Domain-Adaptive Service Evaluation of Mobile Health Application

Xiaojia Wang (xjwang@hfut.edu.cn)
Hefei University of Technology https://orcid.org/0000-0001-5294-1599

Kuo Du
Hefei University of Technology

Wei Xia
Hefei University of Technology

Shanshan Zhang
The First Affiliated Hospital of Anhui University of Chinese Medicine

Weiqun Xu
The First Affiliated Hospital of Anhui University of Chinese Medicine

Yuxiang Guan
The First Affiliated Hospital of Anhui University of Chinese Medicine

Research article

Keywords: Service evaluation, Mobile health application, Domain adaptive, Machine learning, Delphi

Posted Date: December 5th, 2019

DOI: https://doi.org/10.21203/rs.2.18220/v1

·Domain-Adaptive Service Evaluation of Mobile Health Application

Xiaojia Wang[a*], Ph.D., Kuo Du[a], Ph.D., Wei Xia[a], Ph.D., Shanshan Zhang[b,c], M.D., Weiqun Xu[d,e], M.D., Yuxiang Guan[d,e], M.D.

[a]*Department of Information Management, School of Management, Hefei University of Technology, Hefei, Anhui, China,230009. xjwang@hfut.edu.cn (Xiaojia Wang); 2018110802@mail.hfut.edu.cn (Kuo Du); xiawei@hfut.edu.cn (Wei Xia).*

[b]*Department of Clinical Teaching, The First Affiliated Hospital of Anhui University of Chinese Medicine, Hefei, Anhui, China,230031. 1041810968@qq.com (Shanshan Zhang).*

[c]*Healthcare and Public health Information Center, The First Affiliated Hospital of Anhui University of Chinese Medicine, Hefei, Anhui, China,230031.*

[d]*Department of Endocrinology ,The First Affiliated Hospital of Anhui University of Chinese Medicine, Hefei, Anhui, China,230031. 36342375@qq.com (Weiqun Xu); 782365867@qq.com (Yuxiang Guan).*

[e]*The National Chinese Medicine Clinical Research Base- Key Disease of Diabetes Mellitus Study, Hefei, Anhui, China,230031*

[*] Corresponding author. Xiaojia Wang, E-mail address: xjwang@hfut.edu.cn, Mailbox 270, No. 193, Tunxi Road, Hefei, An Hui Province, China, 230009. Tel.: (551) 62904960, Fax: (551) 62901501.

Domain-Adaptive Service Evaluation of Mobile Health Application

ABSTRACT

Background: In the context of "Internet +" medical treatment, mobile health applications provide services for people in a new way, making it possible for people to carry out health management anytime and anywhere. According to the survey data, the most powerful consumers in the field of mobile health applications are those aged 24 to 35. Thus, it can be seen, it is particularly important to study the preferences of young people for mobile health applications.

Methods: This study established a domain-adaptive mobile health application evaluation model based on users' experience, and used an interactive algorithm combining machine learning and Delphi method to calculate the weight distribution of evaluation factors. Compared with previous studies, the establishment of evaluation index based on user experience of youth groups can more comprehensively measure users' demand for mobile health application service quality. Meanwhile, the mobile health application evaluation system established in this study adopts feedback mechanism to realize dynamic evaluation of mobile health applications.

Results: The cognitive level of information (weighting 52%) was only four percentage points higher than the emotional level (weighting 48%). The importance of the four criteria is content information on cognition (weighting 31%), interaction information on emotion (weighting 29%), interaction information on cognition (weighting 21%), and content information on emotion (weighting 19%) in descending order. Among 20 sub-criteria, less disruptive (weighting 17.8%), security (weighting 10.9%), utility (weighting 9.3%), reliability (weighting 8.1%), navigational (weighting 6.7%) occupy an important position.

Conclusion: We find that the weights assigned to sociability, personalization, aesthetics, and interestingness accounted for a significant proportion of the total weights assigned; however, universality and learnability were poorly weighted. These results have important reference value for the development of mobile health applications.

Keywords: Service evaluation; Mobile health application; Domain-adaptive; Machine learning; Delphi

1. Background

This study was conducted by a health management company that recently plans to develop a mobile health app specifically for young people. Mobile health applications are apps that are installed on small mobile devices such as tablets or smartphones. It can interact with users through contact-based interface to realize health management anytime and anywhere, which makes it an important part of modern medicine. [1] In order to take advantage of the development of the Internet and obtain lasting competitiveness, it is necessary to develop mobile health application. For the sake of confidentiality, we use the fictitious name mobile health management (MHM) instead of the company. MHM is a company in the growth stage, mainly providing nutrition and health consulting, health maintenance services and consulting, weight loss services etc. It takes "Identify problems earlier; Solve problems more effectively; Change health management more innovative" as service tenet and is committed to providing a high-level health management service to meet customers' needs. [2] Although the MHM was founded only five years ago, it has achieved good development. According to their five-year experience in service, senior managers perceive that the youth market is a "fat sheep", and that this generation of young people is significantly different from previous generations in the standard of appraising mobile health applications. Accordingly, they decide to use Target Focus Strategy, which means focusing on the youth market to provide mobile health services in line with the characteristics of contemporary youth and striving to gain competitive advantage in the local market.

In order to achieve this goal, it is particularly important to understand the factors that affect the level of evaluation of mobile health application by young people. It is also of great importance to know the necessity of each influencing factor. Therefore, based on users' experience, we evaluate acceptance of mobile health application by young people from cognitive and emotional perspectives. After reading the literature and making appropriate adjustments according to the requirements of this study, we determined the evaluation indexes. [3] In this study, 870 young people were selected for questionnaire survey, including 452 men and 418 women from 8 different

provinces including Beijing, Hebei and Anhui etc. Among them, there were undergraduates, postgraduates, teachers, civil servants and enterprise employees. Different questions are set for each evaluation factor and the acceptance of mobile health applications. Acceptance degree implies the information about the distribution of evaluation factor weight, because users rate mobile health apps based on their experience. Since machine learning algorithm can automatically learn the implicit weight distribution information in sample data set to calculate the weight of evaluation factors, we use it to calculate the weight of indicators. However, an unavoidable problem with the data obtained from the survey method is that it relies too much on the subjectivity of the respondents, so the survey error will be relatively large. Considering that in the domain adaptive evaluation algorithm, the final index weight is calculated by multiplying the index weights at all levels, we use the prior knowledge of experts to assign weights to the first and second level indexes, so as to reduce the influence of subjectivity of questionnaire data and make the weight calculation more accurate.

Mobile health apps can support continuous health monitoring at the individual and population level, encourage health behaviors to prevent or reduce health problems, support self-health management, increase knowledge, and reduce health care visits. [4] [5] It has been proved to be beneficial to people's health in many aspects, in psychology, disease management, health care. [6] [7]In psychology, mental health apps have potentials in improving the monitoring and management of mental health symptoms or disorders. [8] For example, studies show that low-intensity mindfulness intervention through the use of mobile health apps can effectively improve the psychological distress (depression, anxiety, stress) and sleep dysfunction (subjective sleep quality, sleep latency, habitual sleep efficiency) of Chinese college students. [9] [10] In disease management, mobile health apps can improve education and self-care for patients with complex conditions. [11] It has achieved good results in improving the metabolic index of diseases and strengthening self-management and shown overall positive results in improving physical deficiencies in hypertension, weight, tobacco smoke, diabetes, dyslipidemia and so on.[12] For example, brown demonstrated the feasibility and usability of an interactive, immersive 3d iPad health game for cancer patient disease management through experiments. [13] Baker developed a new medical non-contact control system, which greatly helped patients with physical disabilities.

To realize the function of calling nurses and the effect of controlling hospital equipment. [14]In health care, satisfactory treatment compliance and subsequent health maintenance can be achieved with a high level of exposure to mobile health applications [15] Some apps may have positive benefits when used to deliver exercise or gait training interventions, as self-management systems, or as measurement tools. [16] They are also effective adjuncts to the improvement and maintenance of weight loss after bariatric surgery. [17]Mobile health apps play an important role in improving or maintaining people's physical health, but the quality of mobile health apps on the market today varies greatly. Therefore, it is necessary to evaluate them and design applications that are more acceptable to users, thus bringing greater convenience to people.

At present, many researches have been carried out in the field of mobile health application evaluation, and scholars have proposed different evaluation indexes. We can use perceived usefulness to measure attitude, perceived ease of use to measure satisfaction, and behavioral intention to measure interest and importance to evaluate mobile health applications. [18] [19] Awareness of devices and health information plays an important role in influencing how users rate mobile health applications. [20] Flexibility; performance; competency; learnability; completeness; information; other outcomes; error prevention; flexibility can be used to evaluate the usability of mobile health applications. [21] It has been proved that the quality of information and the credibility of sources played a great role in the evaluation of mobile health applications. [22] Privacy also plays an important role, patients want hospitals to assist them with high efficiency without revealing patients' identities. [23]We can evaluate applications by using the following criteria: cost (free or paid); content (medical information, symptom log etc.); relevant public or private health care; target audiences (health care and non-health care professionals; validity (citing medical information that has been peer-reviewed or otherwise verified); and average user ratings (from1 to 5 stars).[11] Georgsson puts forward a multilevel evaluation model: level 1 is the target system specification, which is used to understand user tasks for system development. Level 2 is a test of task performance to evaluate system validation and human-computer interaction in a laboratory environment. Level 3 aims to combine environmental factors to determine the impact of the system in the real environment [24]

According to the above research results, evaluation indexes are mostly specified from the

application itself. Although they are important for evaluating mobile health applications, they mostly use indicators unrelated to the mobile health application service field. Therefore, we add indicators starting from users' experience, which can better reflect the "user-centered" concept.

In the evaluation method of mobile health application, it can be roughly divided into two categories. One is to conduct relatively simple statistical analysis of the data by designing questionnaires，the other is to build a model and then analyze the data. Brown uses Dedoose qualitative data analysis(QDA) software, a web-based qualitative and hybrid data analysis program, to analyze collected data and visualize the results to evaluate mobile health applications.[21] Battle uses a comprehensive analysis, which combine qualitative and quantitative methods to evaluate mobile health applications. In the qualitative part, data are collected by means of targeted sampling and Emi-structured interview, then they do thematic analysis. In the quantitative part, data are collected, calculated with SAS 9.2, and then analyzed and compared. [25] Liu estimated the influence of each quality dimension on continuous use intention，by statistical analysis of the survey responses of 191 users. [26] Meng combined health awareness with refined likelihood model to test people's daily willingness to use mobile health applications. [22] Mattias puts forward a method to enhance cognitive roaming (the CW), which means take the user as the center of the CW (UC - the CW) when solve defects that have been found in the original technology, and verify by thinking about the statement agreement. In order to assess the effectiveness, efficiency, and self-management of mobile health applications among users with diabetes. [24] Mattias tested the usability of mobile health applications by using the framework analysis (FA) approach and usability issues taxonomy (UPT). It is also shown that using this method can help identify usability problems, especially initial usability problems. [27] Paulo used a mixed method of partial least squares and fuzzy set qualitative comparative analysis to evaluate the number of causal condition combinations that promote mHealth adoption. [28] After the above summary, we can find that the current evaluation methods are mainly simple statistical analysis, or the establishment of simple qualitative and quantitative methods to analyze the obtained data, which are traditional and do not reflect intelligence, and do not consider the authenticity of the data.

Therefore, this paper proposes a mobile health application evaluation model based on user

experience, which uses ontology language to describe the relationship between concepts in the evaluation model, makes different evaluation factors through domain experts, and calculates the weight distribution of evaluation factors, according to the knowledge collected by the user system, by machine learning algorithm meanwhile it supports optimize the weight by the prior knowledge of the domain experts through Delphi method .

Our contributions of this paper are mainly reflected in two aspects. On the one hand, in terms of evaluation indexes, the current evaluation models of mobile health application use evaluation factors unrelated to service fields, which cannot comprehensively measure users' requirements for the quality of mobile health APP service. In this study, evaluation indexes are set from the perspective of users' experience, which can better reflect users' recognition degree of mobile health application. On the other hand, in terms of evaluation methods, an interactive algorithm combining machine learning and Delphi method is adopted to calculate the weight distribution of evaluation factors. This algorithm is proved not only to speed up the calculation process of weights, but also to improve the accuracy of weight distribution. Meanwhile, the evaluation framework of this study supports adding the feedback of the results into the new sample set, in order to realize the dynamic adjustment of weight distribution, and adjust the evaluation index with the knowledge of domain experts to realize the dynamic evaluation.

2. Methods

2.1. Mobile health application evaluation system

In order to build the evaluation model of mobile health application, we design a system as the prototype system of the evaluation model mentioned above. We call it Mobile Health Application Evaluation System (MHAES), which can dynamically adjust the evaluation model according to specific conditions.

Fig.1 Mobile health application evaluation system (MHAES)

As shown in Fig. 1, MHAES is mainly composed of three parts. They are the data layer, the expert layer, and the user layer. The first part is the data layer, including ontology query/storage

engine and relational database, which are used to query and store ontology data. The second part is the expert layer, including evaluation model manager, evaluation category constructor, evaluation factor constructor and weight algorithm engine, which is used to construct evaluation model, evaluation category and evaluation factor, and calculate the weight distribution of evaluation category and evaluation factor. The third part is the user layer, including data receiver, result publisher, data preprocessor and evaluation engine for evaluating mobile health applications.

The expert layer of MHAES realizes the construction of the evaluation model, which mainly uses the prior knowledge and sample data acquired by experts to construct the evaluation model. The construction process can be divided into the following steps: a) The domain experts create the evaluation factor in the evaluation model manager. Construct new evaluation factor through the valence factor constructor according to the information contained in the evaluation results output by the user layer. Then save them in the data layer by the ontology query engine. b) Experts score the evaluation categories to obtain the scoring results. Preprocessed the experts' scoring data and sample data and store them in data layer by ontology query engine. c) The weight algorithm engine obtains expert scoring data and sample data from the ontology query engine to calculate the weight of each evaluation category and evaluation factor, and store the calculated results in the ontology storage engine.

The user layer of MHAES implement the function of evaluating mobile health applications. Specifically, the original data used for evaluation is obtained from the user system, and the evaluation engine is used to calculate the quality of mobile health application service, and the evaluation results are given. The data receiver obtains the data provided by the evaluation user system, the data preprocessor completes the preprocessing of the original data, and the evaluation

engine obtain the evaluation factors and the pre-processed evaluation data from the ontology storage engine. After all these preparatory activities are completed, the evaluation engine uses the obtained information to evaluate, then store the results in the database. At the same time, the result publishing module can receive the results and publish them to decision makers, this information can help them improve the functions of the APP. Meanwhile, it can feed back to experts, so that they can dynamically adjust the evaluation model according to the changes of user experience, in order to make the evaluation standard more reasonable.

2.2. Data collection and evaluation indicators system

2.2.1. Data collection

870 young people were selected in this study for questionnaire survey, which includes 452 men and 418 women from 8 different provinces. Among them, there were undergraduates, postgraduates, teachers, civil servants and enterprise employees. At the same time, it is necessary to introduce the evaluation experts in Delphi method. It is composed of 20 experts with rich experience in mobile health application, including experts from Hefei university of technology (5 people), Hebei university (3 people), Beijing University of Technology (5 people), and employees from mobile health apps company (7 people). Some of the data used to support the findings of this study are included within the article.

2.2.2. Evaluation indicators system

In order to facilitate readers to better understand the evaluation model of mobile health application, this part will explain the definitions of terms involved in our model.

Evaluation Factor: Evaluation factor is a basic evaluation unit extracted from users' experience (including cognition and emotion), which is composed of a binary group: $f = (attr, W)$,

where $attr$ represents the name of users' experience, $w \in [0,1]$ represents the weight of evaluation factor, reflecting the importance of this evaluation factor in the evaluation of mobile health application.

In this paper, reliability is an important aspect to describe users' experience and influence users' evaluation of APP, which can be defined as the evaluation factor of mobile health application. Domain experts can customize various types of evaluation factors in this model. Specific explanations of indicators are shown in Table A in the appendix. To facilitate the use of prior knowledge to calculate the weight of evaluation factors, we give the concept of evaluation categories.

Evaluation Category: Evaluation category represents a conceptual clustering of evaluation factors, which is composed of triples: $S = (Sn, Set, W)$,where $Set = \{e_1, e_2, ..., e_k, e_n, 1 \le k \le n\}$, (e_k is the evaluation factor or evaluation category); Sn is the name of the evaluation category; $w \in [0,1]$ represents the weight of the evaluation category , reflecting the importance of the evaluation category in service evaluation.

In this paper, cognition-based information quality evaluation can be used as an evaluation category S , the Sn item of which is "based on cognition", the set item is $\{S(attr = "Cognitive\ \ Content"), S(attr = "Cognitive\ \ Interaction")\}$ including two secondary evaluation categories: the content information that mainly affects cognition and the interaction information that mainly affects cognition.

Evaluation Concept Tree: Evaluation concept tree T is a tree structure composed of evaluation factors and evaluation categories as nodes. Evaluation categories are internal nodes of evaluation concept trees and evaluation factors are leaf nodes of evaluation concept trees. The hierarchy of

evaluation concept tree is composed as follows: Each element in the set item of evaluation category s is the child node of evaluation category s; all nodes have a unique parent except for the evaluation category at the root node. Meanwhile, the weights of nodes in the evaluation concept tree have the following constraints:

a) The weight of root node is 1;

b) The weight of any evaluation class node s is the sum of all the weight of its children.

In this paper, Fig. 2 shows the evaluation concept tree of this study. The root node of the evaluation concept tree has two sub-evaluation categories (based on cognition and based on emotion), which contain different sub-evaluation categories. Weights start at the root node and are distributed layer by layer to the lower nodes and end at the leaf node. Different weight distribution of evaluation factors reflects different evaluation standards of users. Since different types of mobile health applications have different evaluation concept trees, it is necessary for domain experts to customize evaluation concept trees for mobile health applications in this field.

Fig.2 Evaluation concept tree

2.2.3. Semantic description diagram of evaluation indicators system

Mobile health application evaluation model is composed of quintuples $M = [T, SI, DI, ER, P]$ where T represents evaluation concept tree; SI represents the instance of the survey being evaluated; DI represents a data instance of SI, which contains multiple data items. Each data item represents the value of data instance DI on an evaluation factor. ER represents the output of evaluation model and records the quality evaluation result of DI. P is the relation set of the evaluation model, representing various constraint relations among concepts in the evaluation model. Fig.3 shows the semantic description graph of the evaluation model of mobile health

application in this study. Evaluation concept tree represents the tree of evaluation T, category represents the category of evaluation, factor represents the factor of evaluation, and predicate {has Root, consist of} jointly defines hierarchical relation of evaluation concept tree. Survey instance represents the instance of survey SI , data instance represents the instance of data DI, evaluation result represents the result of evaluation ER), predicate {from Tree, has Factor} represents the relationship between Data instance DI and evaluation concept tree T, while predicate from service represents data instance DI from service instance SI and evaluation result ER is evaluation result of service instance SI.

Fig.3 Semantic description diagram of mobile health application evaluation model

According to the above description of the evaluation model and its components. We assume that the evaluation concept tree T has n evaluation factors, and the weight of each evaluation factor is, and the value of DI, a data instance of the investigation instance SI, on each evaluation factor is $\{V_1, V_2, ..., V_n\}$, in that way the evaluation of mobile health application by survey instance SI is the weighted sum of the values of DI on each evaluation factor, given in equation (1)

Calculation formula of evaluation:

$$Quality(SI) = \sum_{i=1}^{n} V_i W_i \tag{1}$$

It can be seen from equation (1) that the evaluation results of mobile health applications are determined by the value of evaluation factors and the weight distribution of evaluation factors. The value of the evaluation factor is derived from user experience, and the weight distribution of the evaluation factor can be calculated according to the user's feelings in the process of use and the prior knowledge of domain experts, which is obtained from the domain experts with the help of the evaluation concept tree. In the next section, we will calculate the weight distribution.

2.3. Method of evaluation

Generally, evaluation factors will have different value ranges. Therefore, in order to ensure the comparability of the weights of evaluation factors, it is necessary to preprocess the evaluation data and adjust the range of all evaluation factors to a uniform interval. However, in this study, due to the scale used in the survey, the number of each evaluation factor is between 1 and 5, so the weight can be calculated directly without standardization.

In the weight calculation, to make the weight distribution of evaluation factors close to the evaluation standard of users, we not only used the machine learning algorithm to calculate the weight of each evaluation factor, but also introduced the prior knowledge of experts by Delphi method to optimize the weight distribution.

2.3.1. Machine learning algorithm

The machine learning algorithm can automatically learn the implicit weight distribution information in sample the data set to calculate the weight of the evaluation factors.

In this study, 870 young people were selected for the questionnaire survey, and the options were set for each evaluation factor and the acceptance level of the mobile health application. Because users rate mobile health apps based on their experience, the acceptance degree implies the information about the weight distribution of the evaluation factor. We represent a group of sample data as a binary group: $S =< DI, d >$; DI represents the evaluation data of a survey instance, and d has two values: d_1=True and d_2=False, thus indicating that respondents like (dislike) this candidate instance. In the machine learning algorithm, information gain represents the degree of influence of different evaluation factors on the service selection results, so this model uses information gain to calculate the weight of the evaluation factors.

Set S_1 as a sample subset of d $= d_1$ in sample dataset S and a sample set of candidate instances favored by the respondents (specific data are shown in Table B.1 in the appendix) and set S_2 (specific data are shown in Table B.2 in the appendix) as a sample subset of d= d_2 in sample dataset S, representing the unselected sample set of service instances; thus, S= S_1 and S_2. We classify the sample data set according to the value of d term, and the expected information obtained is given by equation (2):

$$I(S_1, S_2) = -P_1 Log_2(P_1) - P_2 Log_2(P_2) \tag{2}$$

where p_1 is the probability that d is equal to d_1 for any sample in the sample dataset S and P_2 is the probability that d is equal to d_2 in any sample of S. Use S_1/S and S_2/S to describe the sample subsets. Suppose S^i_{K1} is the sample set of subset S^i_k ,where $"d = d_1"$, and suppose S^i_{k2} is the sample set of subset S^i_k where "d= d_2"; then, the expected information is given by equation (3):

$$E(f_i) = \sum_{k=1}^{m} \frac{S^i_{k1} + S^i_{k2}}{S} I(S^i_{k1}, S^i_{k2}) \tag{3}$$

where, $I(S_0, S_1, ..., S_n, f_i)$ can be calculated according to formula (2), $P_1 = S^i_{k1} / S^i_K$ and $P_2 = S^i_{k2} / S^i_K$ .Then the information gain obtained by the evaluation factor is given by equation (4):

$$Gain(f_i) = I(S_1, S_2) - E(f_i) \tag{4}$$

Equation (4) can be used to calculate the information gain of each evaluation factor, and equation (5) calculates the weight of each evaluation factor based on equation (4):

$$W_i = Gain(f_i) \Big/ \sum_{j=1}^{n} Gain(f_i) \tag{5}$$

By calculating the information gain of the evaluation factors, we can learn the weight distribution of each evaluation factor from the sample set. After sorting out the data of 870 questionnaires, we divide the data set into two categories according to the attitude of the

respondents toward the mobile health application: attitude ratings of 4 and 5 are placed in one category, and those from 1 to 3 are placed in another category. According to the machine learning algorithm, the weights of third-level indexes are calculated, and the results are shown in Fig. 4.

Fig.4 Weight distribution of secondary indicators

2.3.2. Delphi method

Because the information used to construct the sample set is gained by using the knowledge generated from the data collected from the survey, the evaluation result depends heavily on the size of the sample set and the quality of the sample. Therefore, the evaluation effect obtained is somewhat accidental. In fact, the domain experience of domain experts also implies the information of the weight distribution of the evaluation factors; as a result, adding the domain expert's prior knowledge to the calculation not only accelerates the calculation of the weight but also improves the accuracy of the weight distribution. The prior knowledge referred to in this study is the empirical knowledge of domain experts of the importance of different evaluation categories, which can be used to calculate the weight distribution of the evaluation factors. Therefore, we adopt the Delphi method to introduce the knowledge of experts.

We score the indicators by sending emails to experts and then process the scoring results by using a statistical approach. In the expert questionnaire, the importance degree is assigned based on the five subscales method. Because the familiarity of the experts with the five-subscale index varies and because the bases for their judgments are inconsistent, we need to use the degree of familiarity and the basis of judgment to distinguish the authority of experts; namely, the authority coefficient is equal to the average degree of familiarity and determines the arithmetic coefficient. The authority coefficient of experts is generally required to be larger than 0.7, and the degree of

expert authority in this study meets this requirement.

Table 1 Authority of expert

In this study, experts were consulted mainly on the weight of the first- and second-level index. Because of the small number of indicators and the low decision-making complexity, 20 experts are enough. The specific steps were as follows: the questionnaire was developed based on the literature research and evaluation indexes determined by experts in previous fields. After the Delphi questionnaire was collected, the questionnaire was analyzed. The Kendall coordination coefficient of the first-level index in the first round of the letter consultation was 0.716 ($\chi^2 = 42.955,\ \ p < 0.001$), thus indicating a high degree of expert opinion coordination and no need for further letter consultation. The weights of the first- and second-level indicators are calculated; the weights are shown in table 2. (For convenience, we simplify the representation of the indicators: “C” is an abbreviation for “based on cognition”; “CC” is an abbreviation for “Cognitive content”; “CI” is an abbreviation for “Cognitive interaction”; “E” is an abbreviation for “based on emotion”; “EC” is an abbreviation for “emotional content”; and “EI” is an abbreviation for “emotional interaction.”）

Table 2 The weight of primary and secondary indicators

2.3.3. Interactive algorithm combining Machine Learning and the Delphi Method

To effectively use prior knowledge, we define the local weight of the nodes in the evaluation concept tree. The local weight $W'(N,M)$ from node N to its ancestor node M is defined to represent the weight of node N in the evaluation concept subtree with node M as the root node. When node M is the parent node of node N, $W'(N,M)$ is also called the nearest local weight of node N, denoted by the symbol $W'(N)$. Meanwhile, the local weight of the root node and the

nearest local weight of the evaluation concept tree are both 1. The local weight of a node is a measure of its importance in a local range. According to the definition of local weight, we can regard the weight of node N as the local weight $W'(N, root)$ of node N in the evaluation concept tree. The weights of the nodes can be used to calculate the local weights.

***Definition 1*:** If the node N_m is the child node of the node N_0, the sequence of nodes on the path from N_0 to N_m is, $(N_0, N_1, N_2, ..., N_m)$ .In addition, suppose their nearest local weight is $(W_0', W_1', ..., W_m')$;then, $W'(N_m, N_0) = W_1' W_2' ... W_m'$.

In definition 1, if the node N_0 is the root node of the evaluation concept tree, then the node N_0 is the weight of the node. From definition 1 we get definition 2.

***Definition 2*:** The weight of node N is equal to the product of the nearest local weight of each node in the path from node N to the root node.

Set the node sequence on the path from any evaluation factor to root of the root node of the evaluation concept as $(S_0, S_1, ..., S_n, f_i)$ (where n≥0 and S_0=root), and set the nearest local weight of $S_0, S_1, ..., S_n$ as $W_0', W_1', ..., Wn'$ respectively, because S_0=root, W_0'=1. From definition 1, we can obtain the local weight of the node f_i to a node S_k sequence:

$$W'(f_i, s_k) = W_{k+1}' ... W_n' W'(f_i) \quad 0 \le k \le n \tag{6}$$

From definition 2, it can be concluded that the weight of the evaluation factor is

$$W(f_i) = W_0' ... W_n' W'(f_i) \tag{7}$$

By substituting equation (6) for equation (7), a new formula for calculating the weight of the evaluation factors can be obtained:

$$W(f_i) = W_0' ... W_k' W'(f_i, S_k) \tag{8}$$

Formula (8) gives a new method to calculate the weight of the evaluation factor: If the nearest

local weight of the first k ancestor nodes of the evaluation factor and a local weight of the evaluation factor are calculated, the weight value of the evaluation factor can be obtained. The nearest local weight of the k ancestor nodes before the evaluation factor can be calculated by the prior knowledge of domain experts, while the local weight of the sub-tree with the ancestor node of the k layer as the root node can be calculated by machine learning algorithm.

A detailed description of this method is given below. In this model, by referring to the Delphi method, we determine the weights of the first- and second-level indexes by consulting 15 experts by letter. Importance was rated on a five-level scale: 1 means “very unimportant”; 2 means “not important”; 3 means “important”; 4 means “more important”; and 5 means “very important”. Under the condition that the degree of coordination of opinions of all items meets the requirement (Kendall coordination coefficient > 0.7), the average value assigned by each expert for each indicator is calculated as the basis for determining the weight. The average score of each indicator is divided by the average score of all indicators at the same level as the weight. Set the child nodes of node N as $N_1, N_2, \ldots, N_m$, the scores assigned by k experts to the nodes as $L_{i1}, L_{i2}, \ldots, L_{ik}$, the average scores of the nodes N_i as Z_i, and the weight of the nodes N_i as W_i. According to the above description,

$$Z_i = \sum_{b=1}^{k} L_{ib} / K \tag{9}$$

$$W_i = Z_i \Big/ \sum_{i=1}^{m} Z \tag{10}$$

Algorithm implementation:

Step1: Initialization, node variable v=root, current level counting variable c=0, add to H;

Step2: If ((c<k) inverted v (v is not an evaluation factor))

Domain experts set the importance level for each child node of v, and calculate the nearest local

weight of each child node by equations (9) and (10). Traverse each child node of V;

Repeat step 2;

Else {under the subtree T with v as the root node, machine learning algorithm is adopted to calculate the local weight of each evaluation factor on T'}

Ethical considerations

Ethical approval for all data collection was obtained by the ethics committee (EC) of Hefei University of Technology. The EC waived the mandate for obtaining a written informed consent from subjects. Participants were provided with an information sheet which detailed relevant information about the study, potential benefits and risks of participation in this study, the opportunity and means to ask questions, and the options regarding voluntary agreement to participate in this study. Verbal consent was then requested prior to commencement of the survey. This study was provided as an anonymous survey of individuals for which no personal, identifiable information was collected.

3. Results

According to the analysis, the weight distribution of indicators at all levels is obtained, as shown in the table below.

Table 3 Weight distribution of each indicator

To show the results of the weights more clearly and compare the weights of each index more intuitively, the results are expressed in the form of a rectangular tree graph. See Fig 5 for details.

Fig.5. The weight distribution represented by the rectangular tree graph

Different weights represent different degrees of importance. Next, we will analyze the results from the perspective of three levels of indicator results. The first section of the analysis identifies the importance of two primary indicators. Both cognitive and emotional information are important

in influencing users' evaluation of mobile health apps. The cognitive level of information (which was 0.52) was only four percentage points higher than the emotional level (which was 0.48).”

The mobile health application designed for the youth group shows that we should consider not only users' cognitive needs but also their emotional needs. Therefore, on the one hand, we need to be able to cognitively provide users with information that is as they would expect it to be: accurate, reliable, easy to understand, concise, and effective in problem solving. On the other hand, we need to provide users with information that is expressed, as they would expect, in various forms according to different needs. This is a chance to design an app according to users’ preferences and that provides information that indeed provides real benefits to users.

In the second part of the analysis, the importance of four secondary indicators is determined. The effects of content information on cognition, interaction information on emotion, interaction information on cognition, and content information on emotion are, in order of importance, 0.31, 0.29, 0.21, and 0.19, respectively.” At the cognitive level, content information is more important than interactive information. At the emotional level, interactive information is more important than content information. Therefore, when designing mobile health applications, different indicators should be emphasized at different levels.

In the third portion of the analysis, the importance of 20 three-level indicators is determined. These indicators, in descending order of importance, are as follows: less disruptive (0.178), security (0.109), utility (0.093), reliability (0.081), navigational (0.067), sociality (0.061), personalized (0.058), profitable (0.058), accuracy (0.053), beauty (0.046), interesting (0.035), simplicity (0.034), understandability (0.034), fault tolerance (0.023), operational (0.019), fluency (0.015), multiple performance (0.012), correlation (0.009), generality (0.009), and learnability

(0.006).

4. Discussion

The weight distribution is relatively concentrated, with the weight ratio of the first 10 indicators exceeding 0.8 and that of the first 5 indicators exceeding 0.5. Therefore, when designing mobile health applications, it is necessary to pay special attention to the key indicators.

Somewhat surprisingly, user's value having fewer distractions the most. However, because young people are very averse to being interrupted by useless information, it is reasonable that having fewer distractions is such a high priority for them. Therefore, efforts should be made to ensure that there is no distracting information (advertisements, ambiguous information, etc.) in the content provided. If such information must be added because of commercial interests, we should consider the form of the advertising and try our best to add only a small amount of soft advertising. The second most important indicator is security. Therefore, developers need to take effective measures to protect the privacy of users. Developers should reduce system vulnerabilities, improve security, and prevent the system from being breached by criminals. The third most important indicator is utility, which provides information to solve problems faced by users and provides reliable suggestions to users. Utility is the most basic requirement for mobile health apps. If a mobile health app does not have utility, it does not matter how well the app performs in other respects. Therefore, enterprises need to integrate internal and external high-quality resources to improve service quality. The fourth most important indicator is reliability: the source of content must be authoritative and reliable and must clearly label who is the diagnostician and whether that person has the relevant qualifications. Reliability is the basis for achieving utility. Only

information provided from a reliable source can instill in users a sense of security and help them find the right professional in case of an accident. The fifth most important indicator is navigability: the system should have a simple navigation bar or other convenient navigation mode to help users quickly find the information they need.

In this study, we also found that the weights assigned to the four secondary indicators—sociality, personalized, beauty and interesting—accounted for a significant proportion of the total weights assigned, thus exceeding our expectations. The significance of these four secondary indicators is also an important finding of this study and provides a new direction for the design of mobile health applications aimed at young people. Concerning sociality, special modules can be set up in the app to provide opportunities for users to communicate through the app and bridge the communication between users, patients and other relevant persons. “Personalized” indicates that the service and interface of the system can be personalized and optimized according to users' preferences so that users can participate in the design of the interface and information architecture and have some control over the system. Personalization is the embodiment of young people's pursuit of individuality. “Beauty,” indicates that the system interface and information presentation mode are aesthetically rich. Meanwhile, it is also of great importance that color, font, and graphics are used in the proper proportions; beautiful formats are designed so that information can be presented through pictures, cartoons and other interesting forms. “Interesting” indicates that the design and operation mode of the system is interesting and that the makers of the app avoided creating a rigid design and an outdated interaction mode. Makers of interesting apps appropriately adjust and update the system promptly and follow development trends. Meanwhile, the research results also show that the importance of universality and learnability is very low, thus indicating

that the first generation of young people to grow up with the Internet has no difficulty using mobile health apps. Young people do well in using different apps, so they pay less attention to versatility and learnability.

Our study also has some limitations. First, in terms of indicators, we set only 20 indicators and asked respondents to score based on the existing indicators; however, the target valued by the user may not appear in the selected option. Second, in terms of data volume, we collected only 870 questionnaires, so the dataset is not large enough. If existing software data are used to analyze big data, the results will be more accurate. We expect people with access to big data to study these issues in the future and provide better advice for developers, so that they can create more user-friendly products.

5. Conclusion

This paper adopts a comprehensive evaluation method to help a mobile health application company to determine the indicators that should be focused on in the development of mobile health applications targeted at youth. The research methods and ideas in this paper are of great significance and inspiration for other enterprises to make similar decisions. We set up an evaluation index based on the experience of users; therefore, this index better reflects the service concept of “user-centric” Information is collected from both experts and users, and an interactive algorithm combining machine learning and the Delphi method is adopted to assign weights to the established indicators to correct errors in user data by using the prior knowledge of experts. We were surprised to find that the weights assigned to sociability, personalization, aesthetics, and interestingness accounted for a significant proportion of the total weights assigned; however,

universality and learnability were poorly weighted. These results have important reference value for the development of mobile health applications.

6. Declarations

Abbreviations

MHM	Mobile Health Management
CC	Cognitive Content
CI	Cognitive Interaction
EC	Emotional Content
EI	Emotional Interaction

Ethics approval and consent to participate

Ethical approval for all data collection was obtained by the ethics committee (EC) of Hefei University of Technology. The EC waived the mandate for obtaining a written informed consent from subjects. Participants were provided with an information sheet which detailed relevant information about the study, potential benefits and risks of participation in this study, the opportunity and means to ask questions, and the options regarding voluntary agreement to participate in this study. Verbal consent was then requested prior to commencement of the survey. This study was provided as an anonymous survey of individuals for which no personal, identifiable information was collected.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by a grant from the Key Disease of Diabetes Mellitus Study Center at the National Chinese Medicine Clinical Research Base, the China Scholarship Council, the National Natural Science Foundation of China Grant No. 61876055, 71101041, and the National Statistics Research Projects Grant No. 2013LZ07, Basic scientific research and cultivation plan of Central University Grant No. JS2019HGXJ0038 and National Steering Committee for Graduate Education of Chinese Medicine and Traditional Chinese Medicine Grant No. 20190723-FJ-B39. The funders did not have any role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors' contributions

XJW, KD and SSZ conceived of this study and participated in the design and administration of the study. WX and SSZ contributed to data acquisition. WQX and YXG analyzed the data and the results. XJW and KD drafted the manuscript. SSZ, WX and WQX supervised the work and revised the results critically. All authors read and approved the final version of the manuscript.

Acknowledgements

None.

References

1. Appel G, Libai B, Muller E, Shachar R: **On the monetization of mobile apps**. *International Journal of Research in Marketing* 2019.
2. Zhao P, Yoo I, Lancey R, Varghese E: **Mobile applications for pain management: an app analysis for clinical usage**. *BMC Med Inform Decis Mak* 2019, **19**(1):106.
3. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, Gelaude D, Mosley JP, Travers J: **A user-centered model for designing consumer mobile health (mHealth) applications (apps)**. *J Biomed Inform* 2016, **60**:243-251.

4. Huang G, Ren Y: **Linking technological functions of fitness mobile apps with continuance usage among Chinese users: Moderating role of exercise self-efficacy**. *Computers in Human Behavior* 2020, **103**:151-160.
5. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, Riley WT, Shar A, Spring B, Spruijt-Metz D *et al*: **Mobile Health Technology Evaluation**. *American Journal of Preventive Medicine* 2013, **45**(2):228-236.
6. Robillard JM, Feng TL, Sporn AB, Lai JA, Lo C, Ta M, Nadler R: **Availability, readability, and content of privacy policies and terms of agreements of mental health apps**. *Internet Interv* 2019, **17**:100243.
7. Hamblion EL, Burkitt A, Lalor MK, Anderson LF, Thomas HL, Abubakar I, Morton S, Maguire H, Anderson SR, Public Health England TBSTB: **Public health outcome of Tuberculosis Cluster Investigations, England 2010-2013**. *J Infect* 2019, **78**(4):269-274.
8. Wang K, Varma DS, Prosperi M: **A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders**. *J Psychiatr Res* 2018, **107**:73-78.
9. Hall BJ, Xiong P, Guo X, Sou EKL, Chou UI, Shen Z: **An evaluation of a low intensity mHealth enhanced mindfulness intervention for Chinese university students: A randomized controlled trial**. *Psychiatry Res* 2018, **270**:394-403.
10. **<Z1.pdf>**.
11. Ayyaswami V, Padmanabhan DL, Crihalmeanu T, Thelmo F, Prabhu AV, Magnani JW: **Mobile health applications for atrial fibrillation: A readability and quality assessment**. *Int J Cardiol* 2019.
12. de Castro M, Sawatzky J-A: **Mobile Health Interventions for Primary Prevention of Cardiovascular Disease**. *The Journal for Nurse Practitioners* 2018, **14**(8):e165-e168.
13. Brown-Johnson CG, Berrean B, Cataldo JK: **Development and usability evaluation of the mHealth Tool for Lung Cancer (mHealth TLC): a virtual world health game for lung cancer patients**. *Patient Educ Couns* 2015, **98**(4):506-511.
14. Baker SB, Xiang W, Atkinson I: **Internet of Things for Smart Healthcare: Technologies, Challenges, and Opportunities**. *IEEE Access* 2017, **5**:26521-26544.
15. Dounavi K, Tsoumani O: **Mobile Health Applications in Weight Management: A Systematic Literature Review**. *American Journal of Preventive Medicine* 2019, **56**(6):894-903.
16. Nussbaum R, Kelly C, Quinby E, Mac A, Parmanto B, Dicianno BE: **Systematic Review of Mobile Health Applications in Rehabilitation**. *Arch Phys Med Rehabil* 2019, **100**(1):115-127.
17. Mangieri CW, Johnson RJ, Sweeney LB, Choi YU, Wood JC: **Mobile health applications enhance weight loss efficacy following bariatric surgery**. *Obes Res Clin Pract* 2019, **13**(2):176-179.
18. Balapour A, Reychav I, Sabherwal R, Azuri J: **Mobile technology identity and self-efficacy: Implications for the adoption of clinically supported mobile health apps**. *International Journal of Information Management* 2019, **49**:58-68.
19. Kutia S, Chauhdary SH, Iwendi C, Liu L, Yong W, Bashir AK: **Socio-Technological Factors Affecting User's Adoption of eHealth Functionalities: A Case Study of China and Ukraine eHealth Systems**. *IEEE Access* 2019, **7**:90777-90788.

20. Pai RR, Alathur S: **Assessing awareness and use of mobile phone technology for health and wellness: Insights from India**. *Health Policy and Technology* 2019.
21. Brown W, 3rd, Yen PY, Rojas M, Schnall R: **Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology**. *J Biomed Inform* 2013, **46**(6):1080-1087.
22. Meng F, Guo X, Peng Z, Zhang X, Vogel D: **The routine use of mobile health services in the presence of health consciousness**. *Electronic Commerce Research and Applications* 2019, **35**:100847.
23. Mehmood A, Natgunanathan I, Xiang Y, Poston H, Zhang Y: **Anonymous Authentication Scheme for Smart Cloud Based Healthcare Applications**. *IEEE Access* 2018, **6**:33552-33567.
24. Georgsson M, Staggers N, Arsand E, Kushniruk A: **Employing a user-centered cognitive walkthrough to evaluate a mHealth diabetes self-management application: A case study and beginning method validation**. *J Biomed Inform* 2019, **91**:103110.
25. Battle JD, Farrow L, Tibaijuka J, Mitchell M: **mHealth for Safer Deliveries: A mixed methods evaluation of the effect of an integrated mobile health intervention on maternal care utilization**. *Healthc (Amst)* 2015, **3**(4):180-184.
26. Liu F, Ngai E, Ju X: **Understanding mobile health service use: An investigation of routine and emergency use intentions**. *International Journal of Information Management* 2019, **45**:107-117.
27. Georgsson M, Staggers N: **An evaluation of patients' experienced usability of a diabetes mHealth system using a multi-method approach**. *J Biomed Inform* 2016, **59**:115-129.
28. Duarte P, Pinho JC: **A mixed methods UTAUT2-based approach to assess mobile health adoption**. *Journal of Business Research* 2019, **102**:140-150.

Figure legends:

Fig.1 Mobile health application evaluation system (MHAES)

We design a system as the prototype system of the evaluation model mentioned above. We call it Mobile Health Application Evaluation System (MHAES), which consists of three layers can dynamically adjust the evaluation model according to specific conditions.

Fig.2. Evaluation concept tree

The root node of the evaluation concept tree has two sub-evaluation categories (based on cognition and based on emotion), which contain different sub-evaluation categories. Weights start at the root node and are distributed layer by layer to the lower nodes and end at the leaf node.

Fig.3 Semantic description diagram of mobile health application evaluation model

It shows the semantic description graph of the evaluation model of mobile health application in this study.

Fig.4 Weight distribution of secondary indicators

It shows the weights of third-level indexes according to the machine learning algorithm.

Fig.5. The weight distribution represented by the rectangular tree graph

It shows the results of the weight of each evaluation index more clearly and compare the weights of each index more intuitively.

Table list:

Table 1

Authority of expert

	degree of familiarity（Cs）	criterion（Ca）	authority coefficient（Cr）
first round of inquiry	0.88	0.865	0.8725

It shows that the authority coefficient of experts is generally required to be larger than 0.7, and the degree of expert authority in this study meets this requirement.

Table 2

The weight of primary and secondary indicators

control criteria	weight	criteria	weight
C	0.52	CC	0.31
		CI	0.21
E	0.48	EC	0.19
		EI	0.29

The weights of the first- and second-level indicators are calculated by Delphi.

Table 3

Weight distribution of each indicator

control criteria	weight	criteria	weight	sub-criteria	weight
Based on cognition	0.52	Cognitive content	0.31	utility	0.093
				simplicity	0.034
				accuracy	0.053
				reliability	0.081
				understandability	0.034
				learnability	0.006
				correlation	0.009
		Cognitive interaction	0.21	navigational	0.067
				fluency	0.015
				security	0.109
				operational	0.019
Based on emotion	0.48	emotional content	0.19	multiple performance	0.012
				less disruptive	0.178
		emotional interaction	0.29	beauty	0.046
				personalized	0.058
				fault tolerance	0.023
				interesting	0.035

				sociality	0.061
				profitable	0.058
				generality	0.009

According to the analysis, the weight distribution of indicators at all levels is obtained.

Additional Files:

Additional file A.

Table A Explanation of Each Indicator

Description: The table describes the specific meaning of 20 three-level indicators in the evaluation index system, which helps us better understand each indicator.

Additional file B.

Table B.1 Part of the data from set1

Description: The table is excerpted from the data set1. Set1 is related to various indicators of people who scored 4 or 5 according to their attitude towards health application.

Table B.2 Part of the data from set2

Description: The table is excerpted from the data set2. Set2 is related to various indicators of people who scored 1-3 according to their attitude towards health application.

Figures

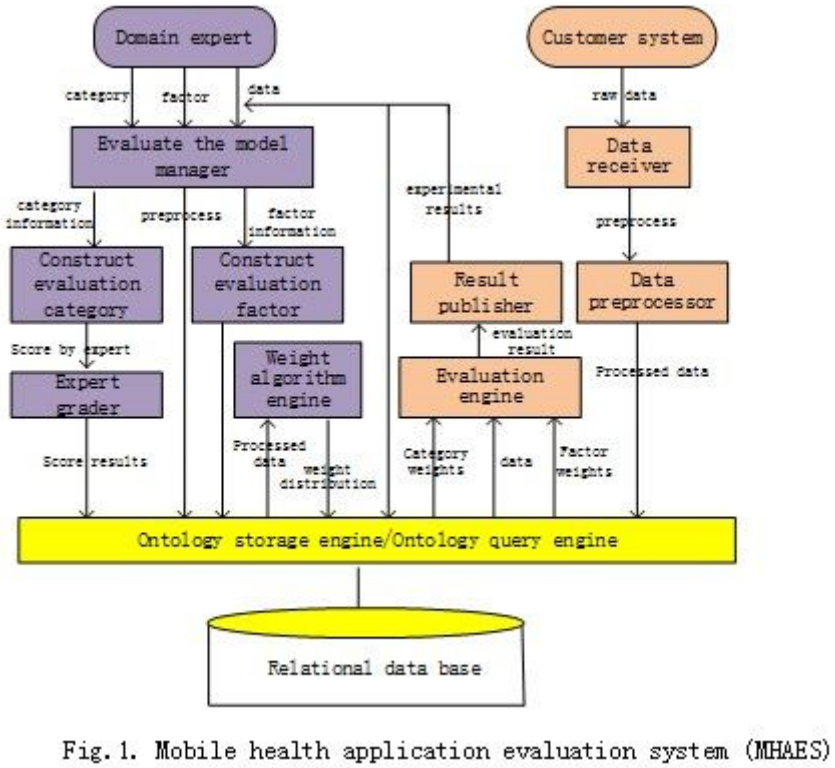

Fig. 1. Mobile health application evaluation system (MHAES)

Figure 1

Mobile health application evaluation system (MHAES). We design a system as the prototype system of the evaluation model mentioned above. We call it Mobile Health Application Evaluation System (MHAES), which consists of three layers can dynamically adjust the evaluation model according to specific conditions.

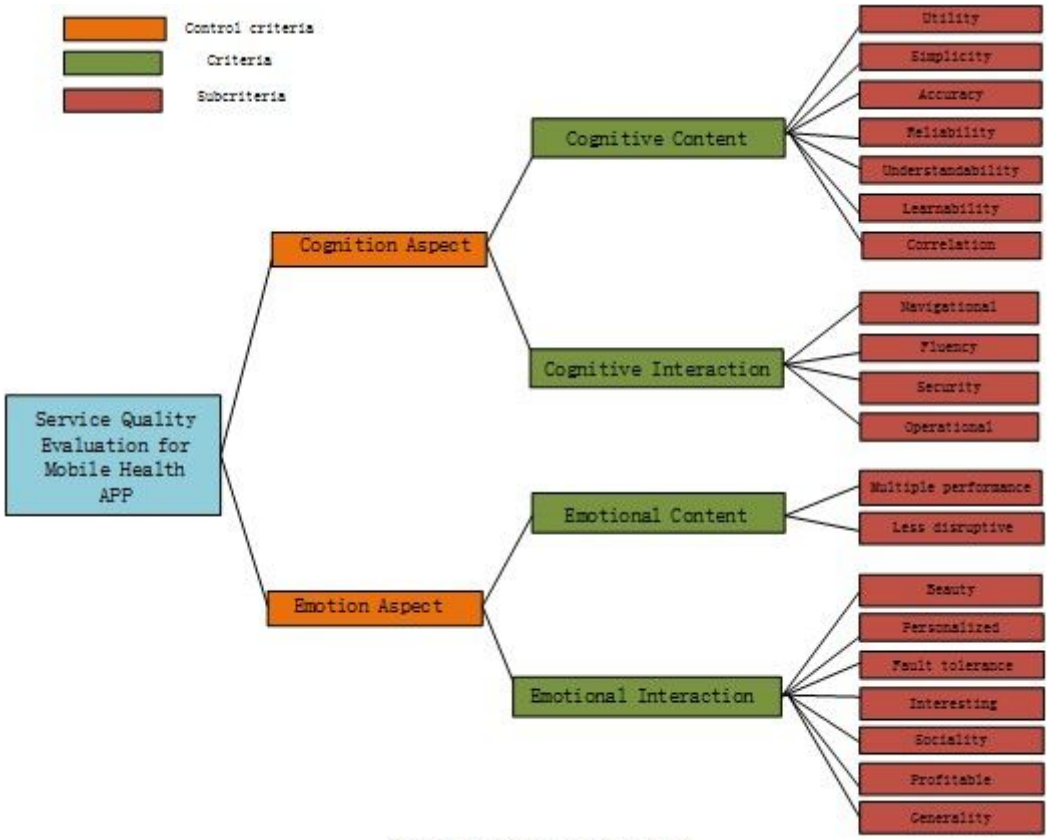

Fig.2. Evaluation concept tree

Figure 2

Evaluation concept tree. The root node of the evaluation concept tree has two sub evaluation categories (based on cognition and based on emotion), which contain different sub evaluation categories. Weights start at the root node and are distributed layer by layer to the lower nodes and end at the leaf node.

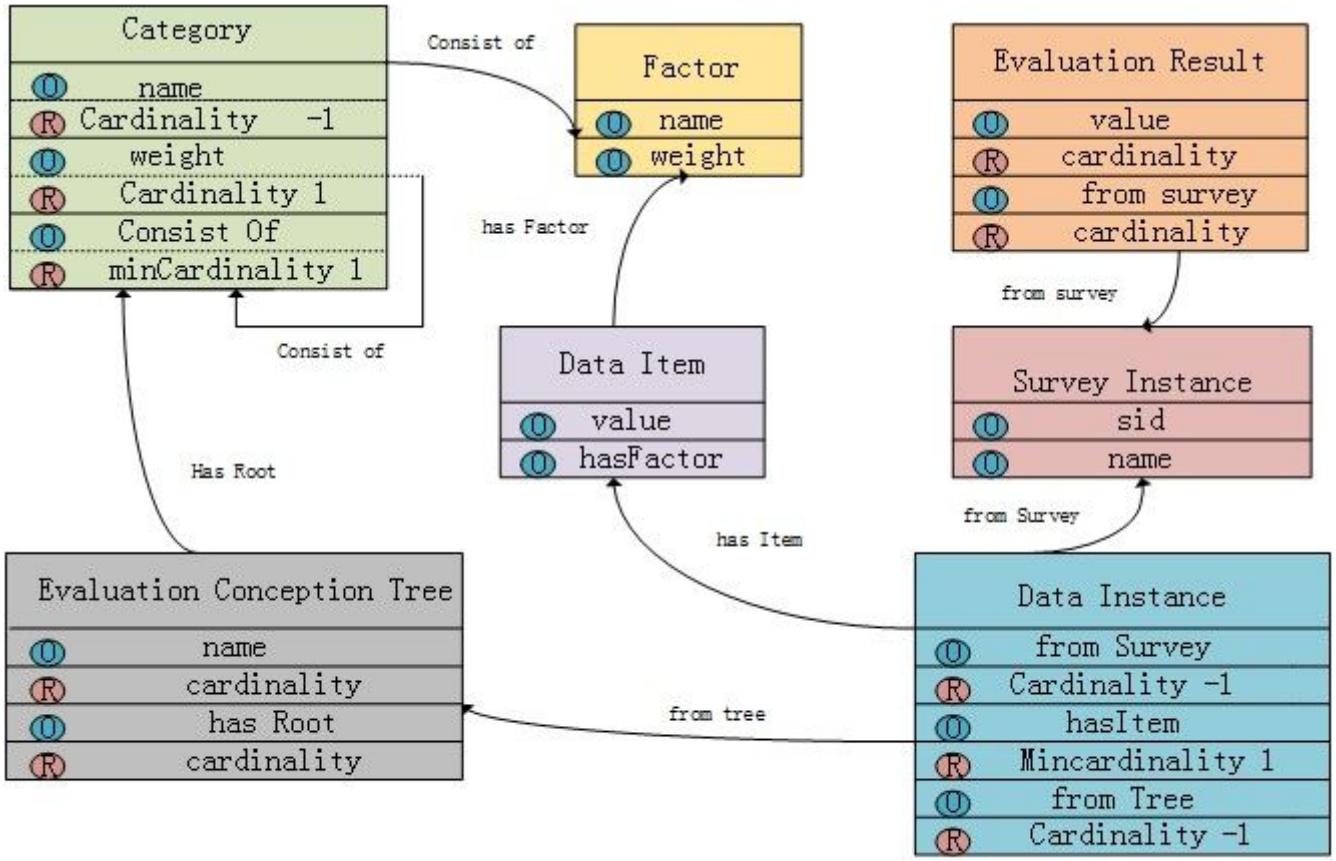

Fig.3. Semantic description diagram of mobile health application evaluation model

Figure 3

Semantic description diagram of mobile health application evaluation model. It shows the semantic description graph of the evaluation model of mobile health application in this study.

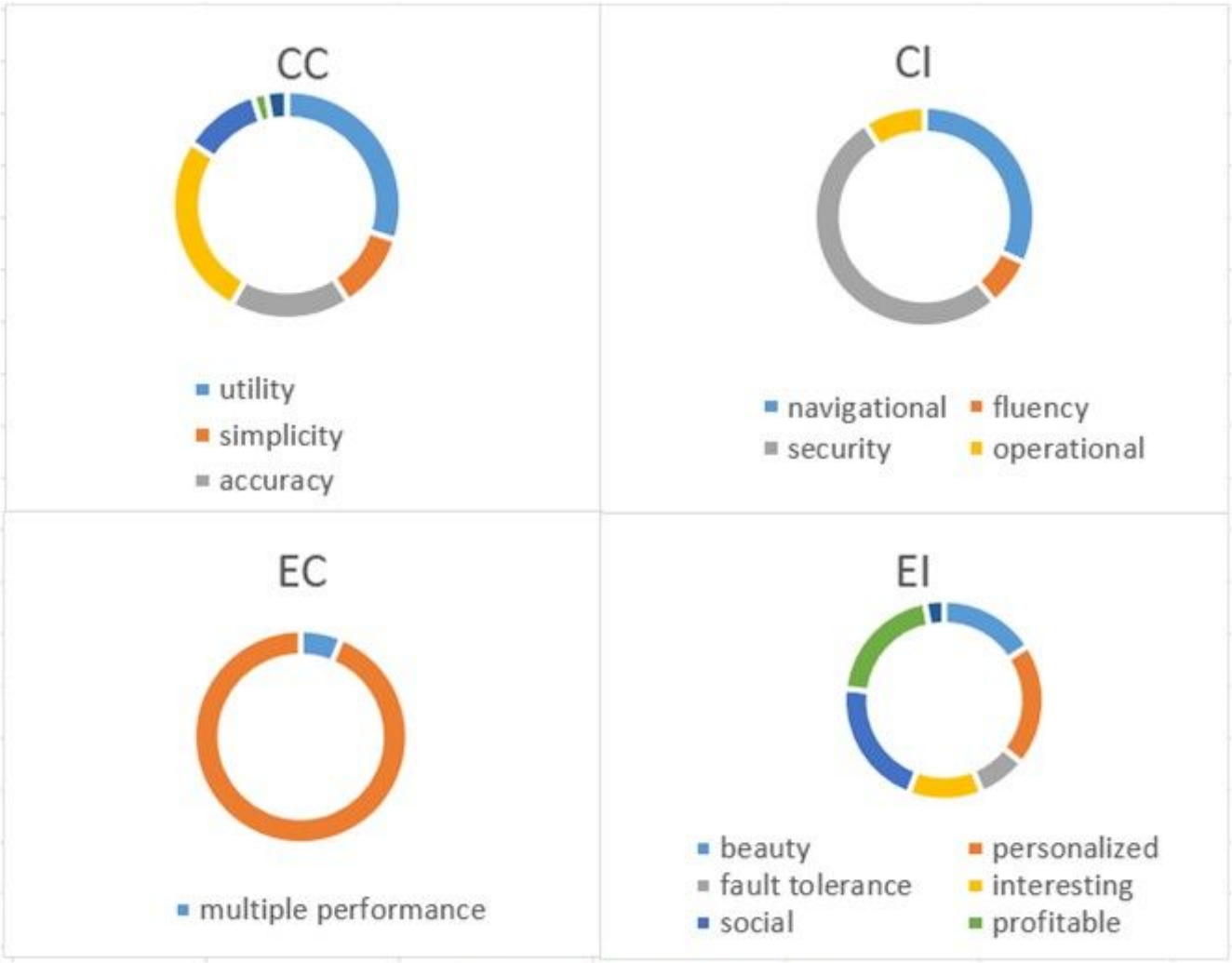

Fig.4. Weight distribution of secondary indicators

Figure 4

Weight distribution of secondary indicators. It shows the weights of third level indexes according to the machine learning algorithm.

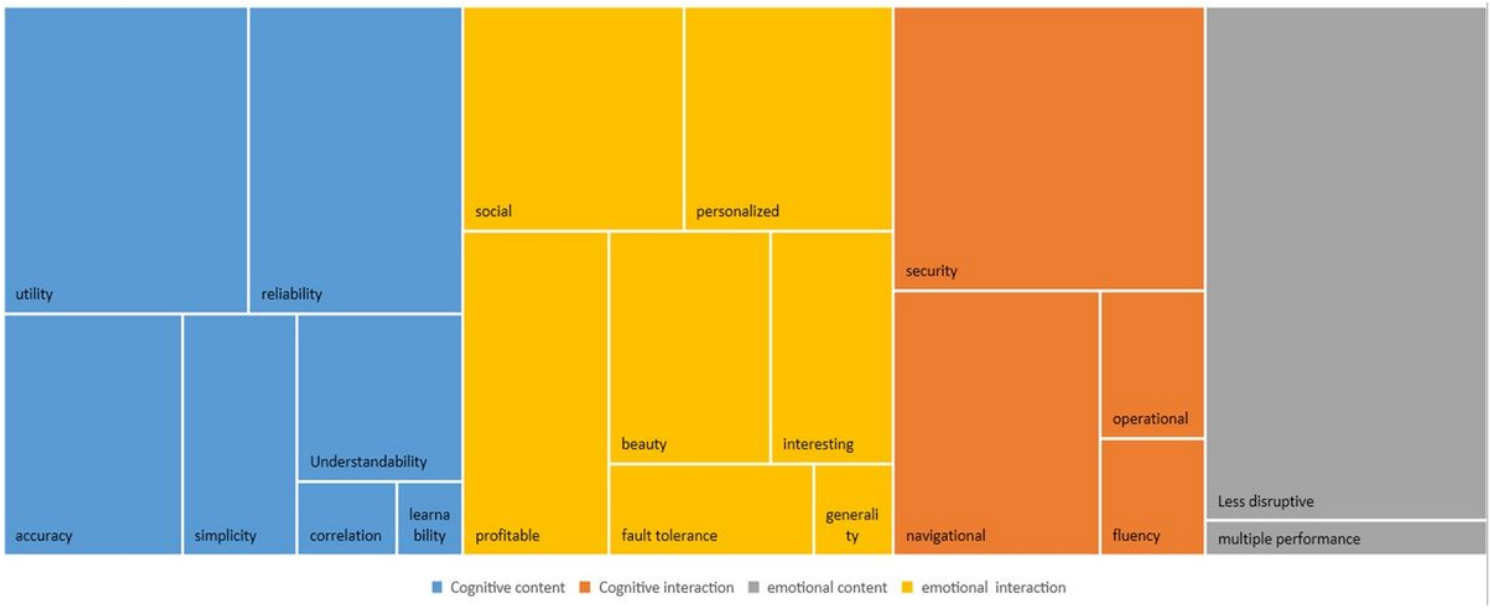

Fig.5 The weight distribution represented by the rectangular tree graph

Figure 5

The weight distribution represented by the rectangular tree graph. It shows the results of the weight of each evaluation index more clearly and compare the weights 20 of each index more intuitively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TABLEI.tif
- TABLEIII.tif
- TABLEII.tif
- SUPPLEMENTARYMATERIAL.docx