A novel evaluation approach to user psychological experience performance based on ELECTRE method

Gao-Ming DU

Psychology of Education, Neijiang Normal University, Neijiang, P. R. China

_____

ABSTRACT

The research to user psychological experience performance has an important theoretical significance. For this reason, a novel evaluation approach is proposed to the user psychological experience performance based on the ELECTRE method. The experimental results suggest that this approach is feasible, correct and valid.

Key words: Evaluation Approach, ELECTRE Method, Experience Performance

_____

INTRODUCTION

The rapid development of information, computer science and technology as well as knowledge-based economy has forced the implementation of experience economy [1]. User experience is a kind of holistic subjective psychological feelings formed when users accept service [2-3]. Researchers in several fields such as computer science, psychology and sociology have made researches on user experience from their own perspectives and application backgrounds. When users are served, various factors in users themselves, computer application systems and current objective environment will all affect the quality of user experience. Most of the current study on user experience focuses on computer application systems rarely considers users' own psychology and cognition and cannot improve user experience from their mind and psychology, which is a question in need of deep exploration.

Researchers in psychology, sociology and computer science have made qualitative and quantitative analysis on factors influencing the quality of user experience from different focuses. In the qualitative aspect, some researchers think that macro factors can be classified from such aspects as surface, framework, structure, range and strategy, and micro factors mainly include organization, reconstruction and expression of service contents. In the quantitative aspect, researchers proposed some quantitative indexes to measure user experience. For example, Jim Lewis designed CSUQ and PSSUQ [4-5]. CSUQ is aimed at e-mail or on-line tests and PSSUQ is aimed at face-to-face tests. These two questionnaires are mainly used to analyze system effectiveness, information quality, interface quality and overall satisfaction [6]. Arnie Lund designed usefulness, satisfaction and ease questionnaires (USE) containing thirty rated items and four features such as usefulness, satisfaction, ease and learnability.

The integrated research result shows that, factors influencing user experience quality include the following aspects. Usability which can be further divided into response rate, easy to learn, easy to navigate, simple operation, consistent operation and visual appeal. Usefulness is further divided into meeting user demand, improving user working efficiency and creating value for users. Other factors such as ones aimed at users' characteristics, individual information data and emotion etc.

It has become the hotspot in this field to study user experience from the aspect of users' cognitive patterns and emotional features. For example, consider users' cognition, behavior and psychology before the consideration of usability and then propose some principles and psychological research approaches qualitatively. However, this is only theoretical and lack of detailed feasible quantitative evaluation approaches and experiments.

_____

EVALUATION INDEX SYSTEMS
There are several factors influencing user experience and the most important are usefulness and usability.

(1) Usefulness, also called effectiveness, refers to whether the service contents systems provide and recommend are useful or not. It is featured with:

1) Resource coverage, namely the coverage rate of the resource that systems provide on the knowledge domain that users should learn.

2) Hit ratio of recommendation. The hit ratio is defined as the ratio between the hit count and the recommendation count during users' one login process. Users have learnt the content recommended by systems, and if the evaluation score on the content is bigger than the corresponding threshold value, it is called a recommendation hit. If the overall count that systems recommend contents to users is n_r and the total count of recommendation hits is n_s, then the hit ratio of recommendation is the ratio between the hit count and recommendation count.

$$R_H = \frac{n_s}{n_r} \tag{1}$$

3) User loyalty. It is defined as the willingness that user uses the system once again and can be measured by calculating the frequency of user's visit. According to the standard time number N defined based on experts' experience, if the user uses the system n_u times each week, the user loyalty is

$$L = \begin{cases} 1, & n_u \geq N \\ \dfrac{n_u}{N}, & n_u < N \end{cases} \tag{2}$$

(2) Usability. It refers to whether the operation, learning, navigation and use that the user should grasp during the system service accord with user's habits. Usability of e-learning system owns following features.

1) Response rate. The time interval between user's service request and service provision of the system is defined as response time. It can be obtained through analyzing system log data and the response rate can be obtained through calculating response time:

$$t_{RS} = \begin{cases} 1, & (\overline{t_{re}} < t_{ts}) \\ \dfrac{t_{ts}}{\overline{t_{re}}}, & (\overline{t_{re}} > t_{ts}) \end{cases} \tag{3}$$

In which, t_{re} is the response time, namely the time that user login spends, and t_{ts} is the standard response time.

2) Navigation definition. During the use of the system, the ratio between the number of visited target knowledge and that of all the target knowledge is the navigation definition. It is defined as:

$$N_C = \frac{N}{M} \tag{4}$$

In which, M is the number of knowledge items visited by the user when he/she has completed the task successfully and N is the number of knowledge items successfully grasped by the user. If the user finds the target knowledge directly, navigation definition is 100%. If the user cannot find the target knowledge, it is 0%. Thus it can be seen that, the smaller M is, the better the system navigation performance is, and users can easily find desired knowledge items.

3) Task-performing efficiency. During the use of the system, the ratio between the total time spent in learning target knowledge and that spent in using the system is task-performing efficiency. T is one learning task for users designed in the e-learning system and users should learn one or more knowledge items, and then the task-performing efficiency is

$$E_T = \frac{\sum t_{ek}}{\sum t_{en}}, e_k \in E \tag{5}$$

Here, t_{ek} is the learning time, namely the time interval between the click on the kth knowledge item e_k and the click

on the next one, also the time spent in learning knowledge item e_k. t_{en} is the time spent in successfully learning the nth knowledge item of N ones. E is the set of target knowledge items, namely the set of knowledge items users should learn during one learning task, $E = \{e_1, e_2 \cdots\}$.

The login time is t_{in}. The task starting time, namely the time when the system completed loading and starts to present knowledge, is t_s. The time interval from system login to the moment when the system is loaded is the response time t_{re}.

$$t_{re} = t_s - t_{in} \tag{6}$$

E-learning system provides users learning resources with knowledge items. The least time users spend in learning target knowledge items is the least time duration. If one user's learning time has not reached this least time duration, the user is considered to have not completed the corresponding learning task. Least time durations corresponding to knowledge items $e_1, e_2, \cdots$ is $t_{e1}, t_{e2}, \cdots$ which is defined by expert teachers or depends on the average learning time duration among user populations.

If the time duration t_{ek} one user spends in learning knowledge item e_k is bigger than the corresponding least time duration t_{ekm} and the user thinks he/she has successfully grasped the knowledge, namely that the learning for knowledge item e_k has been completed, the successful learning completion function is valued with 1, otherwise it is 0. The corresponding expression is

$$S_{ek} = \begin{cases} 1, & (t_{ek} \geq t_{ekm}) \\ 0, & (t_{ek} \leq t_{ekm}) \end{cases} \tag{7}$$

If one user thinks he/she has completed the current learning task when he/she has successfully learned N knowledge items of E, the successful task completion function is valued with 1, otherwise it is 0. The corresponding expression is

$$T_{S_{ek}} = \begin{cases} 1, & \sum S_{ek} \geq N, e_k \in E \\ 0, & \sum S_{ek} < N, e_k \in E \end{cases} \tag{8}$$

The time interval from the moment that the task is successfully loaded to the moment of task completion is the task time, namely

$$T_t = \sum T_{en} \tag{9}$$

(3) Evaluation Index System. E-learning user psychological experience is a kind of holistic subjective psychological feelings formed when users use the system for learning. According to the analysis of this section, the overall user experience quality is featured with usefulness and usability which can be further divided. The hierarchical structure of features of the e-learning user psychological experience quality is shown in Figure 1.
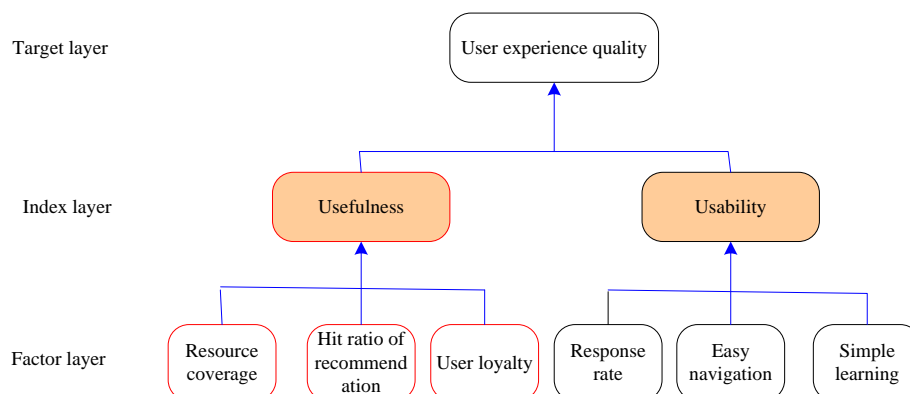

Fig. 1. Overall model of e-learning user experience

_____

THE PROPOSED APPROACH
To ensure the clarity and universality of the evaluation, C_j $(j=1, 2,…,m)$ are supposed to be benefit attributes. p_j, q_j and v_j respectively represent the seriously superior threshold, no difference threshold and the veto threshold between solutions in C_j. As to the first threshold, a_i is considered to be strictly better than a_k if p_j is smaller than the difference between a_i and a_k according to C_j. As to the second threshold, no difference exists between a_i and a_k if q_j is bigger than the difference between a_i and a_k according to C_j. And as to the last threshold, the overall level of a_i is vetoed to be better than a_k if v_j is no bigger than the difference between a_k and a_i according to C_j (a_k is bigger than a_i). Therefore, $0 \leq q_j \leq p_j \leq v_j$.

(1) Definitions of Three Indexes
1) Concordance index, represented with $c(a_i, a_k)$, refers to the supporting degree for the inference "the level of a_i is higher than that of a_k". It can be defined as

$$c(a_i, a_k) = \frac{1}{w} \sum_{j=1}^{m} w_j c_j(a_i, a_k) \tag{10}$$

In which

$$w = \sum_{j=1}^{m} w_j \tag{11}$$

$$c_j(a_i, a_k) = \begin{cases} 0, & if \ \ y_j(a_i) + p_j \leq y_j(a_k) \\ 1, & if \ \ y_j(a_i) + q_j \geq y_j(a_k) \\ \dfrac{y_j(a_i) + p_j - y_j(a_k)}{p_j - q_j}, & other \end{cases} \tag{12}$$

The index $c_j(a_i, a_k)$ refers to the superiority degree of a_i compared with a_k as to attribute C_j.

2) Discordance index, represented with $d_j(a_i, a_k)$, refers to the denying degree for the inference "the level of a_i is higher than that of a_k". It can be defined as

$$d_j(a_i, a_k) = \begin{cases} 0, & if \ \ y_j(a_i) + p_j \geq y_j(a_k) \\ 1, & if \ \ y_j(a_i) + v_j \leq y_j(a_k) \\ \dfrac{y_j(a_k) - y_j(a_i) - p_j}{v_j - p_j}, & other \end{cases} \tag{13}$$

3) Reliability index, represented with $s(a_i, a_k)$, refers to the credibility for the inference "the level of a_i is higher than that of a_k". It can be defined as

$$s(a_i, a_k) = \begin{cases} c(a_i, a_k), & if \ \ \forall C_j, d_j(a_i, a_k) \leq c(a_i, a_k) \\ c(a_i, a_k) \times \prod_{c_j} \dfrac{1 - d_j(a_i, a_k)}{1 - c(a_i, a_k)}, & if \ \ \forall C_j, d_j(a_i, a_k) > c(a_i, a_k) \end{cases} \tag{14}$$

(2) Construct Preference Matrix Based on Qualitative Evaluation. Usually, different decision-makers assess problems from different angles and own different preferences, so they often adopt different measures and standards to make decisions. After sorting alternatives through the ELECTRE-III method, we cannot integrate single decision-making results to generate the group result solely through the average method or Borda method, because it is unrealistic. Therefore, a new method should be proposed in which different measures and standards can be reflected during the group ranking process. These measures and standards are closely related with the ranking order of alternatives and the value of reliability indexes, so we should study the following two problems at the same time: 1) Differences between alternatives when reliability indexes are respectively $s(a_k, a_i)$ and $s(a_i, a_k)$. 2) The ranking order of the alternative when the reliability index is $s(a_i, a_k)$.

Suppose $\Delta Q_l(a_i, a_k)$ represents the ranking order of the alternative as to the reliability index $s(a_i, a_k)$ and $Q_l(a_i)$ refers to the ranking order of a_i made by the lth person. If there are m alternatives for ranking and the lth person considers that a_i is the best one, then $Q_l(a_i)=1$. If a_i is considered to be the worst one, then $Q_l(a_i)=m$. The relative difference between the ranking order of a_i and that of other alternatives $\Delta Q^l(a_i, a_k)$ can be represented with:

$$\Delta Q_l(a_i, a_k) = \begin{cases} Q_l(a_i) \gg Q_l(a_k), & if \quad Q_l(a_k) - Q_l(a_i) > \dfrac{n}{2} \\ Q_l(a_i) > Q_l(a_k), & if \quad 0 < Q_l(a_k) - Q_l(a_i) \le \dfrac{n}{2} \\ Q_l(a_i) = Q_l(a_k), & if \quad Q_l(a_k) - Q_l(a_i) = 0 \\ Q_l(a_i) < Q_l(a_k), & if \quad 0 < Q_l(a_i) - Q_l(a_k) \le \dfrac{n}{2} \\ Q_l(a_i) \ll Q_l(a_k), & if \quad Q_l(a_i) - Q_l(a_k) > \dfrac{n}{2} \end{cases}$$

In which, $Q_l(a_i) \gg Q_l(a_k)$ refers to that the alternative a_i is far better than a_k according to the lth person, $Q_l(a_i) > Q_l(a_k)$ refers to that the alternative a_i is a little better than a_k according to the lth person, $Q_l(a_i) = Q_l(a_k)$ refers to that the alternative a_i is similar to a_k according to the lth person, $Q_l(a_i) < Q_l(a_k)$ refers to that the alternative a_i is a little worse than a_k according to the lth person and $Q_l(a_i) \ll Q_l(a_k)$ refers to that the alternative a_i is far worse than a_k.

In traditional ELECTRE-III method, when the superior sequence of all alternatives should be determined, the outrnaking relation exists only when $s(a_i, a_k) \ge \lambda_{max} - \beta$ in which $\lambda_{max} = \max_{a_i, a_k \in A} s(a_i, a_k)$, β is the capture threshold and its value is set by decision-makers according to their own experience. Among all outranking relations, the quality of alternative a_i depends on the difference between the numuber of its flow-out directed arcs and that of flow-in directed arcs. Therefore, the relative zone between the reliability indexes $s_l(a_i, a_k)$ and $s_l(a_k, a_i)$ as to each pair of alternatives can be defined as:

$$Zone\ \text{I}: \quad s_l(a_i, a_k) \ge \lambda_{max} - \beta, \quad s_l(a_k, a_i) \ge \lambda_{max} - \beta$$
$$Zone\ \text{II}: \quad s_l(a_i, a_k) \ge \lambda_{max} - \beta, \quad s_l(a_k, a_i) < \lambda_{max} - \beta$$
$$Zone\ \text{III}: \quad s_l(a_i, a_k) < \lambda_{max} - \beta, \quad s_l(a_k, a_i) \ge \lambda_{max} - \beta$$
$$Zone\ \text{IV}: \quad s_l(a_i, a_k) < \lambda_{max} - \beta, \quad s_l(a_k, a_i) < \lambda_{max} - \beta$$

2) Construct the preference evaluation matrix (shown in Table 1). It is a matrix used to evaluate the superior sequence of alternative a_i and a_k based on the given relative sequence location $\Delta Q^l(a_i, a_k)$ of (a_i, a_k) and the relative zone between reliability indexes $s_l(a_i, a_k)$ and $s_l(a_k, a_i)$.

Table 1 Preference Evaluation Matrix

	$Zone\ \text{I}$	$Zone\ \text{II}$	$Zone\ \text{III}$	$Zone\ \text{IV}$
$Q_l(a_i) \gg Q_l(a_k)$	$a_i O_s a_k$	$a_i O_s a_k$	$a_i O_w a_k$	$a_i O_s a_k$
$Q_l(a_i) > Q_l(a_k)$	$a_i O_w a_k$	$a_i O_s a_k$	$a_i I a_k$	$a_i O_w a_k$
$Q_l(a_i) = Q_l(a_k)$	$a_i I a_k$	$a_i O_w a_k$	$a_i B_w a_k$	$a_i I a_k$
$Q_l(a_i) < Q_l(a_k)$	$a_i B_w a_k$	$a_i I a_k$	$a_i B_s a_k$	$a_i B_w a_k$
$Q_l(a_i) \ll Q_l(a_k)$	$a_i B_s a_k$	$a_i B_w a_k$	$a_i B_s a_k$	$a_i B_s a_k$

In which, $a_i O_s a_k$ refers to that alternative a_i is far better than (O_s) a_k, $a_i O_w a_k$ refers to that alternative a_i is a little better than (O_w) a_k, $a_i I a_k$ refers to that there is no difference between alternative a_i and a_k (I), $a_i B_w a_k$ refers to that alternative a_i is a little worse than (B_w) a_k and $a_i B_s a_k$ refers to that alternative a_i is far worse than (B_s) a_k.

(3) Consistent reliability, inconsistent reliability and net reliability

1) Consistent reliability $\Phi^+(a_i)$ refers to the superior degree of alternative a_i, defined as:

$$\Phi^+(a_i) = \sum_{a_k \in A} s(a_i, a_k), \forall a_k \in A \tag{15}$$

2) Inconsistent reliability refers to the superior degree of other alternatives as to a_i, defined as:

$$\Phi^-(a_i) = \sum_{a_k \in A} s(a_k, a_i), \forall a_k \in A \tag{16}$$

3) Net reliability $\Phi(a_i)$ is the difference between the consistent reliability and inconsistent reliability, defined as:

$$\Phi(a_i) = \Phi^+(a_i) - \Phi^-(a_i) \tag{17}$$

Quality of all alternatives can be ranked in certain order according to the value of net reliability $\Phi(a_i)$. Generally speaking, the bigger the value of net reliability $\Phi(a_i)$ is, the better the corresponding alternative is.

(4) Construct the group outranking relationship
1) Group concordance index $C_G(a_i, a_k)$ refers to the reliability that the inference "alternative a_i is better than a_k" is supported from the angle of groups. It can be defined as:

$$C_G(a_i, a_k) = \sum_{l=1}^{m} v_l f_l(a_i, a_k) \tag{18}$$

In which

$$f_l(a_i, a_k) = \begin{cases} 1, & if \quad a_i O_s a_k \\ 0.75, & if \quad a_i O_w a_k \\ 0.5, & if \quad a_i I a_k \\ 0.25, & if \quad a_i B_w a_k \\ 0, & if \quad a_i B_s a_k \end{cases} \tag{19}$$

2) Group discordance index $D_G(a_i, a_k)$ refers to the reliability that the inference "alternative a_i is better than a_k" is denied from the angle of groups. It can be defined as:

$$D_G(a_i, a_k) = \sum_{l=1}^{m} v_l g_l(a_i, a_k) \tag{20}$$

In which

$$g_l(a_i, a_k) = \begin{cases} 1, & if \quad a_i B_w a_k \ or \ a_i B_s a_k, \quad Q_l(a_i) << Q_l(a_k) \\ 0, & others \end{cases} \tag{21}$$

3) Group reliability index $S_G(a_i, a_k)$ refers to the reliability of the inference "alternative a_i is better than a_k" and can be defined as:

$$S_G(a_i, a_k) = C_G(a_i, a_k) \cdot (1 - D_G(a_i, a_k)) \tag{22}$$

(5) Ranking of group decision-making
1) Group consistent reliability $\Phi_G^+(a_i)$ refers to the reliability that alternative a_i is better than others from the angle of groups and can be defined as:

$$\Phi_G^+(a_i) = \sum_{k=1}^{n} S_G(a_i, a_k) \tag{23}$$

2) Group inconsistent reliability $\Phi_G^-(a_i)$ refers to the reliability that other alternatives are better than ai from the angle of groups and can be defined as:

$$\Phi_G^-(a_i) = \sum_{k=1}^{n} S_G(a_k, a_i) \tag{24}$$

3) Group net reliability $\Phi_G(a_i)$ refers to the difference between $\Phi_G^+(a_i)$ and $\Phi_G^-(a_i)$ and can be defined as:

$$\Phi_G(a_i) = \Phi_G^+(a_i) - \Phi_G^-(a_i) \tag{25}$$

Quality of all alternatives can be ranked in certain order according to the value of $\Phi_G(a_i)$. Generally speaking, the bigger the value of $\Phi_G(a_i)$ is, the better the corresponding alternative is.

_____

RESULTS

Usually, pairwise comparison is adopted to decide the feature weights, namely obtaining features with high weights through comparing any two features those are attached to some index. Scale like 1~9 is used for measurement during the comparison. Weight judgment matrix of features can be obtained only through experts' experience, but different experts own different views, so consistency check should be made on the matrix.

As to usefulness indexes, the feature vector is $W = (0.38, 0.15, 0.92)$, and after normalization it turns to $W = (0.26, 0.10, 0.64)$ of which the values are respectively corresponding to such relative weights as resource coverage, hit ratio of recommendation and user loyalty. As to usability indexes, the feature vector obtained through normalizing the feature vector is $W = (0.10, 0.64, 0.26)$ of which the values are respectively corresponding to such relative weights as response rate, navigation definition and task-performing efficiency.

Total ranking weight refers to the relative weight of all elements in each layer as to the total target in the hierarchical structure. It is indicated that the consistency passes the test and the relative weight $W = (0.75, 0.25)$.

The approach mentioned in this paper is used to make an analysis on the user psychological experience in e-learning system used in junior classes in a Chinese university. 110 learners are selected stochastically. The learning task required in the experiment is to learn circuit switching and packet switching in e-learning system.

The total target knowledge item set that needs learning is E = [packet switching, message switching, packet, circuit switching, data switching technology]. Suppose that the least time duration learners spend in learning some target knowledge item is equal to half of the average time the learner population spends in learning this item, shown in Table 2.

Table 2 The least time learners spend in learning target knowledge items

Number of elements	Contents of element	t_{ekm}/s	$\overline{t_{ek}}/s$
1	Packet switching	14	28
2	Message switching	8	16
3	Packet	10	20
4	Circuit switching	6	13
5	Data switching technology	3	6

According to experts' experience, it is considered that the learning task is completed when learners have grasped three knowledge items of E. The standard times of loyalty is 2, namely the times learners use e-learning system each week. The standard response time is 10 seconds.

The analytic hierarchy process in section 2.3 is used to build the total quantitative evaluation model of e-learning user psychological experience, constructing user psychological experience quality from two features and indexes like usefulness and usability and then obtaining the total ranking weight vector $W = (0.1937, 0.0785, 0.4778, 0.0262, 0.1593, 0.0646)$ of six child features such as resource coverage, hit ratio of recommendation, user loyalty, response time, navigation definition and task-performing efficiency.

The focus of the experiment is to study and analyze child indexes like user loyalty, response rate, navigation definition and task-performing efficiency as well as the total index of the whole user psychological experience.
The average value is 31%, indicating that, as to learners, the attraction of this system is weaker and fewer users will continue using it for learning.

The response time ranges from 10 seconds to 70 seconds due to the influence of internet environment and the average response rate is 44%. The fact that the system response rate as to 25% users is 0% indicates that users close the system when waiting for the system response and as a result, no response time is obtained.

The average value of navigation is 11%. It shows that the value is 0% as to 51% users, namely 51% users can hardly find required knowledge items when using this system.

It shows that the average task-performing efficiency is 10% and the biggest value is 66%, indicating that some users can complete the learning task efficiently, but still 51% users can not complete the task.

_____

The highest value for experience quality is 62%, the lowest 27% and the average value 38%, indicating that the user psychological experience quality in this e-learning system remains to be further improved.

CONCLUSION

This paper has made a research and analysis on such aspects as feature selection, feature quantification measurement and the quantification evaluation modeling of total user psychological experience that influence e-learning user psychological experience quality. The research focus lies on features and factors of usefulness and usability. Further researches will be made on the influence of other factors, like emotions etc., on the user psychological experience quality, so as to build more integrated quantification evaluation approaches of user psychological experience.

REFERENCES

[1] Lin C; Hu J; Kong XZ, *Chinese Journa.*, **2012**, 35(1), 1-15
[2] Hu CP; Deng SL*, J. INTELL.*, **2006**, 24(3), 321-325
[3] L Dudek, *ACM Trans. Comput. Hum. Interact.*, **2006**, 18(1), 1-30
[4] Li XQ, *J. INTELL.*, **2010**, 28(5), 763-767
[5] Pang QH, *Computer Engineering*, **2007**, 33(18), 59-61
[6] Zhou RG, *Computer Engineering and Applications*, **2007**, 43(31), 102-105