




A sentiment analysis driven method based on public and personal preferences with correlated attributes to select online doctors

Jian Wu^{1,2} · Guangyin Zhang^{1,2} · Yumei Xing^{1,2} · Yujia Liu^{1,2}  · Zhen Zhang³ · Yucheng Dong⁴ · Enrique Herrera-Viedma⁵

Accepted: 23 January 2023 / Published online: 20 February 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

This paper proposes a method to assist patients in finding the most appropriate doctor for online medical consultation. To do that, it constructs an online doctor selection decision-making method that considers the correlation attributes, in which the measure of attribute correlation is derived from the history real decision data. To combine public and personal preference with correlated attributes, it proposes a Choquet integral based comprehensive online doctor ranking method. In detail, a two stage classification model based on BERT (Bidirectional Encoder Representations from Transformers) is used to extract service features from unstructured text reviews. Then, 2-additive fuzzy measure is adopted to represent the patient public group aggregated attribute preference. Next, a novel optimization model is proposed to combine the public preference and personal preference. Finally, a case study of dxy.com is carried out to illustrate the procedure of the method. The comparison result between proposed method and other traditional MADM (multi-attribute decision-making) methods prove its rationality.

Keywords Multi-attribute decision-making · Doctor selection · Online reviews · Sentiment analysis · Choquet integral · 2-additive fuzzy measure

1 Introduction

Recently, the online medical service has showcased the tremendous growth under the continuous impact of COVID-19 epidemic [1]. A growing number of people seek medical services through online health platforms, such as Haodf and Dxy in China; Teladoc, DoctorOnDemand, Amwell in American. The emerging of these platforms improved the patient's access to more medical resources comparing to traditional offline medical visits. Generally, these online platforms provide lots of doctors in most of the specialties and amount of the descriptive information about each doctor, like the certification, professional title, total number of visits and the patients' history reviews of doctors. It allows patients to select and interact with a doctor by using communication technology, such as video, phone or message without any time and place limitation [2].

Plenty of scholars have carried out research of online medicine industry. Some studies investigated the satisfaction influencing factors in the use of telemedicine service from both doctor and patient perspective [3]. Some other studies proposed the concern about the quality of telemedicine [4]. They compared telemedicine with the traditional offline service by the randomized controlled trials or quasi-randomized controlled trials. Meanwhile, there is limited research on patient selection of online doctors. Based on health platform provided massive and heterogeneous information, how to help patients find a high quality and appropriate doctor according to the patient's personal preference is an understudied question. Designing a good method to help patients select appropriate doctors can contribute to increase the overall patient satisfaction level and build harmonious doctor-patient relationships. The application of recommendation technique to online health field will face data sparsity problems since the patients usually have few interactions with doctors in a long time period on the platform [5]. Guo et al. [6] deployed an unsupervised aggregation approach for integrating various ranking features to identify doctors who have the potential to be the key opinion leaders on a range of diseases. The weight of attributes in

✉ Yujia Liu
lyj71@126.com

their ranking method are only derived from the data objectively, i.e., it ignores the patient's personal preference, and they do not consider the interaction relationships among different service features.

People who search for online medical services may have different expectations. Some patients ask for health service through online just for its convenience and not serious conditions. While some other patients may seek for the "second opinion". For example, they want to search an optimal treatment plan through asking several numbers of professionals for its complicated health status. As Lu and Wu [7] indicated, the disease risk can significantly influence patient's choice. Patients who suffer from a high-risk disease will be more worried and hope to be able to find a doctor with higher technical quality, rather than solely a doctor with high caring behaviors. Conversely, patient cares more about functional quality of doctors when affected by relatively low-risk diseases, and caring behaviors make patients more satisfied. In some situations, patients will have additional high demand on response time. Therefore, it is important to consider the patient's personal preference in the selection process. Besides, the information asymmetry between the doctor and the patient is a well known problem in medical industry [8], which also exists in online environment. This may lead some doctors to recommend more/less treatment than the patient's fully informed choice. The patient history reviews displayed in health platforms can alleviate this kind of asymmetry and work as an essential reference source for the quality of the doctor services. Previous studies have also proved the information in online reviews will influence patients' decisions [9].

Based on the above analysis, this study proposes a MADM approach to assist the patient in selecting a high quality and appropriate doctor. The main work of this study are as follows:

- 1) To extract patient opinion from review texts, a two stage classification model based on BERT is devised. The extracted features merge with other attributes which could influence patients decisions to build the complete decision matrix. The Choquet integral with respect to fuzzy measure is suggested to model the attribute interaction relationships and aggregate the attribute information.
- 2) In order to provide ranking for patients with different kinds of requirements, the attribute weight allocation takes both the public and the personalized preferences into account. An optimization model is constructed to identify the public preference based on the doctor consultation volume. Another optimization model is set up to combine the public and the personalized preferences.

The rest of this paper is organized as follows. Related works about health service quality and evaluation, sentiment

analysis in health environment are briefly reviewed in Section 2. Section 3 presents some definitions and theorems of the Choquet integral and fuzzy measure. Section 4 introduces the framework of the proposed method, mainly including four parts, sentiment analysis of patient review, public preference identification, personal preference elicitation and preference integration. Section 5 is a case study using the data from Dxy.com to show the procedure of the proposed method. Section 6 proposed the comparison and discussion of the proposed correlated attributes-based method and some conventional independent attributes-based MADM methods. The conclusion and future work are presented in the last part.

2 Related works

2.1 Healthcare service quality and evaluation

The majority of studies used relatively generic quality models to evaluate health or medical service. One of the most representative is the SERVQUAL model and its variant form, which mainly divides the service quality into tangibility, reliability, responsiveness, assurance and empathy five dimensions, focusing mainly on the perception of the service process [10]. For example, Singh and Prasher [11] combined SERVQUAL and Fuzzy-Analytic Hierarchy Process (AHP) to evaluate the service of the Indian healthcare sector. Grönroos [12] divided service quality perception into technical quality (outcome quality) and functional quality (process quality), in which, technical quality focuses on skills, the precision of practices and procedures, diagnosis accuracy, and treatment outcome, while functional quality refers to the methods through which the services are provided to patients. Lu and Wu [7] used the empirical data to examine the technical quality and functional quality of patient online booking choices. The two factors are directly given by patients who have made a review about a doctor. They use the ordinary least squares to check the assumption and also found that these two factors have an interaction effect. Lu et al. [13] used the SERVQUAL theory to measure customer perception of functional quality and combine it with technical quality to build a relatively comprehensive evaluation index of telemedicine. Sheng and Teo [14] proposed that every experience perceived by customers is a complex feeling, and it is difficult to distinguish among such feelings because they may have relevant overlapping areas and clear interrelations. The concept of Shapley index is an effective method for determining the weight of experts with interactive relationships, which can be represented by the fuzzy measures [15]. However, the fuzzy measures in the existing studies are given by

experts subjectively or experiment designed by researchers. For non-expert decision-making problem based on online reviews, obtaining the fuzzy measures through not just subjective opinion but online history real decision data makes the decision-making process more objective. To aggregate the information of interactive attributes, the Choquet integral is utilized as an effective tool by researchers [16].

Meanwhile, some research constructed more specific evaluation dimensions from the patient's perspective. Reader et al. [17] proposed a patient complaint taxonomy. And late research [18] proved its rationality. They use three concepts to summarize the complaint: clinical, management, relationships. López et al. [19] categorized the patient reviews into two evaluation types, specific description and global remark and identified three topics in specific description; interpersonal manner; technical competence; and system issues. Xu et al. [20] used the text mining technique to identify seven aspects, namely, bedside manner, diagnosis accuracy, waiting time, service time, insurance process, doctor knowledge, and office environment from the online patient review. They also explored the impact of different dimensions in doctor review on the choice of patient booking behavior. Table 1 lists some topics identified by previous works. In addition to above, some researches indicated that since most patients lack the knowledge and information required to evaluate the technical (clinical) quality of services, their assessment of the service quality is shaped on the basis of the non-clinical or functional aspects [21]. Hong et al. [22] summarized

the researches that compared the patient online review with traditional patient satisfaction surveys. They reported that there is a considerable overlap between the narrative comments of patient reviews and the traditional surveys, but it also exists several identified domains of patient experience that were not covered by the traditional surveys.

Online medical services and traditional medical services differ in some ways. As a result, the evaluation indexes between online and offline medical services will be a little different. Lots of studies evaluated online medicine services from the whole level which focus on a range of specialties or diseases, and test its feasibility or reliability rather than single doctor level. Few studies proposed an evaluation index system specifically for the online medical services. Nouri et al. [23] summarized and categorized the evaluation criteria for the m-health apps. Lu et al. [24] referred the mobile service quality evaluation and traditional medical quality evaluation and proposed a service quality evaluation indexes for the online medical services from the perspective of regional doctors.

2.2 Sentiment analysis with online reviews

Sentiment analysis (SA) is a widely used technique to analyze the subjective text reviews of consumers on products or services, which is utilized to transform the text reviews of shopping experience and satisfactory into quantifiable sentiment orientation [31]. Consumers and business managers can obtain objective opinions on

Table 1 The identified topics in patients' reviews

Research	Topics	DataSource
Hao and Zhang [25]	experience of finding doctors, doctors' technical skills and bedside manner, general appreciation from patients, description of various symptoms	Haodf.com
Lagu et al. [26]	patient care experience, facility or office experience and staff characteristics, clinician technical skills, clinician communication and interpersonal skills, overall negative or positive comments, reason for Seeking Care	randomly sampled from health system and 28 commercial websites
He et al. [18], Reader et al. [17], Harrison et al. [27]	clinical: quality and safety; management: institutional issues, timing and access; relationships: communication, caring	research survey; validated in Vitals.com
Liu et al. [28]	competence, communication, treatment, and convenience	Haodf.com
Wan et al. [29]	professional knowledge, platform, response, personality traits, trust, sickness explanation, diagnosis	Dingxiang Doctor www.dxy.com
Ryskina et al. [30]	perceived attitudes of doctors toward the patients and caregivers, doctor communication with patients or caregivers, perceived clinical expertise of the doctors	Yelp.com, 2009–2018 pertain to nursing home (NH) doctors
Xu et al. [20]	bedside manner, diagnostic accuracy, waiting time, service time, insurance process, doctor knowledge, office environment	one of the leading appointment-booking websites in the United States
López et al. [19]	global Remarks, specific Descriptions: interpersonal manner, technical competence, system-Issues	Ratems.com and Yelp.com.

consumption or decision-making by sentiment analysis [32]. Since the selection of online medical services is a non-expert decision making problem, the objective opinions from the reviews provided decision bases for consumers. For example, measuring health care quality [33], monitoring the outcomes and effectiveness of medications or medical treatment [34].

The existing studies of sentiment analysis methods can be categorized into two types: 1) The lexicon based sentiment analysis method, which is suitable for formal and simple pattern sentences; 2) The machine learning based sentiment analysis method, which performs better on relatively flexible and complicated sentences [35]. Some works combined these two methods to get a more complete and accurate result [36]. When referring to the fine-grained sentiment analysis, the additional task of aspect identification is needed. Zunic et al. [37] provided a review of the sentiment analysis in the health care context, which summarized the commonly used method in health care context and found that few research used the deep learning technique to analyze the sentiment orientation in health care, although the deep learning is proved to have a better performance in the sentiment analysis of product or film reviews. Denecke and Deng [38] proposed different facets of sentiment in health-related texts, such as the accuracy and certainty of a diagnosis, the effectiveness of a treatment outcome, changes in health status and experiences of a sort of drug and. They argued that sentiment in the context of medicine sometimes is difficult to capture in a few words. Yadav et al. [39] used CNN (Convolutional Neural Network) structure to build two classification algorithms for the sentiment analysis of two major medical aspects, medical condition and treatment. Meanwhile, they tested the SentiWordNet (SWN) lexicon and found it not working well in the classification task.

In this study, the BERT neural network [40] is used to identify the public opinion. BERT is a pretraining language model proposed by Google AI research institute, which refreshed the best results of 11 natural language processing tasks in GLUE benchmark (General Language Understanding Evaluation) at its release time. Its stacked self-attention layers have a good capability to model the complex interactive relationship among the characters in the series input structure. Compared with traditional machine learning methods, such as decision tree, support vector machines, Naive Bayes used in Chinese context. One of its advantages is that the minimum unit of input is characters rather than words, which means that word segmentation process is unnecessary before send the text to the model. This advantage reduce the impact caused by word segmentation errors to some extent, which makes it a very suitable model for non-standard Chinese online comment texts. The other advantage is that the BERT model

does not need the feature engineering. Preprocessing such as specifically speaking, the stop words removing, n-gram feature constructing are not required for pretrained BERT model as well.

3 Choquet Integral and 2-additive fuzzy measure

Choquet integral with respect to fuzzy measure is an effective and widely used method to deal with the interactions between attributes in MADM problems [15, 41]. This section introduces some definitions and notations related to Choquet integral and fuzzy measure, all of which will be used in the subsequent content of this paper.

Definition 1 [42] Let $X = \{x_1, x_2 \cdots x_n\}$ be any finite set. $P(X)$ is the subset of X . A discrete fuzzy measure on X is a mapping $\mu : P(X) \rightarrow [0, 1]$ satisfying:

$$\mu(\emptyset) = 0, \mu(X) = 1 \quad (1)$$

$$A \subset B \subset X, \text{ implies } \mu(A) \leq \mu(B) \quad (2)$$

Definition 2 [43] Let μ be a fuzzy measure on X . The Möbius transform φ of μ is defined by:

$$\varphi(A) = \sum_{K \subset A} (-1)^{|A \setminus K|} \mu(K), \forall A \subset X \quad (3)$$

Under the Möbius transform, the two constraints in Definition 1 is converted to the following form:

$$\varphi(\emptyset) = 0, \sum_{A \subset X} \varphi(A) = 1 \quad (4)$$

$$\sum_{i \in B \subset A} \varphi(B) > 0, \text{ for all } A \subset X, \text{ for all } i \in A \quad (5)$$

The original fuzzy measure is a monotonous set function with 2^n coefficients to be determined. When the value of n becomes large, the computational complexity will increase significantly. To deal with this problem, based on Möbius transform, k-additive fuzzy measure [44] is proposed to make a good trade-off between computational complexity and presentation capability.

Definition 3 [44] A fuzzy measure μ is said to be k-additive if its Möbius transform satisfies $\varphi(A) = 0$ if $A \subset X$ and $|A| > k$ and there exist at least a subset B such that $|B| = k$ and $\varphi(B) \neq 0$.

2-additive fuzzy measure is a widely used form of k-additive fuzzy measure. It satisfies $\varphi(A) = 0$ if $A \subset X$ and $|A| > 2$ and there exist at least a subset B such that

$|B| = 2$ and $\varphi(B) \neq 0$. It only considers the interactions of pairwise elements and overlooks the higher order interactions. Since there are only $[n(n+1)/2]$ coefficients to be determined, it reaches a best compromise between low complexity and rich presentation capability. In the proposed online doctor selection method, 2-additive fuzzy measure is adopted to represent the attribute weight.

Definition 4 [45] Let φ be the Möbius transform of μ , the Shapley importance index which reveals the average contribution of i to all the relevant subsets can be calculated as:

$$I_i = \sum_{A \subset X \setminus x_i} \frac{1}{|A| + 1} \varphi(A \cup \{x_i\}), \quad \sum_{i=1}^n I_i = 1 \quad (6)$$

In order to intuitively approach the concept of interaction, the Shapley interaction index can be given by:

$$I_{ij} = \sum_{A \subset X \setminus \{x_i, x_j\}} \frac{1}{|A| + 1} \varphi(A \cup \{x_i, x_j\}), \quad I_{ij} \in [-1, 1] \quad (7)$$

In the case of 2-additive fuzzy measure, the Shapley importance index and interaction index will be:

$$I_i = \varphi(\{x_i\}) + \frac{1}{2} \sum_{x_j \in X \setminus x_i} \varphi(\{x_i, x_j\}) \quad (8)$$

$$I_{ij} = \varphi(\{x_i, x_j\}) \quad (9)$$

The interaction index I_{ij} measures the average interaction between two elements i and j . It can be expressed ranging from redundancy (negative) to complementary (positive). The value 1 of I_{ij} represents maximum complementary between element i and j (positive interaction), the value -1 represents maximum substitutivity (negative interaction). Therefore, the shapley interaction index can be used to measure the interaction relationship of the correlated attributes in online doctor selection method and the Shapley importance index is used to reflect the average importance of doctor's each service attribute.

Definition 5 [42] The Choquet Integral of a measurable function $f : X \rightarrow R^+$ in regard to μ is defined by:

$$C_\mu(f) = \sum_{i=1}^n (f(x_{(i)}) - f(x_{(i-1)}))\mu(\{x_{(i)} \dots x_{(n)}\}) \quad (10)$$

where (i) stands for a permutation on i such that $f(x_{(1)}) \leq f(x_{(2)}) \leq \dots \leq f(x_{(n)})$ and $f(x_{(0)}) = 0$. Under the Möbius transform φ , the Choquet Integral can be defined as [45]:

$$C_\varphi(f) = \sum_{A \subset X} \varphi(A) \min_{i \in A} \{f(x_i)\} \quad (11)$$

Based on identified fuzzy measure φ , Choquet Integral can be used to aggregate the attribute information to get the overall evaluation score and the ranking of alternative online doctors.

4 Research framework

4.1 Problem description

The online medical platform which provides online medical consultation services usually shows the doctors' various attributes and patients historical reviews on its webpage. Users need to decide whom to be chosen as a provider of medical consulting services. This study aims to assist a patient to select an online doctor among several potential ones based on these provided information. Due to the limited understanding of online doctors, online text reviews play an important role for obtaining decision-making attributes. At the same time, users usually have their own personal preferences, which also need to be taken into account. An intelligent doctor ranking system can adapt with patients which have different kinds of personal requirement and help them select the most desirable, satisfactory one. The proposed selection method mainly composed of four parts: 1) sentiment analysis of the patient reviews; 2) public preference learning with 2-additive fuzzy measure; 3) personalized preference representation using Fuzzy-BWM (fuzzy best worst method); 4) integration of the two preference and rank doctors. Figure 1 shows the procedure of the proposed method.

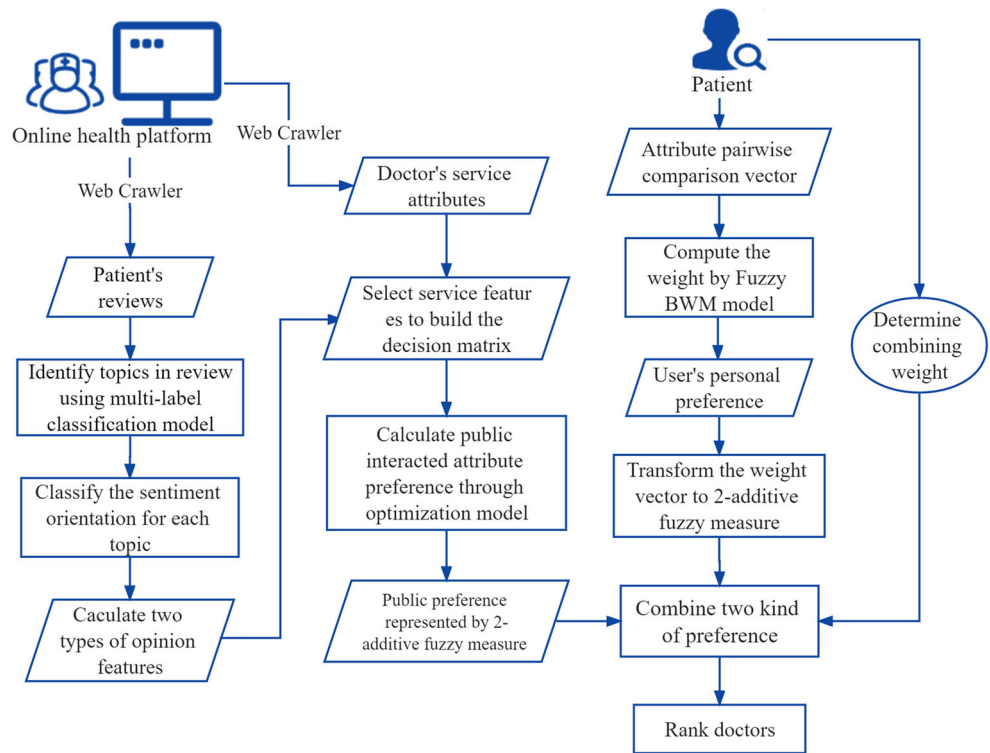
4.2 Data acquisition

A web crawler program written by Python 3.7 is achieved to obtain all the doctors' service attributes (such as consultation price, professional titles, average waiting time etc), and patient historical reviews data from the health platform website. Let $D = \{d_1, d_2, \dots, d_l\}$ be the set of doctors, $C = \{c_1, c_2, \dots, c_m\}$ be the set of the doctor's service attributes, $R = \{R_1, R_2, \dots, R_l\}$ be the set of patients review set, where R_k represents an review set composed of text reviews of doctor d_k .

4.3 Sentiment analysis based on BERT

For the online medical service industry, users' historical reviews are an important resource to reflect the doctor's service quality. To get the sentiment orientation from the reviews, a two-stage sentiment analysis framework is devised. Usually, the user will talk about their consultation experience from several aspects. Thus, the first stage is to predict which aspect each review talks about. The

Fig. 1 The procedure of the proposed method



second stage is to predict the sentiment orientation on each mentioned aspect. Two classification models are constructed for each stage. The deep neural network BERT is adopted to be the encoder function of the two classification model [40]. As a pretrained language model, it is responsible for transforming a string of text into a context-aware high dimensional dense tensors. For a small amount of reviews whose number of characters exceeds the

maximum length 512, its exceeding text part is discarded. This paper initializes the Bert layer of two classification models with the parameters of pretrained “hfl/chinese-bert-wwm-ext” model [46]. And the parameters of other part are randomly initialized. The python and transformers package are used to achieve the sentiment analysis program. Figure 2 shows the framework of the two stage sentiment analysis. The detailed process is described as follows.

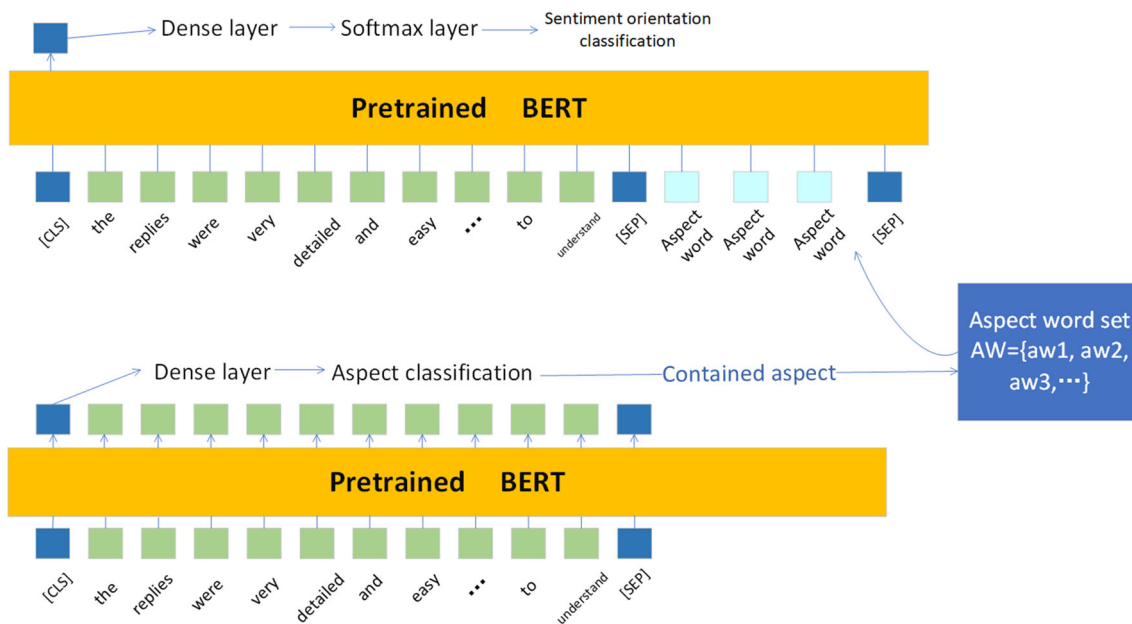


Fig. 2 The framework of the two stage sentiment analysis

4.3.1 Aspect classification

In the first stage, suppose there are n predefined aspects in all the reviews, to identify which aspect that each review in review sets R talked about, a multi-label classification model denoted as $Model_1$ is constructed. Multi-label means that each review may involve with more than one aspect. The model contains a basic Bert layer with default configuration and a dense layer with sigmoid activation function. The input of the basic Bert layer is the raw text of each review. The input of the dense layer is the output of [CLS] token of the basic Bert layer. Each output dimension of the dense layer corresponds to one aspect. The calculation process can be represented in (12). Then, a set of labeled samples need to be prepared to train this model. After training, it can use $Model_1$ to predict the mentioned aspects of all the unlabeled reviews in review set R . The set of aspect vector set $T = \{T_1, T_2, \dots, T_i\}$ can be obtained, where set T_i is composed of the aspect vector $t_j (j = 1, 2, \dots, |R_i|)$ for each review in set R_i . The value of element $t_{jk} (k = 1, 2, \dots, n)$ in aspect vector t_j is assigned as in (13).

$$\begin{aligned} \text{text_vector} &= \text{Character_dictionary}(\text{review text}) \\ \text{aspect_vector} &= \text{Dense_layer}(\text{Bert_layer}(\text{text_vector})) \end{aligned} \quad (12)$$

$$t_{jk} = \begin{cases} 1, & \text{if review } r_j \text{ is predicted to contain the aspect } k. \\ 0, & \text{if review } r_j \text{ is not predicted to contain the aspect } k. \end{cases} \quad (13)$$

$$\begin{aligned} \text{text_vector} &= \text{Character_dictionary}(\text{review text}, \text{aspect word}) \\ \text{sentiment_vector} &= \text{Softmax_layer}(\text{Dense_layer}(\text{Bert_layer}(\text{text_vector}))) \end{aligned} \quad (14)$$

$$s_{jk} = \begin{cases} 1, & \text{if review } r_j \text{ is predicted to be positive toward aspect } k. \\ -1, & \text{if review } r_j \text{ is predicted to be negative toward aspect } k. \\ 0, & \text{if review } r_j \text{ is predicted to be neutral toward aspect } k \text{ or predicted not contain the aspect } k. \end{cases} \quad (15)$$

In the following case study part, when labelling training samples, it turns out that the distribution of categories is extremely unbalanced and most of the reviews in training samples are positive. To enhance the model’s ability to distinguish the negative reviews, the loss of negative class is doubled. It means that the loss of classifying a negative review to a wrong label is twice as to classify a positive review to a wrong label. Some artificial negative samples are also constructed as the supplement by imitating the original negative samples. The pseudo code of the algorithm

4.3.2 Sentiment orientation classification

In the second stage, to predict the sentiment orientation of the identified aspects of each review, the second classification model denoted as $Model_2$ is constructed. It is composed of a basic Bert layer with default configuration, a dense layer and a softmax layer. The input form of Bert layer refers to [47]. For each aspect, they select several words as the aspect word and repeat the aspect word three times to make up the second sentence. The input is goes like “[CLS] raw review text [SEP] aspect word aspect word aspect word”. For example, if $Model_1$ predict that the sentence “The replies were very detailed and easy to understand” talk about the aspect of quality of Question and Answer(Q&A) and the string “Q&A quality” is selected as the aspect word, the input will be “[CLS] The replies were very detailed and easy to understand [SEP] Q&A quality Q&A quality Q&A quality”. The sentiment orientation is set into three types: positive, negative and neutral. The calculation process can be represented in (14). Another set of labeled samples need to be prepared to train $Model_2$. In the predict phrase, $Model_2$ will classify the sentiment orientation of each contained aspect of each unlabeled reviews in review set R . The output is the set of sentiment vector set $S = \{S_1, S_2, \dots, S_i\}$, where set S_i is composed of the sentiment vector $s_j (j = 1, 2, \dots, |R_i|)$ for each review in set R_i . The value of element $s_{jk} (k = 1, 2, \dots, n)$ in sentiment vector s_j is assigned as in (15).

for the above sentiment analysis process is shown in the Algorithm 1.

4.3.3 Defining service features from texts

After getting the aspect vector set T and the sentiment vector set S , two types of features are calculated to reflect the difference between doctors’ review.

For doctor d_i , the first type of features $b_k (k = 1, \dots, n)$ is the proportion of each aspect in doctor d_i ’s reviews. The second type of features $b_{k+n} (k = 1, \dots, n)$ is the proportion of positive reviews in each aspect.

$$b_k = \frac{\sum_{j=1}^{|R_i|} t_{jk}}{|R_i|} \quad (16)$$

$$b_{k+n} = \frac{\sum_{j=1}^{|R_i|} s_{jk} \text{ (if } s_{jk} = 1)}{\sum_{j=1}^{|R_i|} t_{jk}} \quad (17)$$

Input:

Patient's review text set $R = \{R_1, R_2, \dots, R_l\}$

Aspect word set $AW = \{aw_1, aw_2, \dots, aw_n\}$

Output:

Aspect vector set $T = \{T_1, T_2, \dots, T_l\}$, T_i is composed of aspect vector of each review in R_i ; Sentiment vector set $S = \{S_1, S_2, \dots, S_l\}$, S_i is composed of sentiment vector of each review in R_i .

```

1: for each doctor  $d_i$  in D do
2:   for each review  $r_j$  in  $R_i$  do
3:      $t_j = Model\_1(r_j)$ , calculated using (12) and
       (13)
4:     for each  $t_{jk}$  in  $t_j$  do
5:       if  $t_{jk} = 1$  then
6:          $s_{jk} = Model\_2(r_j, \text{aspect word } aw_k)$ ,
           calculated using (14) and (15)
7:       else
8:          $s_{jk} = 0$ 
9:       end if
10:    end for
11:     $s_j = [s_{j1}, s_{j2}, \dots, s_{jn}]$ 
12:  end for
13:   $T_i = [t_1, t_2, \dots]$ 
14:   $S_i = [s_1, s_2, \dots]$ 
15: end for
16: return aspect vector set  $T$  and sentiment vector set  $S$ 

```

Algorithm 1 Sentiment analysis algorithm.

where $|R_i|$ represents the total number of reviews about doctor d_i . Let set $B = \{b_1, b_2, \dots, b_{2n}\}$ represents the features calculated from review texts. Combining set B with the attribute set C , it can get a complete attribute set $X = \{B, C\}$. Let matrix $E_{(m+2n) \times l}$ represents the attributes values of all the doctors.

4.4 Public preference learning based on 2-additive fuzzy measure

To help a user select an appropriate doctor from the online health platform, especially new users which have little knowledge about what kind of doctor is appropriate for herself/himself, it would be helpful to refer to others selection experience. Generally, the preferences of individuals in a group are slightly different from each other. In group decision making field, to get the preference which can represent the whole public, they use different consensus reaching process (CRP) to aggregate the individual's preference [48–50]. In online health platform, the aggregated public's preference reflect their values toward the importance of different service attributes. This can be inferred through each doctor's consultation

volume in a fixed time period. A doctor with high value of consultation volume means that the doctor's service attributes satisfy more users' need. When a user chooses a specific doctor, it represents that this doctor is his best alternative after limited comparison. To deal with the potential complex service attribute relationships, 2-additive fuzzy measure is adopted to represent the public group preference. It has the ability to model attribute interaction relationship among pairs including redundancy and complementary. If no interaction exists among attributes, it will reduce to weighted arithmetic mean. Considering the attribute interaction can increase the accuracy of the learned public's preference. The attributes in set X are the decision attributes. The Choquet integral aggregate the interactive attribute information. An optimization model is constructed here [43] to identify the parameters in 2-additive fuzzy measure. Each doctor's consultation volume in a fixed time period is taken as the doctors' overall evaluation score. The idea is to use the Choquet integral to fit each doctor's consultation volume. Therefore, the objective function of the optimization model is minimizing the sum of the squared difference between all the doctors' Choquet Integral calculated using the attribute values in matrix E and his/her overall evaluation score.

$$Z = \sum_{k=1}^l (C_a(f_1^k, \dots, f_{m+2n}^k) - s_k)^2 \quad (18)$$

where s_k are the overall evaluation score of doctor k . f_1^k, \dots, f_{m+2n}^k are the normalized value of each attribute of doctor k . The constraints are the monotonicity and non-negative constraint of the fuzzy measure as in (4) and (5). Before taking the value in matrix $E_{(m+2n) \times l}$ into the optimization model, min-max normalization is calculated along its column. After transformation, this optimization problem can be reduced to a quadratic programming problem with $2^{m+2n} - 1$ variables and $(m + 2n) * 2^{(m+2n)-1} + 1$ constraints, written as follows [43]:

$$\begin{aligned} \text{minimize } Z &= \frac{1}{2} a^T F_a a + \Gamma_a^T a \\ \text{s.t. } &\begin{cases} f^T a = 1 \\ G a + b \geq 0 \end{cases} \end{aligned} \quad (19)$$

where a is the variable vector containing the value of φ_b for $\forall A \subset X$. F and G are parameter matrix. Γ and f are parameter vectors. Solving the quadratic programming problem, the 2-additive fuzzy measure φ_b representing the aggregated public preference can be obtained.

4.5 Personal preference representation through Fuzzy-BWM

Except for referring to others preference, it is also necessary to consider user's personal preference in the final decision stage. In this step, the Fuzzy-BWM method is adopted to

help users express their own preference. The BWM model was introduced by [51]. Later Guo and Zhao [52] extended it to the fuzzy space and proposed the Fuzzy-BWM method. Compared with previous similar methods, BWM takes both the computational efficiency and the ability to check the consistency into account. The main steps of the Fuzzy-BWM method are as follows:

Step 1 The user determines the most important and least important attribute in the attribute set X .

Step 2 Compare other attributes with the most important and least important attribute, using the linguistic variables listed in the first column in Table 2 to represent the comparison result. The most important to others vector is as $\tilde{A}_B = (\tilde{a}_{B1}, \tilde{a}_{B2}, \dots, \tilde{a}_{Bn})$. The others to the least important vector is as $\tilde{A}_w = (\tilde{a}_{1w}, \tilde{a}_{2w}, \dots, \tilde{a}_{nw})$. \tilde{a}_{Bj} represents the fuzzy preference of the best attribute X_B over the attribute X_j . \tilde{a}_{jw} represents the fuzzy preference of attribute X_j over the worst attribute X_w . $\tilde{a}_{BB} = (1, 1, 1)$ and $\tilde{a}_{ww} = (1, 1, 1)$.

Step 3 Transform the linguistic variables to fuzzy ratings represented by triangular fuzzy number (TFN), and the rules of transformation are listed in Table 2.

Step 4 Use the optimization model to get the best weights. Considering \tilde{w}_j , \tilde{w}_w and \tilde{w}_B as TFNs. Use $\tilde{w}_j = (l_j^w, m_j^w, u_j^w)$ to represent the fuzzy weight of attribute j . The optimization problem (20) can be constructed.

$$\min \max_j \{ |\frac{\tilde{w}_B}{\tilde{w}_j} - \tilde{a}_{Bj}|, |\frac{\tilde{w}_j}{\tilde{w}_w} - \tilde{a}_{jw}| \}$$

$$s.t. \begin{cases} \sum_{j=1}^n R(\tilde{w}_j) = 1 \\ l_j^w \leq m_j^w \leq u_j^w \\ l_j^w \geq 0 \\ j = 1, 2, \dots, n \end{cases} \quad (20)$$

It can be transferred to the optimization form as in (21).

$$s.t. \begin{cases} \min \tilde{\xi} \\ |\frac{\tilde{w}_B}{\tilde{w}_j} - \tilde{a}_{Bj}| \leq \tilde{\xi} \\ |\frac{\tilde{w}_j}{\tilde{w}_w} - \tilde{a}_{jw}| \leq \tilde{\xi} \\ \sum_{j=1}^n R(\tilde{w}_j) = 1 \\ l_j^w \leq m_j^w \leq u_j^w \\ l_j^w \geq 0 \\ j = 1, 2, \dots, n \end{cases} \quad (21)$$

Solving the above optimization problem, the optimal fuzzy weights $(\tilde{w}_1^*, \tilde{w}_2^*, \dots, \tilde{w}_n^*)$ can be obtained. Next the graded mean integration representation (GMIR) (22) is used

Table 2 The linguistic term set

Linguistic Terms	Membership function
Equally importance(EI)	(1,1,1)
Weakly important(WI)	(2/3,1,3/2)
Fairly Important(FI)	(3/2,2,5/2)
Very important(VI)	(5/2,3,7/2)
Absolutely important(AI)	(7/2,4,9/2)

to transform fuzzy weight of attribute to crisp weights. Let $W_p = \{w_1, w_2, \dots, w_{m+2n}\}$ represents the calculated user's personal attribute weight.

$$R(\tilde{a}_i) = \frac{l_i + 4m_i + u_i}{6} \quad (22)$$

Step 5 Calculate the consistency ratio (CR) to check the consistency degree of pairwise comparison. According to [52], following equation is used to calculate the consistency index.

$$\xi^2 - (1 + 2u_{BW})\xi + (u^2 - u_{BW}) = 0 \quad (23)$$

where u_{BW} is the upper boundary of \tilde{a}_{BW} . By solving the (23), it can get the consistency index ξ . The CR can be calculated as:

$$consistency\ ratio = \frac{\xi^*}{consistency\ index} \quad (24)$$

where $CR \in [0, 1]$. The lower the CR the more consistent the user's given comparisons.

4.6 Integrating public preference and personal preference

In this part, the public's preference φ_b acquired in Section 4.4 needs to integrate with the personal preference W_p acquired in Section 4.5. Since the public's preference is represented by the 2-additive fuzzy measure and the personal preference is represented by the common weight vector. To fuse these two kind of preference structure, an optimization model is build to transform the common weight vector W_p to the 2-additive fuzzy measure form φ_p . The Shapley importance index reflects the overall importance of an attribute and has the same form with the common weight vector. Thus, the first constraint is that the Shapley importance index vector I_b of φ_b should equal to the weight vector W_p at each attribute. The second constraint is the monotonicity constraint of the 2-additive fuzzy measure. Because the Shapley importance index and the 2-additive fuzzy measure are not one to one correspondence, to get a single solution, the new fuzzy measure φ_p is expected to be as close as possible to φ_b . For the attribute set with two elements, the new fuzzy measure

φ_p is expected to have the difference with φ_b only in the set which weight is not zero in φ_b . Hence, the objective function can be divided into two terms. The first term is defined as minimize the squared distance between φ_p and φ_b in attribute set which has one element or whose value is not zero in φ_b . The second loss term is defined as minimize the weight of attribute set whose value is zero in φ_b . The optimization problem can be represented as follows:

$$\min \sum_{A \subset Set1} (\varphi_b(A) - \varphi_p(A))^2 + 2 * \sum_{A \subset Set2} \varphi_p(A)^2$$

$$s.t. \begin{cases} \sum_{A \subset X \setminus x_i} \frac{1}{|A|+1} \varphi_p(A \cup \{x_i\}) = w_i \text{ for all } i \in X \\ \sum_{A \subset X} \varphi_p(A) = 1 \\ \sum_{i \in B \subset A} \varphi_p(B) \geq 0 \text{ for all } A \subset X, \text{ for all } i \in A \end{cases}$$

(25)

where $Set1 = \{A \mid A \subset X \text{ and } |A| = 1 \text{ or } |\varphi_b(A)| > 0\}$, $Set2 = \{A \mid A \subset X \text{ and } |A| = 2 \text{ and } \varphi_b(A) = 0\}$. Fuzzy measure φ_p contains the set weight that need to be determined. Solving the problem (25), the user's personal preference with 2-additive fuzzy measure form can be obtained. Then (26) is used to combine it with the public preference φ_b .

$$\varphi(A) = (1 - \gamma) * \varphi_b(A) + \gamma * \varphi_p(A) \quad \forall A \subset X \quad (26)$$

$\gamma (0 \leq \gamma \leq 1)$ is a parameter specified by the user to determine the relative importance of these two kind of weights. Patients with chronic diseases or long

term conditions may familiar with which kind of doctors are suitable for them since they have a long interaction experience with doctors. They can set γ to a low value and rank doctors based more on their own preference. People who seldom get sick may not have a clear, firm idea about which kind of doctor are appropriate for them and their conditions, which doctor attribute is more important and which is less important. They can set γ to a high value and rank doctors based more on other's selection experiences. It's easy to prove that φ still satisfies the monotonicity constraint. Based on the final fuzzy measure φ , the Choquet Integral as the overall evaluation score can be calculated using (11). The user can get a rank of his doctor alternatives through sort of the overall evaluation score.

5 Case study: doctor selection in Dxy.com

5.1 Case description

This section presents a numerical case study to illustrate the proposed method. Dxy.com is one of the relatively large platforms in China, providing online medical service to the public. Its service has covered most of the department that the large offline hospital has. Its doctors mainly come from the offline tertiary hospital, a few of them are owned by the platform. Figure 3 shows two representation pages of a doctor in Dxy APP. The left one includes the doctor's name, title, hospital, price, skills etc, while the right one is the patient reviews page.

Fig. 3 Screenshots of doctor representation page on dxy APP



According to [53], dermatology, ophthalmology-otorhinolaryngology and gynecology obstetrics pediatrics are most likely be consulted department across all platforms. Based on the websites provided information, four departments the pediatrics, gynaecology, neurology and gastroenterology are selected as experiment. Using the web crawler, it gets all the doctors information, including their objective attributes and patient reviews. In preliminary screening, it deletes the doctors whose monthly reply is lower than 8. Since this group of doctors are busy with handling the offline work affairs and have little time to respond to the health service request from online. Finally, there are 271, 87, 123, 226 doctors in pediatrics, neurology, gynaecology and gastroenterology respectively.

5.2 Decision process

Step 1 Sentiment analysis According to the method described in Section 4.3, it first needs to determine the topics in the review. Referring to previous relevant literature [29] and validated in training dataset annotation process,

it selects and categorizes the evaluation information into four topics. Each topic will be briefly introduced in the following. Table 3 shows some review examples of each topic.

- (1) The first topic is about the doctor's medical ability. It mainly involves the patients description about diagnostic accuracy, health status after following the doctor's advice or prescription.
- (2) The second topic is about the quality of question and answer (Q&A) between patients and doctors online. Since the main interaction form in this platform are the non-instant message or voice, the quality of the information encoding and decoding are very important. Patients often talk about whether the suggestions given by doctors are detailed, clear and easy to understand.
- (3) The third topic is about the doctor's service attitude. A typical negative example is using templates to answer users' questions. In some cases, users will complain that doctors selectively answer their questions and

Table 3 Expression examples in four topics

Topic	Typical examples
Medical ability	<p>Very professional, Very authoritative.</p> <p>The effect of taking medicine according to the doctors advice is very good.</p> <p>It's really helpful to me.</p> <p>The advice given is very effective.</p> <p>Doctors have rich clinical experience.</p>
Service attitude	<p>Very patient.</p> <p>Gentle tone and good attitude.</p> <p>The doctor always answer carefully</p> <p>A very serious and responsible doctor.</p>
Quality Q&A	<p>The answer is so comprehensive that I'm no longer so anxious because of his cough.</p> <p>The replies are very detailed and easy to understand.</p> <p>Explained in great detail and clearly!</p> <p>Organized answer.</p>
General evaluation	<p>A very good doctor.</p> <p>Very satisfied.</p> <p>The doctor is excellent.</p> <p>Flawless.</p> <p>Very reliable.</p> <p>Highly recommended</p>

Table 4 Evaluation of two stage classification result

	General evaluation	Medical ability	Quality of Q&A	Attitude
Number of categorical samples	178	516	644	760
First stage F1-score	0.71	0.85	0.88	0.91
Second stage accuracy	0.88	0.97	0.99	0.97

are unwilling to answer all the questions. When patients use voice to communicate with doctors, he/she can feel the doctor's patience and kindness. A good doctor should be understanding, know about what the patients are worried about and reduce patient anxiety, according to a reviewer saying in his comment.

- (4) The last topic is the comprehensive evaluation about the service. When manually annotating review samples for the training of the classification model, similar to [19], we also find that some users tend to evaluate the services from the whole, while others tend to assess the doctors from one or two specific aspects. The typical expression is like very satisfied, very nice, very recommended. In the attribute calculation process, when a review has been assigned this label, we take it as mentioned all the above three subtopics.

To train two classification models, 2000 reviews are manually annotated to be the training dataset and 271 reviews to be the validation set. The evaluation metric of two stage classification result is shown in Table 4. We choose F1-score and accuracy to reflect the performance of the two model in first stage and the second stage

respectively. The first column in Table 4 are the number of different topic examples in training dataset. Figure 4 shows the average topic distribution in four departments, calculated from the output of *Model*₁. Across these four departments, it can be seen that the quality of Q&A is the most mentioned topic. This is mostly due to that from the patient's perspective, quality of Q&A is the most easily perceived aspect in the interaction process with doctors. Medical ability mentioned frequency in Neurology is little bit higher than other three sectors. This is consistent with our daily medical experience. Since most diseases in this department are chronic and hard to cure, like insomnia, convulsions, epilepsy etc. When searching for professional help, patients tend to focus more on the doctor's medical ability. It also found that in most cases, a doctor will be evaluated to be good at all the three aspects, few of the doctors are evaluated perform well in only one or two aspects.

Step 2 Identification of public preference with 2-additive fuzzy measure

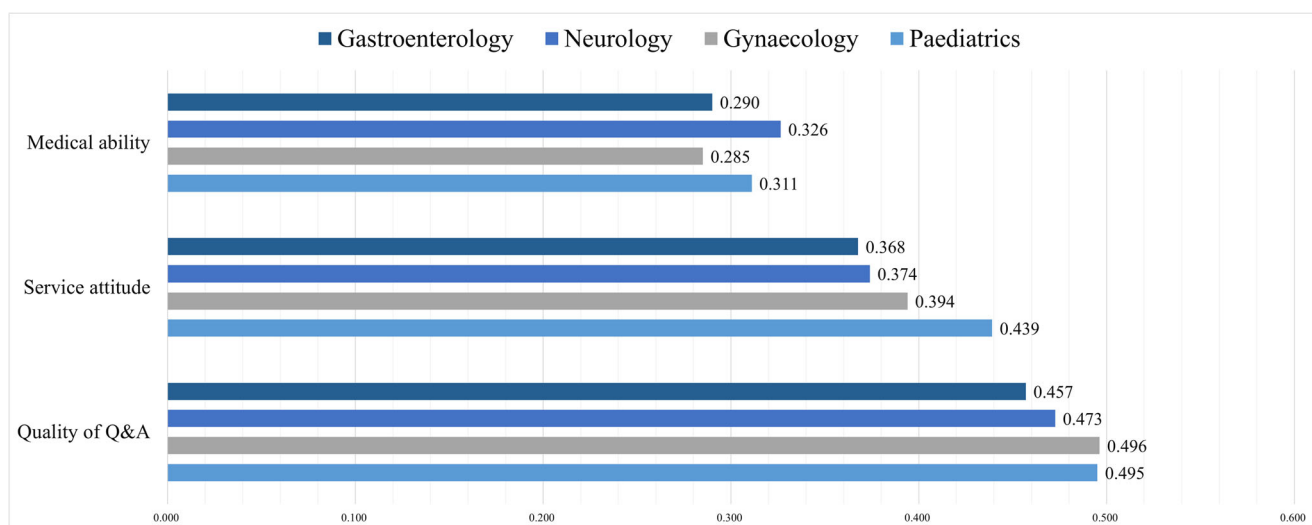
**Fig. 4** Topic proportion in four departments

Table 5 The statistical description of the nine variables in pediatrics before normalization

Attribute	Count	Mean	Std	Min	25%	50%	75%	Max	Coefficient of Variation
Quality of Q&A	271	0.52	0.13	0.00	0.46	0.51	0.57	0.87	0.25
price	271	67.37	49.64	10.00	39.00	60.00	79.00	650.00	0.74
Medical ability	271	0.44	0.18	0.00	0.30	0.40	0.58	1.04	0.41
Review number	271	356.30	597.17	1.00	39.00	167.00	375.00	4531.00	1.68
Job title	271	0.59	0.18	0.30	0.50	0.50	0.75	1.00	0.30
Monthly reply	271	92.02	145.20	10.00	21.00	41.00	94.50	1442.00	1.58
Service attitude	271	0.44	0.12	0.00	0.38	0.45	0.50	1.00	0.27
Weekly praise rate	271	0.04	0.19	0	0	0	0	1	5.12
Avg reply time	271	52.83	65.57	2	13.5	26	58	540	1.24

In this step, it need to build a decision matrix to represent the patients decision situation and identify the 2-additive fuzzy measure based on it. The following eight indexes are selected as the attribute:

- **Review number**, the total number of patient reviews of the doctor.
- **Consultation price**, the price the user has to pay if he/she want to consult the doctor online.
- **Weekly praise rate**, on the doctors’ presentation page, users can see labels the website has assigned to the doctor, each doctor has several labels, like “National top 100 hospitals”, “Professional excellence”, “Doctor Degree of medicine”, “Weekly praise rate top 10”, “charity ambassador” etc. In all these labels, the weekly praise rate(top10, top3, top2, top1) is found to highly correlated to the monthly consultation volume, so the 0-1 variable is build to indicate whether the doctor has this label.
- **Average(Avg) reply time**, average waiting time from the user asking doctor the first question to the doctor’s first response.
- **Job title**, including chief doctors, associate chief doctors, attending doctors and resident doctors. In the

data preprocessing stage, the chief doctors, associate chief doctors, attending doctors and resident doctors are transformed to the numerical scale 1, 0.75, 0.5, 0.25.

- **Medical ability**, this index is obtained by adding two parts together. The first part is given by the label on the doctors home page, if a doctor has been given the label of “Professional excellence”, or “experts in this field”, each label will add 0.25 score to this index. The second part is the topic ratio of medical ability in reviews.
- **Service attitude**, this index corresponds to one of the topics in sentiment analysis, it mainly involves the patients perceived working attitude, attentiveness etc.
- **Quality of Q&A**, this index is also obtained by summing two parts. Like medical ability, the first part is given by the labels on the doctors home page, the second part is topic ratio of quality of Q&A in reviews. If a doctor has been given the label of “patient and careful”, it will add 0.25 score to this index.
- **Monthly consultation volume**, this index is the result variable, the aggregated result of different users decisions and be taken as the evaluation score of each doctor.

The medical ability, service attitude and quality of Q&A need to be calculated by the sentiment analysis.

Table 6 The normalized performance of ten doctors in pediatrics

Doctor index	Medical ability	Avg reply time	Job title	Price	Weekly praise rate	Service attitude	Quality of Q&A	Review number	Monthly reply
d_1	0.871	0.542	0.750	0.323	1	0.524	0.869	0.971	0.765
d_2	0.565	0.640	0.500	0.537	1	0.458	0.497	0.805	0.813
d_3	0.556	0.752	0.750	0.387	0	0.466	0.740	0.740	0.736
d_4	0.600	0.598	0.750	0.619	0	0.549	0.780	0.871	0.606
d_5	0.595	0.804	0.500	0.537	1	0.530	0.720	0.770	0.840
d_6	0.545	0.752	0.750	0.575	0	0.460	0.510	0.791	0.772
d_7	0.591	0.598	1.000	0.308	0	0.505	0.713	0.714	0.473
d_8	0.894	0.928	0.500	0.409	0	0.528	0.799	0.719	0.542
d_9	0.633	0.776	0.500	0.505	0	0.612	0.765	0.644	0.542
d_{10}	0.538	0.776	0.750	0.537	0	0.503	0.749	0.722	0.747

Table 7 The identified public preference in four departments

Attribute set	Pediatrics	Neurology	Gynaecology	Gastroenterology
(medical ability)	0.000	0.000	0.000	0.079
(quality Q&A)	0.000	0.000	0.000	0.000
(service attitude)	0.000	0.000	0.000	0.000
(job title)	0.000	0.000	0.000	0.000
(avg waiting time)	0.135	0.214	0.672	0.395
(price)	0.090	0.062	0.037	0.026
(review num)	0.269	0.074	0.472	0.061
(weekly praise rate)	0.289	0.423	0.411	0.706
(medical ability, quality Q&A)	0.000	0.000	0.000	0.000
(medical ability, service attitude)	0.000	0.000	0.000	0.000
(medical ability, job title)	0.000	0.000	0.000	0.000
(medical ability, avg waiting time)	0.000	0.000	0.000	-0.079
(medical ability, price)	0.000	0.000	0.000	0.000
(medical ability, review num)	0.000	0.000	0.000	0.000
(medical ability, weekly praise rate)	0.000	0.000	0.000	0.000
(quality Q&A, service attitude)	0.000	0.000	0.000	0.000
(quality Q&A, job title)	0.000	0.000	0.000	0.000
(quality Q&A, avg waiting time)	0.117	0.109	0.000	0.000
(quality Q&A, price)	0.000	0.000	-0.000	0.000
(quality Q&A, review num)	0.000	0.211	0.000	0.000
(quality Q&A, weekly praise rate)	0.078	0.000	0.000	0.000
(service attitude, job title)	0.000	0.000	0.000	0.000
(service attitude, avg waiting time)	0.000	0.000	0.000	0.000
(service attitude, price)	0.000	0.000	0.000	0.000
(service attitude, review num)	0.000	0.000	0.000	0.000
(service attitude, weekly praise rate)	0.000	0.000	0.000	0.000
(job title, avg waiting time)	0.000	0.000	0.000	0.000
(job title, price)	0.000	0.000	0.000	0.000
(job title, review num)	0.000	0.000	0.000	0.000
(job title, weekly praise rate)	0.056	0.000	0.000	0.000
(avg waiting time, price)	0.000	0.000	0.000	0.000
(avg waiting time, review num)	0.047	0.026	-0.413	0.190
(avg waiting time, weekly praise rate)	-0.081	-0.214	-0.259	-0.317
(price, review num)	-0.090	-0.062	-0.036	-0.026
(price, weekly praise rate)	0.144	0.169	0.139	0.000
(review num, weekly praise rate)	-0.053	-0.012	-0.023	-0.035

Table 8 The Shapley index for eight attributes in four departments

Attribute	Pediatrics	Neurology	Gynaecology	Gastroenterology
Medical ability	0.0000	0.000	0.000	0.039
Quality Q&A	0.098	0.160	0.000	0.000
Service attitude	0.000	0.000	0.000	0.000
Job title	0.028	0.000	0.000	0.000
Avg waiting time	0.177	0.175	0.336	0.292
Price	0.117	0.115	0.088	0.013
Review num	0.221	0.156	0.236	0.125
Weekly praise rate	0.360	0.394	0.340	0.530
Sum of the Choquet Integral	39.3704	13.7483	52.4191	24.6840
Choquet Integral fitting error	8.2992	2.9878	10.8426	6.2420

The other six indicators can directly be obtained from the representation page. Previously mentioned proportion of positive reviews in each topic got from sentiment result were excluded. Since it found that in each department for more than 90% of the doctors, the negative percent is lower than 1%. The influence of this slight proportion of negative reviews to the user’s decisions can be totally ignored. The situation in reality is that when users browsing the doctor’s review page, it’s hard to find a negative review.

Table 5 shows a statistic description of the nine variables in pediatrics. We found that in all these four departments, the average reply time, monthly reply and review number dimension has the form of long tailed, J-shaped distribution. The consultation price, medical ability, service attitude and quality of Q&A has the form of normal distribution. To enhance the models fitting ability, the log-transformation is applied to the monthly reply dimension, average reply time. Then, it makes a min-max normalization to each variable. The price and average reply time are a cost variable which means the smaller value is better than the bigger value. So these two variables are normalized in the opposite direction. After the above preprocess, a complete decision matrix is constructed. Table 6 shows part of the normalized data in pediatrics.

Next according to Section 4.4, it needs to construct an optimization model to determine the public preference. One optimization model is set up for each department and the Gurobi software is used to solve it. Table 7 shows the optimization result of the Möbius transform fuzzy measure. For the 2-additive fuzzy measure with 8 attributes, it has 36 variables in total. To check the result more intuitively, the Möbius form is transformed into the Shapley index. The result is shown in Table 8. The Shapley index value gives an overall importance of each attribute after considering interaction behaviors between attributes. The last two lines in Table 8 show the Choquet Integral’s fitting ability of the doctor’s assigned overall evaluation score. The second last row is the sum of the evaluation score of all the doctors in each department and the last row is the sum of square prediction error in each department.

Step 3 Obtaining the individual preference

In this step, the patient needs to use the linguistic terms listed in Table 2 to express his/her personal preference toward the attributes. Suppose a patient wants to select a doctor from the Table 6, he/she gives his/her two pairwise weight comparison vectors as in Table 9.

According to (21), it can construct the optimization problem (27). Using the commercial programming solver Gurobi to solve (27), the attribute weight represented by triangular fuzzy number can be obtained. The optimization result of minimum k is 0.814. Next, using (22) to transform the triangular fuzzy number into crisp number,

the result is shown in Table 10. Finally, it can use (23) and (24) to calculate the consistency ration. The result is $0.814/8.04=0.101$, which indicates a high consistency.

$$\begin{aligned}
 & \min k \\
 & l_5 - 3.5 * u_1 \leq k * u_1, l_5 - 3.5 * u_1 \geq -k * u_1 \\
 & m_5 - 4 * m_1 \leq k * m_1, m_5 - 4 * m_1 \geq -k * m_1 \\
 & u_5 - 4.5 * l_1 \leq k * l_1, u_5 - 4.5 * l_1 \geq -k * l_1 \\
 & l_5 - 1 * u_2 \leq k * u_2, l_5 - 1 * u_2 \geq -k * u_2 \\
 & m_5 - 1 * m_2 \leq k * m_2, m_5 - 1 * m_2 \geq -k * m_2 \\
 & u_5 - 1 * l_2 \leq k * l_2, u_5 - 1 * l_2 \geq -k * l_2 \\
 & l_5 - 2/3 * u_3 \leq k * u_3, l_5 - 2/3 * u_3 \geq -k * u_3 \\
 & m_5 - 1 * m_3 \leq k * m_3, m_5 - 1 * m_3 \geq -k * m_3 \\
 & u_5 - 1.5 * l_3 \leq k * l_3, u_5 - 1.5 * l_3 \geq -k * l_3 \\
 & l_5 - 2/3 * u_4 \leq k * u_4, l_5 - 2/3 * u_4 \geq -k * u_4 \\
 & m_5 - 1 * m_4 \leq k * m_4, m_5 - 1 * m_4 \geq -k * m_4 \\
 & u_5 - 1.5 * l_4 \leq k * l_4, u_5 - 1.5 * l_4 \geq -k * l_4 \\
 & l_5 - 1 * u_6 \leq k * u_6, l_5 - 1 * u_6 \geq -k * u_6 \\
 & m_5 - 1 * m_6 \leq k * m_6, m_5 - 1 * m_6 \geq -k * m_6 \\
 & u_5 - 1 * l_6 \leq k * l_6, u_5 - 1 * l_6 \geq -k * l_6 \\
 & l_5 - 2/3 * u_7 \leq k * u_7, l_5 - 2/3 * u_7 \geq -k * u_7 \\
 & m_5 - 1 * m_7 \leq k * m_7, m_5 - 1 * m_7 \geq -k * m_7 \\
 & u_5 - 1.5 * l_7 \leq k * l_7, u_5 - 1.5 * l_7 \geq -k * l_7 \\
 & l_5 - 2.5 * u_8 \leq k * u_8, l_5 - 2.5 * u_8 \geq -k * u_8 \\
 & m_5 - 3 * m_8 \leq k * m_8, m_5 - 3 * m_8 \geq -k * m_8 \\
 & u_5 - 3.5 * l_8 \leq k * l_8, u_5 - 3.5 * l_8 \geq -k * l_8 \\
 & l_2 - 3.5 * u_1 \leq k * u_1, l_2 - 3.5 * u_1 \geq -k * u_1 \\
 & m_2 - 4 * m_1 \leq k * m_1, m_2 - 4 * m_1 \geq -k * m_1 \\
 & u_2 - 4.5 * l_1 \leq k * l_1, u_2 - 4.5 * l_1 \geq -k * l_1 \\
 & l_3 - 1.5 * u_1 \leq k * u_1, l_3 - 1.5 * u_1 \geq -k * u_1 \\
 & m_3 - 2 * m_1 \leq k * m_1, m_3 - 2 * m_1 \geq -k * m_1 \\
 & u_3 - 2.5 * l_1 \leq k * l_1, u_3 - 2.5 * l_1 \geq -k * l_1 \\
 & l_4 - 1.5 * u_1 \leq k * u_1, l_4 - 1.5 * u_1 \geq -k * u_1 \\
 & m_4 - 2 * m_1 \leq k * m_1, m_4 - 2 * m_1 \geq -k * m_1 \\
 & u_4 - 2.5 * l_1 \leq k * l_1, u_4 - 2.5 * l_1 \geq -k * l_1 \\
 & l_6 - 2/3 * u_1 \leq k * u_1, l_6 - 2/3 * u_1 \geq -k * u_1 \\
 & m_6 - 1 * m_1 \leq k * m_1, m_6 - 1 * m_1 \geq -k * m_1 \\
 & u_6 - 1.5 * l_1 \leq k * l_1, u_6 - 1.5 * l_1 \geq -k * l_1 \\
 & l_7 - 2.5 * u_1 \leq k * u_1, l_7 - 2.5 * u_1 \geq -k * u_1 \\
 & m_7 - 3 * m_1 \leq k * m_1, m_7 - 3 * m_1 \geq -k * m_1 \\
 & u_7 - 3.5 * l_1 \leq k * l_1, u_7 - 3.5 * l_1 \geq -k * l_1 \\
 & l_8 - 2/3 * u_1 \leq k * u_1, l_8 - 2/3 * u_1 \geq -k * u_1 \\
 & m_8 - 1 * m_1 \leq k * m_1, m_8 - 1 * m_1 \geq -k * m_1 \\
 & u_8 - 1.5 * l_1 \leq k * l_1, u_8 - 1.5 * l_1 \geq -k * l_1 \\
 & 1/6 * l_1 + 4/6 * m_1 + 1/6 * u_1 + 1/6 * l_2 + 4/6 * m_2 + \\
 & 1/6 * u_2 + 1/6 * l_3 + 4/6 * m_3 + 1/6 * u_3 + 1/6 * l_4 + \\
 & 4/6 * m_4 + 1/6 * u_4 + 1/6 * l_5 + 4/6 * m_5 + 1/6 * u_5 + \\
 & 1/6 * l_6 + 4/6 * m_6 + 1/6 * u_6 + 1/6 * l_7 + 4/6 * m_7 \\
 & + 1/6 * u_7 + 1/6 * l_8 + 4/6 * m_8 + 1/6 * u_8 = 1 \\
 & l_1 \leq m_1 \leq u_1 \\
 & l_2 \leq m_2 \leq u_2 \\
 & l_3 \leq m_3 \leq u_3 \\
 & l_4 \leq m_4 \leq u_4 \\
 & l_5 \leq m_5 \leq u_5 \\
 & l_6 \leq m_6 \leq u_6 \\
 & l_7 \leq m_7 \leq u_7 \\
 & l_8 \leq m_8 \leq u_8 \\
 & l_1 \geq 0, l_2 \geq 0, l_3 \geq 0, l_4 \geq 0, l_5 \geq 0, l_6 \geq 0, l_7 \geq 0, l_8 \geq 0
 \end{aligned}
 \tag{27}$$

Table 9 User's two comparison vectors over eight attributes

Attribute	Job title	Avg reply time	Price	Weekly praise rate	Medical ability	Service attitude	Quality of Q&A	Review number
Medical ability(best attribute)	AI	EI	WI	FI	EI	EI	WI	VI
Job title(worst attribute)	EI	AI	FI	VI	AI	WI	VI	WI

Step 4 Integrating Preference and Ranking In this step, attribute weights obtained in step 2 and step 3 need to be integrated. The user has to specify the parameter representing the relative importance of these two parts. For comparison, γ is set to four values 0.25, 0.5, 0.75 and 1. Using Gurobi to solve the optimization model (25), the transformed fuzzy measure can be obtained. Then the combined fuzzy measure can be calculated using (26). The result is shown in Table 11.

The comprehensive score of each doctor in Table 6 can be calculated through the Choquet integral (11) with the 2-additive fuzzy measure in Table 11. The result is shown in Table 12. Figure 5 shows these doctors' ranking position variation diagram. The attribute weights gradually change from depend only on public preference to depend only on personal preference.

5.3 Result analysis

From Table 7, it can be seen that based on the decision matrix, the optimization model do identified some non-zero attribute interaction index values. A typical positive pair is price and weekly praise rate, which means only when these two attributes get high values simultaneously, they can contribute to a high overall evaluation score. A typical negative pair is the review number and weekly praise rate. Negative interaction means that the high performance of both attributes will not increase the overall evaluation score too much than only one of them get high performance. These can demonstrate that in the practical selection process, the patient is likely to consider the redundancy and complementary attribute relationships. In Table 8, the Shapely value of these attributes reflect their importance in

a more intuitive way. Across these four departments, weekly praise rate, review number and avg waiting time are always the three most important attributes. The weights of attributes calculated based on the reviews are relatively small. This is possibly due to that from the user's perspective, it's hard to tell a remarkable difference between doctors' online review texts. Benefiting from the careful management and refined operation of the platform, doctors with obvious negative or abnormal reviews will be ordered to make adjustments. After a long time, the distribution of the sentiment of patient reviews among doctors become similar.

In Fig. 5, the ranking order of doctors slightly changed with different value of parameter γ . But in general, the ranking are stable without drastic changes. This can prove that the parameter γ can appropriately adjust the rankings of doctors according to the patient's confidence of his own preference.

All the information used in decision process has public access without privacy concern, which is very practical in reality.

5.4 Computational complexity

- (1) The proposed two stage BERT based opinion extraction method is relatively robust and with high accuracy. The BERT encoder works as a strong base-line method by many tasks in natural language processing field. After training, it can be efficiently calculated without take too much time and space complexity.
- (2) What's more, for n attributes, BWM only need $2n$ pairwise comparisons, which is less than the AHP's n^2 pairwise comparisons. In other words, Comparing to AHP method, the fuzzy-BWM used in proposed

Table 10 Personal preference represented in triangular fuzzy number and crisp number form

Attribute	Triangular fuzzy weight	Crisp weight
job title	(0.046, 0.053, 0.064)	0.054
Avg reply time	(0.172, 0.172, 0.172)	0.172
Price	(0.146, 0.146, 0.146)	0.146
Weekly praise rate	(0.114, 0.114, 0.146)	0.120
Medical ability	(0.172, 0.172, 0.172)	0.172
Service attitude	(0.095, 0.095, 0.102)	0.096
Quality of Q&A	(0.114, 0.172, 0.172)	0.162
Review number	(0.064, 0.079, 0.102)	0.080

Table 11 Integrated attribute preference with different γ

Attribute set	$\gamma = 1$	$\gamma = 0.75$	$\gamma = 0.5$	$\gamma = 0.25$
(medical ability,)	0.095	0.071	0.048	0.024
(quality Q&A,)	0.045	0.034	0.023	0.011
(service attitude,)	0.045	0.034	0.023	0.011
(job title,)	0.017	0.013	0.009	0.004
(avg waiting time,)	0.131	0.132	0.133	0.134
(price,)	0.116	0.110	0.103	0.097
(review num,)	0.160	0.187	0.215	0.242
(weekly praise rate,)	0.175	0.204	0.232	0.260
(medical ability, quality Q&A)	0.035	0.026	0.018	0.009
(medical ability, service attitude)	0.035	0.026	0.018	0.009
(medical ability, job title)	0.028	0.021	0.014	0.007
(medical ability, avg waiting time)	0.023	0.017	0.011	0.006
(medical ability, price)	0.030	0.023	0.015	0.008
(medical ability, review num)	0.000	0.000	0.000	0.000
(medical ability, weekly praise rate)	0.000	0.000	0.000	0.000
(quality Q&A, service attitude)	0.023	0.017	0.011	0.006
(quality Q&A, job title)	0.016	0.012	0.008	0.004
(quality Q&A, avg waiting time)	0.138	0.133	0.128	0.123
(quality Q&A, price)	0.018	0.013	0.009	0.004
(quality Q&A, review num)	0.000	0.000	0.000	0.000
(quality Q&A, weekly praise rate)	0.004	0.022	0.041	0.059
(service attitude, job title)	0.016	0.012	0.008	0.004
(service attitude, avg waiting time)	0.010	0.008	0.005	0.003
(service attitude, price)	0.018	0.013	0.009	0.004
(service attitude, review num)	0.000	0.000	0.000	0.000
(service attitude, weekly praise rate)	0.000	0.000	0.000	0.000
(job title, avg waiting time)	0.003	0.003	0.002	0.001
(job title, price)	0.011	0.008	0.005	0.003
(job title, review num)	0.000	0.000	0.000	0.000
(job title, weekly praise rate)	0.000	0.014	0.028	0.042
(avg waiting time, price)	0.006	0.004	0.003	0.001
(avg waiting time, review num)	0.000	0.012	0.023	0.035
(avg waiting time, weekly praise rate)	-0.099	-0.095	-0.090	-0.085
(price, review num)	-0.084	-0.085	-0.087	-0.088
(price, weekly praise rate)	0.061	0.081	0.102	0.123
(review num, weekly praise rate)	-0.076	-0.070	-0.065	-0.059

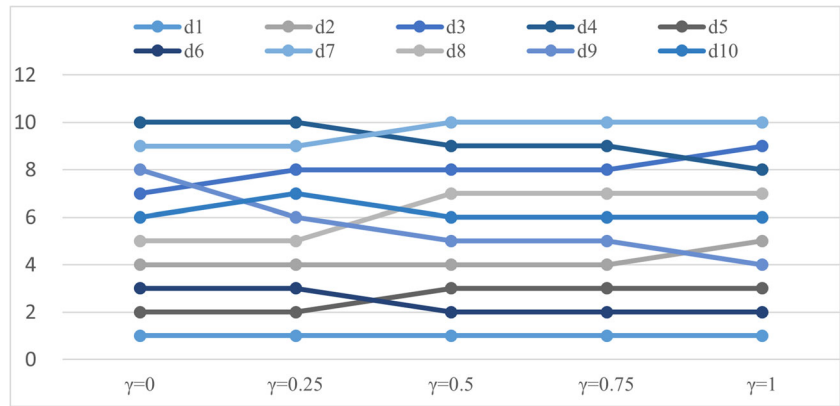
method have less comparison times. Hence, it's more convenient for patients to express his/her values of the importance of attributes.

- (3) The three involved optimization problems in decision process are linear programming or quadratic programming problem. Using existing commercial solver, these problems can converge fast and get their optimal result.

6 Comparasion analysis

To check the influence of considering the attribute interactions, comparison analysis is conducted. The first comparison method uses linear weighted aggregation to replace the Choquet integral, in which, the attribute weight representing the public preference is replaced with the result of linear least squares under the constraint that the sum of

Fig. 5 Doctor ranking with different γ



weights equals to one. The personal weight is the same as case study part, and parameter γ is set to 0.5. In previous studies, considering the privacy protection problem, Hu et al. [54] combined the intuitionistic fuzzy sets (IFSs) and Visekriterijumsko KOMpromisno Rangiranje (VIKOR) to build a doctor ranking method. IFS is a kind of set to represent the fuzzy and uncertain evaluation values, and VIKOR is a MADM method which features in providing a set of compromise solutions under a series of conflicting criteria. Sun et al. [55] proposed two doctor selection method based on single-valued neutrosophic sets (SVNSs) and prospect theory (PT). The extended TODIM (an Acronym in Portuguese of Interactive and MADM) and extended ELECTRE (elimination and choice translating reality) III, another two kinds of MADM method, were used as the ranking function for the two methods respectively. To make these three methods comparable with the method in this paper, their fuzzy evaluation set is replaced with crisp value. And the attribute weight allocation is also replaced with linear-based as in first comparison method. They are denoted as linear+VIKOR, linear+TODIM, linear+ELECTRE III respectively in Table 13. Additional

parameters of these three methods are identical with original paper's settings. Based on the above, using the data from pediatrics department, the proposed method compare with four methods to verify the influence of attributes' correlation, and illustrate the advantage of the proposed method. Table 13 shows these methods' ranking results.

As for the ability to fitting the public preference, the final optimization value of objective Z is 8.299, which can be calculated by (18) and model (19), while the linear least squares based method is 8.449. This can show the 2-additive fuzzy measure's better capacity to capture the public's attribute preference. From Table 13, it could be seen that compared with proposed method, the order of d_1 and d_5 , d_4 and d_{10} exchanged in both linear and linear+TODIM method. This can demonstrate that the attribute interactions can make an impact on final rankings. And it has the necessity to consider the interaction relationships during decision process. The result of linear+ELECTRE III method is quite different from others, this mainly can be ascribed to its sensitivity to its three threshold parameters. The result of linear+VIKOR is most similar to the proposed method, but the parameter setting in this method is focused more

Table 12 The final evaluation score of ten doctors in Table 6

Doctor index	$\gamma = 0$	$\gamma = 0.25$	$\gamma = 0.5$	$\gamma = 0.75$	$\gamma = 1$
d_1	0.737	0.731	0.725	0.719	0.714
d_2	0.458	0.505	0.551	0.598	0.645
d_3	0.302	0.367	0.433	0.498	0.564
d_4	0.291	0.360	0.429	0.497	0.566
d_5	0.734	0.722	0.711	0.699	0.687
d_6	0.720	0.718	0.715	0.713	0.710
d_7	0.299	0.362	0.425	0.488	0.551
d_8	0.338	0.396	0.453	0.511	0.569
d_9	0.301	0.390	0.478	0.567	0.656
d_{10}	0.316	0.387	0.458	0.529	0.601

Table 13 Ranking results of different methods

Methods	Ranking result
Proposed method	$d_5 > d_1 > d_2 > d_8 > d_{10} > d_4 > d_9 > d_3 > d_6 > d_7$
Linear	$d_1 > d_5 > d_2 > d_8 > d_4 > d_{10} > d_9 > d_3 > d_6 > d_7$
Linear+VIKOR [54]	$d_5 > d_1 > d_2 > d_8 > d_4 > d_9 > d_{10} > d_3 > d_6 > d_7$
Linear+TODIM [55]	$d_1 > d_5 > d_2 > d_8 > d_4 > d_9 > d_{10} > d_6 > d_3 > d_7$
Linear+ELECTRE III [55]	$d_{10} > d_5 > d_8 > d_1 = d_4 = d_7 > d_3 > d_6 > d_2 > d_9$

on modeling the behavior preference when facing risks and uncertain conditions instead of complicated attribute preference relationships.

In addition to the above, the proposed method have the following two advantages:

- (1) Firstly, the original weight allocation method in similar studies [54, 55] require decision makers to give the importance of attributes before or weight compare information in pairs, which is unrealistic in some special situations. The proposed method can obtain the weight of attributes by learning the public’s history decision information, and then it can deal with the problem where decision makers have no information of attributes importance.
- (2) Secondly, our method can integrate the public weight and personal weight. To do that, an optimizaiton model (25) is build to transform the the personal weight to the same form with public weight. A parameter γ is used to balance these two kind of weights, making the ranking method has the ability to provide different ranking result for different kind of users. Higher γ means that the user has a high demand and confident of his/her own preference. Lower γ means that the user has little selection experience, and the ranking need to based more on public weight. Figure 5 shows our ranking order is more flexible than the results of previous methods in Table 13.

7 Conclusion

With the continuous development and improvement of the online medical platform and relevant standards, growing number of people will turn to online requesting medical and health information services. This study proposes a doctor ranking method based on the information in online health platform. The proposed method can help patients select appropriate doctors and increase the overall patient satisfaction level. For healthcare providers and platforms, it can also help them better understand the patient’s need and preference. This method can also extend to some other related field [56, 57] with proper adjustment.

The main contribution of this study is presented as follows:

- (1) This study proposes an online text review-based online doctor selection method which considers the correlated attributes. Existing online reviews-based MADM methods assume that attributes are independent of each other, which is hard to match the reality of online doctor selection. On the other hand, the traditional MADM method that considers the correlation of attributes measures the correlation subjectively, rather than from actual data. Moreover, this paper addresses a MADM method that considers correlation attributes based on realdata, and therefore it can regarded a complete new typed decision making problem.
- (2) It proposes a MADM method based on both public preference and personal preference. Existing studies ignore the personal preferences that have same importance with the public preference derived from the online text reviews. To resolve this problem, the public preference is identified through an optimization model with the provided doctor consultation volume information. Then, a novel optimization model is proposed to combine the public preference and personal preference. Therefore, two kind of preference the patient’s personal preference and the public’s preference jointly determine the weight of the attribute to help patients with different degree of experience.

There are several potential directions that can be improved in this paper. For example, in above, the utility function on each single attribute is assumed to be monotonically increasing or decreasing. As Kwok and Lau [58] indicated, the price can be a typical non-monotonic attribute. Some users just want to find an intermediate level price. They don’t want to spend too much or too low. Future work can also try to make a smaller granularity level decomposition of the patient’s final choice, separate the decision makers automatically and provide more fine-grained ranking recommendations.

Acknowledgements The authors are very grateful to the anonymous referees for their valuable comments and suggestions.

Funding This work was supported by National Natural Science Foundation of China(NSFC) (Nos.72001134, 71971135, 71971039, 71910107002)

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

References


- Omboni S, Padwal RS, Alessa T et al (2022) The worldwide impact of telemedicine during COVID-19: current evidence and recommendations for the future. *Connect Health* 1:7–35. <https://doi.org/10.20517/ch.2021.03>
- Uscher-Pines L, Mulcahy A, Cowling D et al (2016) Access and quality of care in direct-to-consumer telemedicine. *Telemed J e-health* 22(4):282–287. <https://doi.org/10.1089/tmj.2015.0079>
- Nguyen M, Waller M, Pandya A et al (2020) A review of patient and provider satisfaction with telemedicine. *Curr Allergy and Asthma Rep* 20(11):72–72. <https://doi.org/10.1007/s11882-020-00969-7>
- Hertzog R, Johnson J, Smith J et al (2019) Diagnostic accuracy in primary care e-visits: Evaluation of a large integrated health care delivery system's experience. *Mayo Clin Proc* 94(6):976–984. <https://doi.org/10.1016/j.mayocp.2019.02.011>
- Yuan H, Deng W (2021) Doctor recommendation on health-care consultation platforms: an integrated framework of knowledge graph and deep learning. *Internet Res* 32(2):454–476. <https://doi.org/10.1108/intr-07-2020-0379>
- Guo L, Jin B, Yao C et al (2016) Which doctor to trust: a recommender system for identifying the right doctors. *J Med Internet Res* 18(7):e186. <https://doi.org/10.2196/jmir.6015>
- Lu N, Wu H (2016) Exploring the impact of word-of-mouth about physicians' service quality on patient choice based on online health communities. *BMC Med Inform Decis Making*, 16(1). <https://doi.org/10.1186/s12911-016-0386-0>
- Lupton M (2005) Informed consent : can a patient ever be fully informed? *Current opinion in obstetrics & gynecology* 17(6):601–604. <https://doi.org/10.1097/01.gco.0000191900.61697.74>
- Han X, Qu J, Zhang T (2019) Exploring the impact of review valence, disease risk, and trust on patient choice based on online physician reviews. *Telematics Inform* 45:101,276. <https://doi.org/10.1016/j.tele.2019.101276>
- Matin BK, Rezaei S, Moradinazar M et al (2016) Measurement of quality of primary health services by SERVQUAL model: evidence from urban health centers in west of iran. *Res J Med Sci* 10(5):475–480
- Singh A, Prasher A (2019) Measuring healthcare service quality from patients' perspective: using fuzzy AHP application. *Total Qual Manag Bus Excell* 30(3-4):284–300. <https://doi.org/10.1080/14783363.2017.1302794>
- Grönroos C (1984) A service quality model and its marketing implications. *Eur J Mark* 18(4):36–44. <https://doi.org/10.1108/EUM000000004784>
- Lu W, Hou H, Ma R et al (2021) Influencing factors of patient satisfaction in teleconsultation: a cross-sectional study. *Technol Forecast Soc Chang* 168:120,775. <https://doi.org/10.1016/j.techfore.2021.120775>
- Sheng ML, Teo TSH (2012) Product attributes and brand equity in the mobile domain : the mediating role of customer experience. *Int J Inf Manag* 32(2):139–146. <https://doi.org/10.1016/j.ijinfomgt.2011.11.017>
- Xing Y, Cao M, Liu Y et al (2022) A choquet integral based interval type-2 trapezoidal fuzzy multiple attribute group decision making for sustainable supplier selection. *Comput Ind Eng* 165:107,935. <https://doi.org/10.1016/j.cie.2022.107935>
- Büyükköçkan G, Göçer F, Uztürk D (2021) A novel pythagorean fuzzy set integrated choquet integral approach for vertical farming technology assessment. *Comput Ind Eng* 158:107,384. <https://doi.org/10.1016/j.cie.2021.107384>
- Reader TW, Gillespie A, Roberts J (2014) Patient complaints in healthcare systems : a systematic review and coding taxonomy. *BMJ Qual Saf* 23(8):678–89. <https://doi.org/10.1136/bmjqs-2013-002437>
- He L, He C, Wang Y et al (2020) What do patients care about? mining fine-grained patient concerns from online physician reviews through computer-assisted multi-level qualitative analysis. In: *AMIA annual symposium proceedings, american medical informatics association*, pp 544
- López A, Detz A, Ratanawongsa N et al (2012) What patients say about their doctors online : a qualitative content analysis. *J Gen Intern Med* 27(6):685–692. <https://doi.org/10.1007/s11606-011-1958-4>
- Xu Y, Armony M, Ghose A (2021) The interplay between online reviews and physician demand: an empirical investigation. *Manag Sci*, 67(12). <https://doi.org/10.1287/mnsc.2020.3879>
- James TL, Villacis Calderon ED, Cook DF (2017) Exploring patient perceptions of healthcare service quality through analysis of unstructured feedback. *Expert Syst Appl* 71:479–492. <https://doi.org/10.1016/j.eswa.2016.11.004>
- Hong YA, Liang C, Radcliff TA et al (2019) What do patients say about doctors online? a systematic review of studies on patient online reviews. *J Med Internet Res* 21(4):e12,521. <https://doi.org/10.2196/12521>
- Nouri R, R Niakan Kalhori S, Ghazisaeedi M et al (2018) Criteria for assessing the quality of mhealth apps: a systematic review. *J Am Med Inform Assoc* 25(8):1089–1098. <https://doi.org/10.1093/jamia/ocy050>
- Lu W, Wang XP, Zhao J et al (2020) Research on teleconsultation service quality based on multi-granularity linguistic information : the perspective of regional doctors. *BMC Med Informat Decis Making* 20(1):113–113. <https://doi.org/10.1186/s12911-020-01155-5>
- Hao H, Zhang K (2016) The voice of chinese health consumers : a text mining approach to web-based physician reviews. *J Med Internet Res* 18(5):e108–e108. <https://doi.org/10.2196/jmir.4430>
- Lagu T, Norton CM, Russo LM et al (2019) Reporting of patient experience data on health systems' websites and commercial physician-rating websites : Mixed-methods analysis. *J Med Internet Res* 21(3):e12,007–e12,007. <https://doi.org/10.2196/12007>
- Harrison R, Walton M, Healy J et al (2016) Patient complaints about hospital services : applying a complaint taxonomy to analyse and respond to complaints. *Int. J. Qual. Health Care : J Int Soc Qual Health Care* 28(2):240–5. <https://doi.org/10.1093/intqhc/mzw003>
- Liu J, Zhang W, Jiang X et al (2020) Data mining of the reviews from online private doctors. *Telemed e-Health* 26(9):1157–1166. <https://doi.org/10.1089/tmj.2019.0159>
- Wan Y, Peng Z, Wang Y et al (2021) Influencing factors and mechanism of doctor consultation volume on online medical consultation platforms based on physician review analysis. *Internet Res* 31(6):2055–2075. <https://doi.org/10.1108/INTR-10-2020-0589>
- Ryskina KL, Wang H, Foley KA et al (2020) Patient and caregiver perceptions of nursing home physicians: insight from

- yelp reviews, 2009–2018. *J Am Geriatr Soc* 68(9):2101–2105. <https://doi.org/10.1111/jgs.16634>
31. Ji F, Cao Q, Li H et al (2022) An online reviews-driven large-scale group decision making approach for evaluating user satisfaction of sharing accommodation. *Expert Syst Appl* 213:118,875. <https://doi.org/10.1016/j.eswa.2022.118875>
 32. Liang X, Liu P, Wang Z (2019) Hotel selection utilizing online reviews: a novel decision support model based on sentiment analysis and DL-VIKOR method. *Technol Econ Dev Econ* 25(6):1139–1161. <https://doi.org/10.3846/tede.2019.10766>
 33. Cambria E, Benson T, Eckl C et al (2012) Sentic PROMs: application of sentic computing to the development of a novel unified framework for measuring health-care quality. *Expert Syst Appl* 39(12):10,533–10,543. <https://doi.org/10.1016/j.eswa.2012.02.120>
 34. Adams DZ, Gruss R, Abrahams AS (2017) Automated discovery of safety and efficacy concerns for joint & muscle pain relief treatments from online reviews. *Int J Med Inform (Shannon, Ireland)* 100:108–120. <https://doi.org/10.1016/j.ijmedinf.2017.01.005>
 35. Fan ZP, Li GM, Liu Y (2020) Processes and methods of information fusion for ranking products based on online reviews : an overview. *Inf Fusion* 60:87–97. <https://doi.org/10.1016/j.inffus.2020.02.007>
 36. Khan FH, Qamar U, Bashir S (2016) Swims : semi-supervised subjective feature weighting and intelligent model selection for sentiment analysis. *Knowledge-based systems* 100:97–111. <https://doi.org/10.1016/j.knsys.2016.02.011>
 37. Zunic A, Corcoran P, Spasic I (2020) Sentiment analysis in health and well-being: systematic review. *JMIR Med Inform* 8(1):e16,023. <https://doi.org/10.2196/16023>
 38. Denecke K, Deng Y (2015) Sentiment analysis in medical settings : new opportunities and challenges. *Artif Intell Med* 64(1):17–27. <https://doi.org/10.1016/j.artmed.2015.03.006>
 39. Yadav S, Ekbal A, Saha S et al (2018) Medical sentiment analysis using social media: Towards building a patient assisted system. In: Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018), European Language Resources Association (ELRA), Miyazaki, Japan
 40. Devlin J, Chang MW, Lee K et al (2019) BERT : Pre-training Of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies, vol 1(Long and Short Papers). Association for Computational Linguistics, Minneapolis, Minnesota, pp 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
 41. Lou S, Feng Y, Li Z et al (2021) Two-additive fuzzy measure-based information integration approach to product design alternative evaluation. *J Ind Inf Integr* 25:100,247. <https://doi.org/10.1016/j.jii.2021.100247>
 42. Grabisch M (1995) Fuzzy integral in multicriteria decision making. *Fuzzy Sets Syst* 69(3):279–298. [https://doi.org/10.1016/0165-0114\(94\)00174-6](https://doi.org/10.1016/0165-0114(94)00174-6)
 43. Miranda P, Grabisch M (1999) Optimization issues for fuzzy measures. *International Journal of Uncertainty, Fuzziness Knowl.-Based Syst* 07(06):545–560
 44. Grabisch M (1997) K-order additive discrete fuzzy measures and their representation. *Fuzzy Sets Syst* 92(2):167–189. [https://doi.org/10.1016/S0165-0114\(97\)00168-1](https://doi.org/10.1016/S0165-0114(97)00168-1)
 45. Grabisch M, Kojadinovic I, Meyer P (2008) A review of methods for capacity identification in choquet integral based multi-attribute utility theory applications of the kappalab r package. *Eur J Oper Res* 186(2):766–785. <https://doi.org/10.1016/j.ejor.2007.02.025>
 46. Cui Y, Che W, Liu T et al (2020) Revisiting pre-trained models for Chinese natural language processing. In: Findings of the association for computational linguistics: EMNLP 2020. Association for Computational Linguistics Online, pp 657–668. <https://doi.org/10.18653/v1/2020.findings-emnlp.58>
 47. Jiang Q, Chen L, Xu R et al (2019). In: A challenge dataset and effective models for aspect-based sentiment analysis. In: Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP). Association for Computational Linguistics, Hong Kong, China, pp 6280–6285. <https://doi.org/10.18653/v1/D19-1654>
 48. Wu J, Wang S, Chiclana F et al (2022) Two-fold personalized feedback mechanism for social network consensus by uninorm interval trust propagation. *IEEE Trans Cybern* 52(10):11,081–11,092. <https://doi.org/10.1109/TCYB.2021.3076420>
 49. Gai T, Cao M, Chiclana F et al (2022) Consensus-trust driven bidirectional feedback mechanism for improving consensus in social network large-group decision making. *Group Decision and Negotiation* <https://doi.org/10.1007/s10726-022-09798-7>
 50. Chang JL, Li H, Wu J (2022) How tourist group books hotels meeting the majority affective expectations. A group selection frame with kansei text mining and consensus coordinating. *Group Decision and Negotiation*. <https://doi.org/10.1007/s10726-022-09810-0>
 51. Rezaei J (2015) Best-worst multi-criteria decision-making method. *Omega* 53:49–57. <https://doi.org/10.1016/j.omega.2014.11.009>
 52. Guo S, Zhao H (2017) Fuzzy best-worst multi-criteria decision-making method and its applications. *Knowl-Based Syst* 121:23–31. <https://doi.org/10.1016/j.knsys.2017.01.010>
 53. Li Y, Yan X, Song X (2019) Provision of paid web-based medical consultation in China: Cross-sectional analysis of data from a medical consultation website. *Journal of Medical Internet Research* 21(6) e12:126. <https://doi.org/10.2196/12126>
 54. Hu J, Zhang X, Yang Y et al (2019) New doctors ranking system based on vikor method. *Int Trans Oper Res* 27(2):1236–1261. <https://doi.org/10.1111/itor.12569>
 55. Sun R, Hu J, Chen X (2017) Novel single-valued neutrosophic decision-making approaches based on prospect theory and their applications in physician selection. *Soft Comput (Berlin Germany)* 23(1):211–225. <https://doi.org/10.1007/s00500-017-2949-0>
 56. Zhao M, Gao H, Wei G et al (2022) Model for network security service provider selection with probabilistic uncertain linguistic todim method based on prospect theory. *Technol Econ Dev Econ* 28(3):638–654. <https://doi.org/10.3846/tede.2022.16483>
 57. Zhang Z, Li Z (2022) Personalized individual semantics-based consistency control and consensus reaching in linguistic group decision making. *IEEE Transactions on Systems, Man, Cybern Syst* 52(9):5623–5635. <https://doi.org/10.1109/TSMC.2021.3129510>
 58. Kwok PK, Lau HYK (2019) Hotel selection using a modified TOPSIS-based decision support algorithm. *Decis Support Syst* 120:95–105. <https://doi.org/10.1016/j.dss.2019.02.004>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Affiliations

Jian Wu^{1,2} · Guangyin Zhang^{1,2} · Yumei Xing^{1,2} · Yujia Liu^{1,2}  · Zhen Zhang³ · Yucheng Dong⁴ · Enrique Herrera-Viedma⁵

Jian Wu
jyajian@163.com

Guangyin Zhang
aaiuuui@163.com

Yumei Xing
yumeixing2021@163.com

Zhen Zhang
zhen.zhang@dlut.edu.cn

Yucheng Dong
ycdong@scu.edu.cn

Enrique Herrera-Viedma
viedma@decsai.ugr.es

- ¹ School of Economics and Management, Shanghai Maritime University, Shanghai, 201306, China
- ² Center for Artificial Intelligence and Decision Sciences, Shanghai Maritime University, Shanghai, 201306, China
- ³ Institute of Systems Engineering, Dalian University of Technology, Dalian, 116024, China
- ⁴ Business School, Sichuan University, Chengdu, 610065, China
- ⁵ Department of Computer Science and AI, University of Granada, Granada, 18071, Spain