective, and can meet the individual decision-making needs. The specific content can be understood from two aspects: rationality and individuality.

4.2.1 Rationality Analysis

In terms of the indicator of taste, the empirical results are basically consistent with the results of Dianping, but due to different data sources, the results will also be kind of different. The result of Dianping is R4 > R5 > R2 > R3 > R1, while the result of empirical research is R5 > R4 > R2 > R1 > R3. The restaurants with better taste are R4 and R5, while the restaurants with worse taste are R1 and R3, and the taste of R2 belongs to the intermediate level. However, there are some differences between empirical results and Dianping inside the echelon. R4 in Dianping has higher taste scores than R5 and R3 has higher taste scores than R1. However, the empirical research results show that R5 has better taste than R4 and R1 has better taste than R3. The main reason is that the data in the study are from Meituan and it is the different data source that result in the difference On Meituan website, the overall score of R1 is 5, R3 is only 3.9, R4 is 4.5 and R5 is 5. Therefore, it is reasonable that the taste of R1 is better than R3 and R5 is better than R4.

In terms of the indicator of environment, there is little difference between the empirical results and Dianping. The environmental ranking of the 5 restaurants in Dianping is R5 > R3 = R2 > R1 = R4. In the empirical study, R5 ranks the highest among the four sub-items of environmental indicators and R4 ranks the lowest, which is consistent with Dianping. In the indicator of cleanness, which takes the highest proportion, the ranking result is R5 > R3 > R2 > R1 > R4, while in the indicator of space, a relatively higher proportion of environmental indicator, the ranking result is R5 > R2 > R3 > R1 > R4. By synthesizing 4 sub-indicators, the ranking of environmental indicators of 5 restaurants in the empirical study is R5 > R3 > R2 > R1 > R4, which is basically consistent with the results of Dianping.

In terms of the indicator of service, the empirical results are consistent with the general trend of Dianping. The ranking result of the 5 restaurants in Dianping is R4 > R5 = R3 > R2 > R1. In the empirical research, the importance of service attitude is obviously higher than other factors among the four sub-indicators related to service, therefore, the ranking result of service attitude should be like the ranking of service attitude. In the experiment, the ranking of service attitude is R4 > R5 > R2 > R3 > R1, which is generally consistent with that of the Dianping. Basically, the results of R2 and R3 are somewhat different from those of Dianping, but the service scores of R2 and R3 restaurants in Dianping don't vary too much, so it is normal that such differences exist.

4.2.2 Individuality Analysis

At present, Dianping is one of the most authoritative websites in the catering industry. It accumulates a large amount of restaurant reviews and becomes an important source of information for many customers to choose restaurants. On the website of Dianping, comment information is mainly divided into two parts: structured scoring and review texts. When portraying restaurants, the website mainly uses structured scoring information to evaluate restaurants from three aspects of taste, service and environment. The scope of evaluation is limited, the granularity is relatively coarse, and it is unable to understand more detailed information, therefore, it cannot fully meet personalized needs.

In the empirical research, 4 categories (service, price, environment and dishes) are extracted, and 15 categories of indicators affecting restaurant service quality were used to evaluate service quality. Compared with the results of Dianping, the dimensions of evaluation in empirical research are broader and the granularity is finer. For example, the coarse-grained evaluation indicator of service can be subdivided into 4 aspects: waiting time, service attitude, parking service and service speed.

In the empirical research, customers and supervisors can choose the appropriate dimension to make personalized decision and supervision according to their own needs from the indicators that affect the quality of restaurant service. Compared with the customized scoring method in Dianping, the method in the empirical research is more flexible and can provide more decision-making information for customers and supervisors, so that customers and supervisors can have a better understanding of the service quality of restaurants in overall aspects and make more scientific and reasonable decisions and judgments.

4.3 Comprehensive Evaluation Results

Single factor evaluation can help learn the service of restaurant in a more specific attribute, yet it is also important to understand the overall evaluation of service quality of restaurant, because the overall service quality represent the image of the restaurant and it matters a lot when customers choose restaurants. In order to obtain the overall service quality, weight set and single-factor matrices are combined by the synthesis operator. The final service quality evaluation results of 5 restaurants are shown in Table 5.

Table 5 Comprehensive Service Quality Results

	Results	Meituan	Dianping
R1	(4.2226,0.4459,0.1070)	5.0	(7.4,7.7,7.4)
R2	(4.3341,0.4461,0.1071)	3.7	(8.3,8.3,8.2)
R3	(4.0102,0.4527,0.1052)	3.9	(8.0,8.3,8.3)
R4	(4.2241,0.4252,0.0961)	4.5	(9.0,7.7,8.5)
R5	(4.5001,0.4495,0.1085)	5.0	(8.5,8.4,8.3)

According to Ex, the service performance of 5 restaurants is R5 > R2 > R4 > R1 > R3. The service quality of R2, R4 and R5, is better than other 2 restaurants, because the expectation values of R2 and R5 are higher, while the average value of R4 is also at a higher level and is stable, and the random subjectivity is weaker. R1 and R3, have relatively poor service quality in terms of both average and stability.

4.3.1 Comparative Analysis

Whether in the result of this study or the website of Meituan and Dianping, R5 performs best of all the restaurants, and R3 performances not so good in all 3 evaluation methods. R1 and R2 perform differently in 3 methods. On the whole, the results of this study are generally consistent with the results of the website of Dianping and the results of Meituan performs worse than other 2 evaluation methods.

The average ranking result of 5 restaurants in the Dianping is R5 = R4 > R2 > R3 > R1. The service quality of R2, R4 and R5, is better than that of R1 and R3. This result is consistent with the empirical research results, but the results of R1 and R4 are slightly different from those of the empirical research, which is mainly caused by 2 reasons. One reason is that the origin data is different，and the origin reviews of this study comes from Meituan, so it is normal that slight differences occur between the study and Dianping. The other reason is that dimensions in the research are more detailed than that of Dianping, and the weight of indicators is considered.

The ranking of the results of Meituan is R5 = R1 > R4 > R3 > R2, which is different from the results in the empirical research. Comparing the 2 ranking sequences, it is found that the major difference lies in R1 and R2. In order to compare the real service quality of R1 and R2, the comprehensive score of Dianping is introduced into the comparison. The comprehensive score of Dianping of R1 restaurant is the lowest among the 5 restaurants, which shows that the method of this research is closer to the result of Dianping than that of the Meituan. On the website of Dianping, the taste of R2 ranks third, lower than R4 and R5; the environment ranks second, lower than R5, service ranks fourth, lower than R3, R4 and R5. Therefore, R2, ranking third after R4 and R5, is closer to the results of this research than that of Meituan.

4.3.2 Case Analysis

In order to analyze the difference between the empirical results and the results of Meituan, R1 is taken as a case study to analyze the content and scoring of the reviews, and explore the reasons for the deviation between the results of Meituan and the evaluation results in the empirical research. Besides, the reliability of the evaluation results can also be proved by qualitative analysis.

1) Historical data leads to deviation

Historically accumulated data lead to differences in evaluation results. On the website of Meituan, the score of R1 is 5, and to better analyze the service quality of R1, 100 reviews on the first 10 pages of the group are selected, including 12 reviews by 1 star, 13 reviews by 2 stars, 23 reviews by 3 stars, 21 reviews by 4 stars and 31 reviews by 5 stars. Through the 100 reviews, it is obvious that the service quality of R1 is not as good as the overall score of 5. The reason why the restaurant scores 5 points is that there were a lot of 5-star reviews before, so the sensitivity of the score of the middle and bad reviews is not enough, and the reliability of the score obtained by simple average is not enough. In addition, there are many non-commentary ratings on Meituan score, which also results in the difference between the empirical results and Meituan. This part of the score is mostly 5 points, but this part of the score is less reliable than the score with text content.

2) The inconsistency data results in deviation

The inconsistent data between scoring and text results in different results. In some cases, the score of the restaurant is high, but the text content is of negative sentiments, which is obviously inconsistent between scores and text. This part of the user's score data is wrong, and this kind of score will have a certain impact on the quality of service evaluation results. There are also reviews which are obviously not related to the service quality of restaurants and the reliability of the evaluation content is not high. On the website of Meituan, this kind of score will still be included in the final score. However, in the method proposed in the study, the emotional label of each factor is identified as -2, indicating that no evaluation attribute is mentioned, which will be more reasonable.

3) Inconsistent dimensions lead to deviation

In some mainstream websites, the indicators are limited and some of the indicators does not have enough data to perform better model. However, the empirical research selects 4 coarse-grained indicators and 15 sub-indicators to evaluate service quality, while there is only overall score on Meituan. Besides, the empirical research has finer granularity, and considers the importance of different indicators on service quality, while the score of Meituan is the overall score, without considering the importance of indicators, which also results in the deviation of the evaluation results. The research evaluation is more comprehensive and meticulous.

By comprehensive analysis the similarities and cases, compared with the results of websites, the empirical results have greater reference value for customer decision-making, market supervision and service problem discovery, which eliminate historical data impact, have more fine-grained indicators and take the advantages of fuzzy operation on simple weighted average.

5 Conclusion and Discussion

5.1 Conclusion

With the continuous development of e-commerce and online comment websites, online comment has brought tremendous changes in the field of service quality management. Online reviews expand the data sources of service quality research, and the data acquisition mode of service quality research has changed from limited and inefficient offline mode to convenient, fast, wide-ranging and large-volume online mode, which greatly improves the efficiency and objectivity of service quality evaluation.

Based on the current situation of service quality evaluation, combining analytic hierarchy process, fuzzy comprehensive evaluation method and cloud model in the field of management, the fastText and dependency parsing technology in the field of natural language processing, this paper puts forward the service quality evaluation framework which integrates automatic data acquisition, intelligent data processing and real-time data analysis. The proposed service quality architecture is mainly divided into three modules: text data preparation, fine-grained sentiment analysis and fuzzy cloud evaluation. Through the empirical research, the following conclusions can be drawn.

1) Feasibility of Model

Through empirical analysis, compared with the evaluation of mainstream websites, the results are basically consistent with them, which prove the feasibility of the model proposed in this study. Besides more fine-grained evaluation factors and the importance of evaluation factors are considered in the study, so the model proposed in this study is also more reasonable

2) Efficiency of Model

It is very efficient to do the empirical research for it only takes about 30 minutes to collect 28358 reviews, and 48 minutes to pre-train the model in data processing, which is a one-time cost. It only needs training in the first use, and then the model can be directly read. In the experiment, it takes about 5 minutes to predict the sentiment labels on 15 evaluation attributes. In the process of evaluation, the most time-consuming is the process of determining the initial weight by analytic hierarchy process. It takes 0.5 days to invite 10 users to compare the importance and process the data of analytic hierarchy process. But it is not always necessary to spend the time. If the conditions do not change, this result can be reused. Therefore, it takes up to 0.5 days to 1 day to evaluate the service quality by using the empirical model, and the service quality evaluation can be completed in one hour in best condition. Compared with the traditional service quality evaluation methods, the efficiency of the model proposed in the research is greatly improved.

3) Dynamicity of Model

For restaurants with more historical data, the sensitivity of the newly added data in the reviews of the Meituan and Dianping is relatively small. It is not easy to detect the fluctuation of service quality in restaurants during a certain period. The model architecture proposed in this research can easily carry out real-time dynamic data analysis. It can evaluate the quality of service in a certain period (in days, in weeks or in months, etc.) and by comparing the results of each service quality evaluation, it can easily capture the fluctuation of service quality. Compared with the total evaluation, it is more likely to find the problems of service by incremental analysis.

4) Intelligence of Model

There are only two links in the construction of the empirical model that need manual participation. One is to train the fine-grained emotional analysis model in advance. And to balance the data set, we need to determine the seed sentences manually. The other is to determine the initial weights of each factor by using the analytic hierarchy process, which requires 10 users to compare the importance of factors, and then input the data into the system manually.

Only when the first evaluation of service quality or the change of evaluation rules and domains occurs, the above two links need manual participation, which are essentially one-time expenditure. In most cases, the evaluation of service quality does not need the above two links, and is a fully automatic process. Data acquisition is carried out by the crawler program, which can trigger automatically and regularly. The data processing process uses the pre-trained fine-grained emotional analysis model to automatically convert the collected data into sentiment labels. Then the data is automatically transmitted to the fuzzy cloud evaluation process to obtain the final evaluation results.

5.2 Deficiencies and Prospects

Although this study proposes a new model framework to solve the problem of service quality evaluation, and proves the rationality and superiority of the model, there are still some deficiencies or areas to be further studied in the research.

Firstly, the emotional tendency during data annotation is only divided into three emotional categories: negative, neutral and positive, in future research, we can refer to the Likert scale and the practices on mainstream comment websites to further mark the data to distinguish the emotional strength, which will be more reasonable.

Secondly, the model has a certain optimization space, and subsequent studies can try more in-depth learning models, or make other improvements to fasttext model, compare more methods, and get the optimal evaluation model for service quality evaluation.

Finally, in the process of empirical research, the research does not pay attention to the quality of the evaluation indicators, which also have a certain impact on the results of the model. Subsequent research can continue to pay attention to the impact of the quality of the evaluation indicators on the service quality and continue to optimize the model architecture

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by National Natural Science Foundation of China [grant number 70972138].

References

[1]. http://www.cca.cn/En/Index.html
[2]. https://www.iso.org/home.html
[3]. Prakash A and Mohanty R P (2013) Understanding service quality. Production Planning & Control 24(12): 1050-1065.
[4]. Garvin D A (1983) Quality on the Line. Harvard Business Review (61): 65-73.
[5]. Levitt T (1972) Production-Line Approach to Service. Harvard Business Review 50(5): 41-52.
[6]. Sasser W E, Olsen R P and Wyckoff D D (1978) Management of service operations: Text, cases, and readings. Allyn & Bacon.
[7]. Swan J E and Combs L J (1976) Product Performance and Consumer Satisfaction: A New Concept. Journal of Marketing 40(2): 25-33.
[8]. Gronroos C (1984) A Service Quality Model and its Marketing Implications. European Journal of Marketing 18(4): 36-44.
[9]. Parasuraman A, Zeithaml V A and Berry L L (1985) A Conceptual Model of Service Quality and Its Implications for Future Research. Journal of Marketing 49(4): 41-50.
[10]. Parasuraman A (1988) SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality. Journal of Retailing 64(1): 12-37.
[11]. Parasuraman A, Berry L L and Zeithaml V A (1991) Refinement and Reassessment of the SERVQUAL Scale. Journal of Retailing 67(8): 1463-1467.
[12]. Parasuraman A, Zeithaml V A and Berry L L (1994) Reassessment of Expectations as a Comparison Standard in Measuring Service Quality: Implications for Further Research. Journal of Marketing 58(1): 111-124.

[13].Cronin J J and Taylor S A (1992) Measuring Service Quality: A Reexamination and Extension. Journal of Marketing 56(3): 55-68.

[14].Cronin J J and Taylor S A (1994) SERVPERF Versus Servqual: Reconciling Performance-Based and Perceptions-Minus-Expectations Measurement of Service Quality. Journal of Marketing 58(1): 125-131.

[15].Benítez J M, Martín J C and Román C (2007) Using fuzzy number for measuring quality of service in the hotel industry. Tourism Management 28(2):544-555.

[16].Wei X, Luo X, Li Q, et al (2015) Online Comment-Based Hotel Quality Automatic Assessment Using Improved Fuzzy Comprehensive Evaluation and Fuzzy Cognitive Map. IEEE Transactions on Fuzzy Systems 23(1): 72-84.

[17].Aydin N (2017) A Fuzzy-based Multi-dimensional and Multi-period Service Quality Evaluation Outline for Rail Transit Systems. Transport Policy 55:87-98.

[18].You L P and Bai X Y (2017) Hotel Service Quality Measurement Based on Online Reviews Sentiment Analysis and TOPSIS. Information Science 35(10):13-17(in Chinese).

[19].Wang H L, Wu M Y (2015) Evaluating Home Inns in China Based on Social Network Analysis. Chinese Journal of Management 12(03):433-438(in Chinese).

[20].Chen L F, Chen S C and Su C T (2017) An innovative service quality evaluation and improvement model. The Service Industries Journal 1-22.

[21].Joulin A, Grave E, Bojanowski P, et al (2017) Bag of Tricks for Efficient Text Classification. Conference of the European Chapter of the Association for Computational Linguistic 427-431.

[22].Bengio Y, Ducharme R, Vincent P, et al. (2003) A Neural Probabilistic Language Model. Journal of Machine Learning Research 3(6): 1137-1155.

[23].Chen D, Manning C D (2014) A Fast and Accurate Dependency Parser using Neural Networks. empirical methods in natural language processing 740-750.

[24].Jiang T J, Wan C X, Liu D X, et al. (2017) Extracting Target-Opinion Pairs Based on Semantic Analysis. Chinese Journal of Computers 40(03):617-633(in Chinese).

[25].Shen J C, Du S X, Luo Y, et al. (2012) Method and Application Research on Fuzzy Comprehensive Evaluation Base on Cloud Model. Fuzzy Systems and Mathematics 26(06):115-123(in Chinese).

[26].Wall E A and Berry L L (2016) The combined effects of the physical environment and employee behavior on customer perception of restaurant service quality. Cornell Hospitality Quarterly 48(48):59-69.

Appendix A Complete Results of Sentiment Distribution by DP-FastText

	R1	R2	R3	R4	R5
C1	[9,11,86]	[29,33,117]	[15,28,62]	[9,14,14]	[65,47,190]
C2	[123,112,473]	[176,190,1654]	[184,144,838]	[21,37,387]	[243,224,3009]
C3	[0,2,11]	[3,0,1]	[0,0,3]	[4,0,0]	[5,2,3]
C4	[163,21,192]	[281,12,356]	[111,13,105]	[20,5,31]	[332,25,305]
C5	[52,241,383]	[175,311,364]	[439,126,57]	[12,48,101]	[59,91,10]
C6	[13,12,381]	[18,10,645]	[43,14,89]	[0,2,167]	[11,7,650]
C7	[10,83,9]	[14,412,63]	[9,88,7]	[2,91,10]	[7,295,35]
C8	[12,63,446]	[41,119,857]	[5,47,396]	[24,59,32]	[17,59,961]
C9	[12,72,409]	[43,132,717]	[18,66,368]	[6,61,30]	[16,69,895]
C10	[18,60,391]	[23,116,679]	[15,52,345]	[38,86,26]	[42,59,834]
C11	[28,80,393]	[63,122,729]	[17,64,373]	[19,56,41]	[35,70,901]
C12	[226,82,230]	[593,161,1195]	[137,35,193]	[124,44,390]	[145,47,679]
C13	[145,563,2187]	[297,947,5584]	[290,757,2323]	[43,188,1912]	[172,614,7443]
C14	[6,7,14]	[6,10,42]	[3,5,14]	[0,1,8]	[7,3,83]
C15	[6,3,149]	[19,1,282]	[18,1,130]	[2,1,94]	[8,2,618]

Appendix B Evaluation Result of 5 restaurant

$$R_1 = \begin{bmatrix} 4.45 & 1.11 & 0.49 \\ 3.99 & 1.69 & 0.69 \\ 4.69 & 0.65 & 0.31 \\ \cdots & \cdots & \cdots \\ 4.41 & 1.12 & 0.11 \\ 3.59 & 1.83 & 0.86 \\ 4.81 & 0.45 & 0.67 \end{bmatrix} R_2 = \begin{bmatrix} 3.98 & 1.67 & 0.69 \\ 4.46 & 1.10 & 0.52 \\ 2.00 & 1.88 & 0.73 \\ \cdots & \cdots & \cdots \\ 4.55 & 0.93 & 0.44 \\ 4.24 & 1.38 & 0.36 \\ 4.74 & 0.61 & 0.77 \end{bmatrix} R_3 = \begin{bmatrix} 3.90 & 1.64 & 0.74 \\ 4.12 & 1.58 & 0.51 \\ 5.00 & 0.00 & 0.00 \\ \cdots & \cdots & \cdots \\ 4.21 & 1.37 & 0.48 \\ 4.00 & 1.60 & 0.67 \\ 4.50 & 1.09 & 0.73 \end{bmatrix}$$

$$R_4 = \begin{bmatrix} 3.27 & 1.64 & 0.53 \\ 4.64 & 0.77 & 0.60 \\ 1.00 & 0.00 & 0.00 \\ \cdots & \cdots & \cdots \\ 4.74 & 0.57 & 0.53 \\ 4.78 & 0.50 & 0.39 \\ 4.90 & 0.25 & 0.55 \end{bmatrix} R_5 = \begin{bmatrix} 3.83 & 1.85 & 0.85 \\ 4.59 & 0.89 & 0.65 \\ 2.60 & 2.01 & 0.99 \\ \cdots & \cdots & \cdots \\ 4.77 & 0.53 & 0.55 \\ 4.63 & 0.82 & 0.73 \\ 4.94 & 0.14 & 0.44 \end{bmatrix}$$