Decision-oriented Usability Evaluation of an Operation Interface: Model and Application

LI Xiao-jun[1,2,3], *SUN Lin-yan*[1,2], *LI Min*[3]

[1] School of Management, Xi´an Jiaotong University, Xi´an 710049, P. R. China

[2] The State Key Lab for Manufacturing, Xi´an 710049, P. R. China

[3] The Second Artillery Engineering University, Xi´an 710025, P. R. China

Abstract: *To support multi-factor decision problems about usability evaluation, especially when studies fall short of comparable objects, a fuzzy synthetic evaluation model is explored in this paper. Grey relational analysis (GRA) is brought in the model to calculate weight vectors of the usability factors. And membership functions of a remark vector are constructed in the context of use of the operation interface. The present method is applied in usability evaluation of operation interface and is proved to be effective. The comprehensive usability gradation of the operation interface to good is* 0.616 4 *that meets the requirements in practice.*

Key words: usability evaluation; fuzzy synthetic evaluation method; grey relational analysis; membership function

1 Introduction

To improve the usability quality of an operation interface, the user-centered design (UCD) method has been universally applied in industry. One of the major challenges in applying this method concerns the models and process of usability evaluation. Evaluation is an integral part of any development process and sometimes discipline specific. In the human-computer interaction (HCI) field, Whitefield, Wilson and Dowell suggest that evaluation involves an assessment of the conformity between a system´s performance and its desired performance [1]. Usability evaluation in essence is a continual process of refining the design of the user interaction component based on frequent inputs from stakeholders and its results feed back into modifications to design [2,3]. In this paper, usability evaluation infers testing system performance against operational requirements at any stage of its development. Thus, heuristic evaluation, lab-based usability testing and other expert-based usability inspection methods are included in the scope of improving usability evaluation.

The usability evaluation of an operation interface is a rather large research area with many different goals, methods, and implications. These usability evaluation methods (UEMs) can be divided into three traditional categories [2]:

1) empirical methods that include controlled experiments, formal lab-based usability testing and field testing/operational evaluation.

2) expert-based usability inspection methods that include guideline reviews, heuristic evaluation, cognitive walkthroughs, formal usability inspections, usability walkthroughs and heuristic walkthroughs.

3) model-based approaches that include stages of user activity analysis and model-mismatch analysis. Practitioners usually use many of the methods through the various stages of design cycle as well as with different levels of formality.

As so far, practices in usability evaluation have the following characteristics:

1) Most studies focus on only one or two aspects of usability. According to the ISO 9241-11 (1998) standard, the components of usability involve effectiveness, efficiency and satisfaction [4]. However,

Received 13 January 2012

This paper is supported by the Ph. D. Programs Foundation of Ministry of Education of China under Grant No. 20070968063

Erik, Morten and Kasper reviewed 19 usability studies between years 1998-2000 and found that 11 experiments pay attention to only one or two factors of usability [5]. This phenomenon also exists in present practices.

2) Comparative evaluations among model, product and improved design account for most of the proportion of current studies. For each study, the result is described as this design/interface improves users' satisfaction or reduces the error rate of users' action. Therefore, the existence of comparative objects is necessary for this kind of research.

3) In previous studies of usability, subjective or objective data are collected and dealt with separately. However, it is controversial that whether subjective or objective measures are appropriate for a specific design [6]. Differences between subjective and objective data may lead to diverse conclusions and suggestions for improvement regarding the usability of an interface.

From the above reviews, it is found that current models and methods are ineffective to support multi-factor decision problems about usability evaluation, especially when the study falls short of a comparable design/interface. The purpose in this paper is to propose a method of usability evaluation based on fuzzy theory and grey related analysis method, which aims at decision support rather than problem inspection about usability of an operation interface.

2 Fuzzy synthetic usability evaluation model

Usability evaluation is usually conducted in a highly dynamic environment, involving complex tradeoffs and uncertainty. For instance, value of satisfaction that is one of the basic attributes of usability, is fuzzy mostly. And related factor weights are difficult to describe accurately. The uncertainty and fuzziness inherent in usability evaluation makes the use of a precise model problematic in practice. Under this condition, the fuzzy synthetic evaluation model and grey related analysis method are brought in usability evaluation of an operation interface.

2.1 Fuzzy multi-factor evaluation model

The fuzzy multi-factor evaluation model converts fuzzy values of evaluation indexes into quantities by constructing a fuzzy subclass, and then integrates these indexes through a fuzzy transform [7].

Giving two finite groups

$$\boldsymbol{U} = [u_1, u_2, \cdots, u_p]$$
$$\boldsymbol{V} = [v_1, v_2, \cdots, v_m]$$

$\boldsymbol{U}$ is the set compos of all the evaluation factors, $\boldsymbol{V}$ is the set composed of all the remark grades. r_{ij} is the judge result of evaluation factor u_i to remark v_j. So the decision-making matrix of p evaluation factors is as follows:

$$\boldsymbol{R} = \begin{bmatrix} R_1 \mid & u_1 \\ R_2 \mid & u_2 \\ \vdots & \\ R_p \mid & u_p \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pm} \end{bmatrix}_{p,m} \quad (1)$$

Where $\boldsymbol{R}$ is the fuzzy connection of $\boldsymbol{U}$ to $\boldsymbol{V}$, and R_i is the fuzzy connection of u_i to $\boldsymbol{V}$. If the weight of each evaluation factor is $\boldsymbol{A} = [a_1, a_2, \cdots, a_p]$ ($\boldsymbol{A}$ is in essence a fuzzy subclass of set $\boldsymbol{U}$, $0 \leqslant a_i \leqslant 1$, and the sum of a_i is 1). One fuzzy subclass of set $\boldsymbol{V}$ can be calculated by applying the synthetic operation of a fuzzy transform, which is the comprehensive evaluation result:

$$\boldsymbol{B} = \boldsymbol{A} \cdot \boldsymbol{R} = [b_1, b_2, \cdots, b_m] \quad (2)$$

Where $\boldsymbol{B}$ is a fuzzy set of $\boldsymbol{V}$. Fuzzy transform $\boldsymbol{A} \cdot \boldsymbol{R}$ changes into a common matrix calculation, which refers to many factors in all directions and is suitable for a multi-factors sequence. The calculation can be described as follows:

$$b_i = \sum_{i=1}^{p} (a_i \cdot r_{ij}) = \min\left(1, \sum_{i=1}^{p} a_i \cdot r_{ij}\right)$$
$$j = 1, 2, \cdots, m \quad (3)$$

2.2 Determination of evaluation factor weights

Since the weight of each factor is the keystone of the fuzzy multi-factor evaluation method, it is vital to correctly calculate the priority vector of usability factors, namely $\boldsymbol{A}$. But it is very difficult to achieve the perception of the priority vector directly in terms of absolute values.

The analytic hierarchy process (AHP) method, using the eigenvector approach to reconcile inconsistent subjective inputs, is popular in determination of evaluation weights [8~10]. And there is a precondition that each evaluation factor must be independent in the application of the AHP. However, most evaluation factors of usability are correlative. Grey relational analysis (GRA) is presented to determine the weights of usability factors.

GRA is a quantitative method to explore the similarity and dissimilarity among factors. And its core is that the closeness of a relationship is judged based on the similarity level of the geometric patterns of sequence curves [11,12]. GRA models are defined in three ways, with the distance of sequence, and with the difference or ratio of the corresponding slopes [13]. In our study, the general relational degree [11] method is used to calculate the weight of each factor. The steps involved in the grey relational analysis are as follows [14].

Step 1 Determination of reference sequence and comparative sequences

Assume $\boldsymbol{X}_0 = [x_0(1), x_0(2), \cdots, x_0(n)]$ is the reference sequence, which refers to the most important usability factor chosen by experts. And comparative sequences referring to other usability factors are

$$\begin{cases} \boldsymbol{X}_1 = [x_1(1), x_1(2), \cdots, x_1(n)] \\ \boldsymbol{X}_2 = [x_2(1), x_2(2), \cdots, x_2(n)] \\ \boldsymbol{X}_m = [x_m(1), x_m(2), \cdots, x_m(n)] \end{cases} \tag{4}$$

Where $m+1$ is the number of usability factors, and n is the number of tested objects similar to the operation interface in our study. $x_i(j)$ ($i=0,1,\cdots,m$; $j=1,2,\cdots,n$) denotes the evaluation value of the j-th interface of the i-th factor.

Step 2 Normalization of data sequences

The measures of different usability factors are various. Normalization of value sequences is necessary to make these factors comparable. The normalization model is as follows.

$$\begin{aligned} y_i(j) &= \frac{x_i(j)}{x_i(1)} \\ i &= 0,1,\cdots,m; \\ j &= 1,2,\cdots,n \end{aligned} \tag{5}$$

Where $y_i(j)$ is the j-th element in Y_i. Y_i is the normalization result of X_i.

Step 3 Calculation of general relational degree

$$\gamma(Y_0, Y_i) = \frac{1}{n}\sum_{k=1}^{n}\gamma(y_0(k), y_i(k)) \tag{6}$$

Where

$$\gamma(y_0(k), y_i(k)) = \frac{\min\limits_i \min\limits_k |y_0(k) - y_i(k)| + \rho \max\limits_i \max\limits_k |y_0(k) - y_i(k)|}{|y_0(k) - y_i(k)| + \rho \max\limits_i \max\limits_k |y_0(k) - y_i(k)|} \tag{7}$$

The distinguishing coefficient $\rho \in [0,1]$. In our study $\rho = 0.5$.

Step 4 Determination of evaluation factor weights

The weight of i-th factor is

$$\mu_i = \frac{\gamma(Y_0, Y_i)}{\sum\limits_{i=0}^{m}\gamma(Y_0, Y_i)} \tag{8}$$

Where $i = 0, 1, \cdots, m$, and

$$\gamma(Y_0, Y_0) = 1 \tag{9}$$

$$\sum_{i=0}^{m}\mu_i = 1 \tag{10}$$

3 Case study

Operation interface is a vital component of weapon systems. And its quality of usability affects operational effectiveness of the whole system. Our study on usability evaluation of operator interface aims to achieve the whole gradation of usability and support corresponding decisions.

3.1 Usability factors of operation interface

In the ISO 9241-11 (1998) standard, usability is considered as the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use [4]. In the process of usability evaluation, effectiveness, efficiency, and satisfaction are decomposed into factors. The four steps of decomposition involve definition of issue, definition of the context of use, decomposition of basic factors (namely effectiveness, efficiency, and satisfaction), and establishment of guidelines for interpretation of the data collected under it.

In our study, the operation interface and environment are described in detail combined with the specific model. And the intended users are young men who are 20 ~ 30 years old. Interaction task is given by a set of operation rules. Given the context of use, completeness of task, number of errors, time cost and perception of users' satisfaction are selected as evaluation factors. Measures of the above factors are percentage, number, minute and scale correspondingly. The scale of users' satisfaction in our study is depicted in Table 1.

Table 1 Scale of relative satisfaction

Intensity of satisfaction	Definition
1	Badly
3	Unsatisfied
5	Neutral
7	Satisfied
9	Strongly satisfied
2,4,6,8	Intermediate values between the two adjacent judgments

3.2 Calculation of usability factor weights

According to the past results of usability evaluation of four interfaces, original data sequences are as follows (Table 2). X_0 denotes value of perception of users' satisfaction, which is taken as the reference sequence. X_1, X_2, X_3 represents values of time cost, number of errors, completeness of task, which are considered as the comparative sequences.

Table 2 Scale of relative satisfaction

	First interface	Second interface	Third interface	Fourth interface
X_0	6	7	8	7
X_1	16	16	13	20
X_2	4	3	1	2
X_3	0.85	0.90	1.00	0.90

Normalization of these sequences is depicted in Table 3.

Table 3 Normalization of data sequences

	First interface	Second interface	Third interface	Fourth interface
$\boldsymbol{Y}_0$	1.00	1.17	1.33	1.17
$\boldsymbol{Y}_1$	1.00	1.00	0.81	1.25
$\boldsymbol{Y}_2$	1.00	0.75	0.25	0.50
$\boldsymbol{Y}_3$	1.00	1.06	1.18	1.06

In order to make cost-sorted factors (time cost, number of errors) and benefit-sorted factors (perception of users' satisfaction, completeness of task) comparable, $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are transformed in the following way.

$$y_i'(j) = \frac{1}{y_i(j)} \tag{11}$$

Where $i = 1, 2$; $j = 1, 2, 3, 4$. And the results are described in Table 4.

Table 4 Results of transformation

	First interface	Second interface	Third interface	Fourth interface
$\boldsymbol{Y}_0$	1.00	1.17	1.33	1.17
$\boldsymbol{Y}_1$	1.00	1.00	1.23	0.80
$\boldsymbol{Y}_2$	1.00	1.33	4.00	2.00
$\boldsymbol{Y}_3$	1.00	1.06	1.18	1.06

The general relational degree

$$\begin{cases} \gamma(\boldsymbol{Y}_0, \boldsymbol{Y}_1) = 0.90 \\ \gamma(\boldsymbol{Y}_0, \boldsymbol{Y}_2) = 0.71 \\ \gamma(\boldsymbol{Y}_0, \boldsymbol{Y}_3) = 0.94 \end{cases} \tag{12}$$

And then the weights vector of usability factors

$$\boldsymbol{A} = [0.28, 0.25, 0.20, 0.27] \tag{13}$$

3.3 Membership functions of fuzzy set V

The fuzzy set of gradation $\boldsymbol{V}$ = (poor, fair, good, excellent) is characterized by its membership function μ_V, which maps each element of the universe $\boldsymbol{X}$ to the interval [0, 1]. This function indicates the degree of belonging to $\boldsymbol{V}$ for each element of $\boldsymbol{X}$. The applicability of fuzzy technology depends on the ability to construct membership functions that appropriately represent various concepts in different contexts [15].

Membership functions can be constructed from data when it is available. In this approach, known as data-driven membership function estimation [16], the membership function that describes the underlying concept is fitted to the collected data points. When data are not available in the form of value-membership pairs, the conventional approach which is also adopted in our study, is to first pick the shape of the membership function from a list of families, and then to fine-tune the values of the parameters. In the context of use, ridge-shaped distribution is chosen as the reference function in our study.

Take number of errors for example, the membership function of poor is constructed in the following way. Its reference function is

$$\mu_{poor}(x) = \begin{cases} 0 & x \leqslant a_1 \\ \dfrac{1}{2} + \dfrac{1}{2}\sin\left[\dfrac{\pi}{a_2 - a_1}\left[x - \dfrac{a_1 + a_2}{2}\right]\right] & a_1 < x \leqslant a_2 \\ 1 & x > a_2 \end{cases} \quad (14)$$

And the corresponding shape is shown in Figure 1.

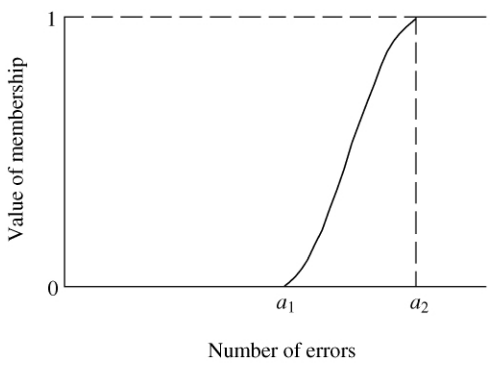

Figure 1 Shape of the reference function

In Figure 1, a_1, a_2 are the critical points. If number of errors is less than a_1, the interface is completely not poor from the point of view. And the interface is completely poor when number of errors is more than a_2. According to rules of operation, $a_1 = 2$, $a_2 = 5$.

Then the membership function of poor is

$$\mu_{poor}(x) = \begin{cases} 0 & x \leqslant 2 \\ \dfrac{1}{2} + \dfrac{1}{2}\sin\left[\dfrac{\pi}{3}\left(x - \dfrac{7}{2}\right)\right] & 2 < x \leqslant 5 \\ 1 & x > 5 \end{cases} \quad (15)$$

In our usability tests, the mean of number of errors is 2.5. And then $\mu_{poor}(x) = 0.067$. Namely, the degree of belonging to poor for number of errors is 0.067.

For number of errors, the membership functions of fair, good and excellent are constructed similarly.

$$\mu_{fair}(x) = \begin{cases} 0 & x \leqslant 2 \\ \dfrac{1}{2} + \dfrac{1}{2}\sin\left(\pi\left(x - \dfrac{9}{2}\right)\right) & 2 < x \leqslant 3 \\ 1 & 3 < x \leqslant 4 \\ \dfrac{1}{2} - \dfrac{1}{2}\sin\left(\pi\left(x - \dfrac{9}{2}\right)\right) & 4 < x \leqslant 5 \\ 0 & x > 5 \end{cases} \quad (16)$$

$$\mu_{good}(x) = \begin{cases} 0 & x \leqslant 1 \\ \dfrac{1}{2} + \dfrac{1}{2}\sin\left(\pi\left(x - \dfrac{7}{2}\right)\right) & 1 < x \leqslant 2 \\ 1 & 2 < x \leqslant 3 \\ \dfrac{1}{2} - \dfrac{1}{2}\sin\left(\pi\left(x - \dfrac{7}{2}\right)\right) & 3 < x \leqslant 4 \\ 0 & x > 4 \end{cases} \quad (17)$$

$$\mu_{excellent}(x) = \begin{cases} 1 & x \leqslant 1 \\ \dfrac{1}{2} - \dfrac{1}{2}\sin\left(\pi\left(x - \dfrac{3}{2}\right)\right) & 1 < x \leqslant 2 \\ 0 & x > 2 \end{cases} \quad (18)$$

And then the evaluation of number of errors to $\boldsymbol{V}$ is (0.067, 0.5, 1, 0). The result of normalization is (0.04, 0.32, 0.64, 0).

The decision-making matrix $\boldsymbol{R}$ of $\boldsymbol{U}$ to $\boldsymbol{V}$ can be obtained by integrating the results of other factors.

$$\boldsymbol{R} = \begin{bmatrix} 0 & 0.31 & 0.56 & 0.13 \\ 0.02 & 0.28 & 0.70 & 0 \\ 0.04 & 0.32 & 0.64 & 0 \\ 0.01 & 0.36 & 0.58 & 0.05 \end{bmatrix} \quad (19)$$

The comprehensive usability evaluation of the operation interface is

$$\boldsymbol{B} = \boldsymbol{A} \cdot \boldsymbol{R} = \begin{bmatrix} 0.28 \\ 0.25 \\ 0.20 \\ 0.27 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 0 & 0.31 & 0.56 & 0.13 \\ 0.02 & 0.28 & 0.70 & 0 \\ 0.04 & 0.32 & 0.64 & 0 \\ 0.01 & 0.36 & 0.58 & 0.05 \end{bmatrix} =$$

$$[0.0157, 0.318, 0.6164, 0.0499] \quad (20)$$

4 Conclusions

It is a benefit to evaluate the whole usability degree for improvement of operation interface. Combined with the GRA model, the fuzzy synthetic evaluation method is effective on decision support related with usability in practice. In our study, the comprehensive gradation of usability to good is 0.616 4. The whole usability degree of the operation interface meets expected demand. However, it is obvious that improvement of the design is necessary to enhance usability. The study has not taken account of learnability because participant users have operation related experience. However, the factor of learnability is crucial for usability evaluation especially where the interface is intensively used and the users should be able to learn quickly. To gain a more complete picture of usability, other factors should be paid attention to and introduced in evaluation models.

References

[1] Whitefield A, Wilson F, Dowell J, *A framework for human factors evaluation.* Behavior and Information Technology, 10(1): 65 ~ 79, 1991

[2] Terence S, *Andre, determining the effectiveness of the usability problem inspectors: a theory-based model and tool for finding usability problems.* Dissertation for Doctor Degree in Virginia Polytechnic Institute and State University, 2000

[3] Dix A, Abowd G, Beale R. *Human-computer Interaction.* Prentice Hall, Nj, 1993

[4] International Organization for Standardization. Iso 9241-11: Ergonomic Requirements for Office Work with Visual Display Terminals (Vdts), Part Ⅱ: Guidance on Usability, 1998

[5] Erik Frøkjær, Morten Hertzum, Kasper Hornbek, *Measuring usability: effectiveness, efficiency, and satisfaction are really correlated.* Proceedings of the Acm Chi 2000 Conference on Human Factors in Computing Systems, 345 ~ 352, 2000

[6] Hornbek K, *Current practice in measuring usability: challenges to usability studies and research.* International Journal Human-computer Studies, (64): 79 ~ 102, 2006

[7] Wang X Z, *Information process in fuzzy space*, Wuhan University Press, Wuhan, 130 ~ 131, 2003 (In Chinese)

[8] Xu W X, Zhang Q S, *An algorithm of meta synthesis based on the grey theory and fuzzy mathematics.* Systems Engineering-Theory and Practice, 21(4): 114 ~ 119, 2001 (In Chinese)

[9] Chen D L, Li Z C, *The appraisement index in suring of information system investment and method for grey comprehensive appraisement.* Systems Engineering-Theory and Practice, 22(2): 100 ~ 103, 2002 (In Chinese)

[10] Guo P, Shi P G, *Research on fuzzy grey comprehensive evaluation method of project risk.* Journal of Xi´an University of Technology, 21(1): 106 ~ 109, 2005 (In Chinese)

[11] Deng J L, *Spread of grey relational space.* Journal of Grey Systems, 7(3): 96 ~ 100,

1995 (In Chinese)

[12] Deng J L, *Grey group decision in grey rationale space*. Journal of Grey System, 10(3) : 177 ~ 182, 1998 (In Chinese)

[13] Xie N N, Liu S F, *Research on evaluation of several grey relational models adapt to grey relational axioms*. Journal of Systems Engineering and Electronics, 20(2) : 304 ~ 309, 2009 (In Chinese)

[14] Gu H, Song B F, *Study on effectiveness evaluation of weapon systems based on grey relational analysis and topsis*. Journal of Systems Engineering and Electronics, 20(1) : 106 ~ 111, 2009 (In Chinese)

[15] Klir G J, Yuan B, *Fuzzy sets and fuzzy logic: theory and applications*. Prentice-Hall, Upper Saddle River, Nj, 1995

[16] Andres L, Medagklia, Fang S C, Henry L W, Nuttle, James R. Wilson, *An efficient and flexible mechanism for constructing membership functions*. European Journal of Operational Research, (139) : 84 ~ 95, 2002

Brief Biographies

LI Xiao-jun is a Ph. D in the School of Management of Xi´an Jiaotong University. His research interests include usability, system science and system engineering. xi_anlxj @ 126. com

SUN Lin-yan is a Ph. D in the School of Management of Xi´an Jiaotong University, and a professor. His research interests include human factors engineering and advanced manufacture management.

LI Min is a Ph. D in the Second Artillery Engineering University, and a professor. His research interests include missile weapon systems and engineering.