IASDR07
INTERNATIONAL ASSOCIATION OF SOCIETIES OF DESIGN RESEARCH THE HONG KONG POLYTECHNIC UNIVERSITY 12TH TO 15TH NOVEMBER 2007

A PROPOSAL OF USABILITY EVALUATION METHOD BY ROUGH SET THEORY

Katsuo Inoue[1], Masayuki Sakai[2] and Yusuke Kinoshita[3]

[1]Department of Kansei Design, Hiroshima International University, k-inoue@he.hirokoku-u.ac.jp

[2] School of Design, Sapporo City University, m.sakai@scu.ac.jp

[3] Graduate School, Kyushu Institute of Technology, kinoshita-yusuke@edu.life.kyutech.ac.jp

ABSTRACT:

We herein propose the usability method as a means of extracting the advantages and disadvantages of an interface design. The proposed method is characterized by the combination of the rough set theory, which is a type of quantitative method, and the conventional qualitative method. In addition, we considered and verified the usefulness of the proposed method through a case study involving a digital music player. In the case study, we first calculated the evaluation ranking of five digital music players by the AHP based on rough approximations, based on which a non-transitive relation was calculated. Samples of five digital music players were divided into two groups of high rank and low rank based on the results of evaluation. We then conducted a task analysis and a protocol analysis. As a result, we extracted the advantages and

disadvantages of the digital music players reported for each group. Furthermore, we calculated rough sets using the table of cognitive parts of the Graphical User Interfaces of the digital music players and the decision class of two groups in order to determine whether the GUI is ease to use. Finally, combinations of cognitive advantages and disadvantages were extracted from the analysis results.

1. BACKGROUND AND PURPOSE

The methods of usability evaluation can be classified into two main categories: quantitative methods and qualitative methods. Quantitative methods are used primarily in selecting the best prototype from a number of prototypes. In contrast, qualitative methods are generally used to investigate problems related to interface design. However, since qualitative data is not always objective, methods of usability evaluation that combine qualitative and quantitative methods are expected. Qualitative methods improve the user interface by identifying problems. However, methods that can determine both the advantages and disadvantages of a user interface design are desired.

The present paper proposes a new usability evaluation method by which both the advantages and disadvantages of a user interface design can be determined. The proposed method is characterized by combining the rough set theory (Mori, Tanaka and Inoue 2004), which is a type of quantitative method, and the conventional qualitative method. We considered and verified the usefulness of the proposed method through a case study involving a digital music player.

2. EXPERIMENT

In an experiment using the KJ method, we first classified twelve digital music players into five samples (A, P, H, S, and V), each of which had a typical interface. The subjects were ten students (five male and five female). Each subject was asked to evaluate five samples with respect to visual ease of use by means of a five-grade paired comparison. The subjects were not allowed to touch the samples.

We then conducted a usability evaluation of task analysis with the same subjects. Specifically, we asked the subjects to evaluate each item of the three tasks shown in Table 1 for five real samples. We recorded the experiment by video camera in order to later analyze protocol data that were specific to their comments on advantages and disadvantages of these samples. After these tasks were completed, each subject evaluated 31 items in the three tasks (Task 1: fifteen items, Task 2: two items, and Task 3: fourteen items) into five ranks in terms of ease of use.

Finally, each subject evaluated the synthetic ease of use into five ranks by paired comparison. In order to confirm the statistical validity, we conducted an additional experiment in which twenty students (ten males and ten females) evaluated the visual ease of use into five ranks by paired comparison.

Table 1: Contents of Task Analysis.

■**Task1** （**Beginner-class level**）

Turn on the power supply. Confirm that the battery is in place and listen to Song B by Singer A. Adjust to the volume to be suitable, and fast forward to approximately 20 seconds of 3 minutes. Stop musical playback after confirm main part, and return to the main screen.

■**Task2** （**Middle-class level**）

Set the player to play one of your favorite songs repeatedly, and then stop playback and return to the main screen.

■**Task3** （**Upper-class level**）

Set the player to play songs randomly. Fasting forward to find your favorite song and listen to this song. After listening briefly to the song, turn off the power and return the player to its original position.

3. ANALYSIS METHOD AND CONSIDERATIONS

3. 1. CONSIDERATION OF PAIRED COMPARISON EVALUATION

We analyzed both the evaluation of paired comparison of visual ease of use before the experiment and the synthetic usability evaluation conducted after the experiment by the AHP method based on rough approximations (Tanaka and Sugihara 2001), and the obtained results are shown in Fig. 1. Specifically, we applied the AHP method using the geometric averages for thirty subjects and ten subjects, respectively.

In addition, the non-transitive relation “if A>B and B>C then A<C” was not calculated by the conventional AHP method. However, the non-transitive relation could be calculated by the AHP method based on rough approximations proposed by Prof. Tanaka using the linear programming problem (Tanaka, Entani and Sugihara 2005). Therefore, we adopted this method in order to take advantage of the higher accuracy offered by this method, as compared to the conventional method.

The transitive relation, excluding sample H, is shown in the evaluation of paired comparison of visual ease of use before the experiment, as illustrated in Fig. 1. Moreover, the width of the interval of each sample is approximately the same. In particular, since the interval between the first and second samples does not overlap, each ranking has been decided. However, since the samples after the third sample overlap, each ranking has not been decided clearly.

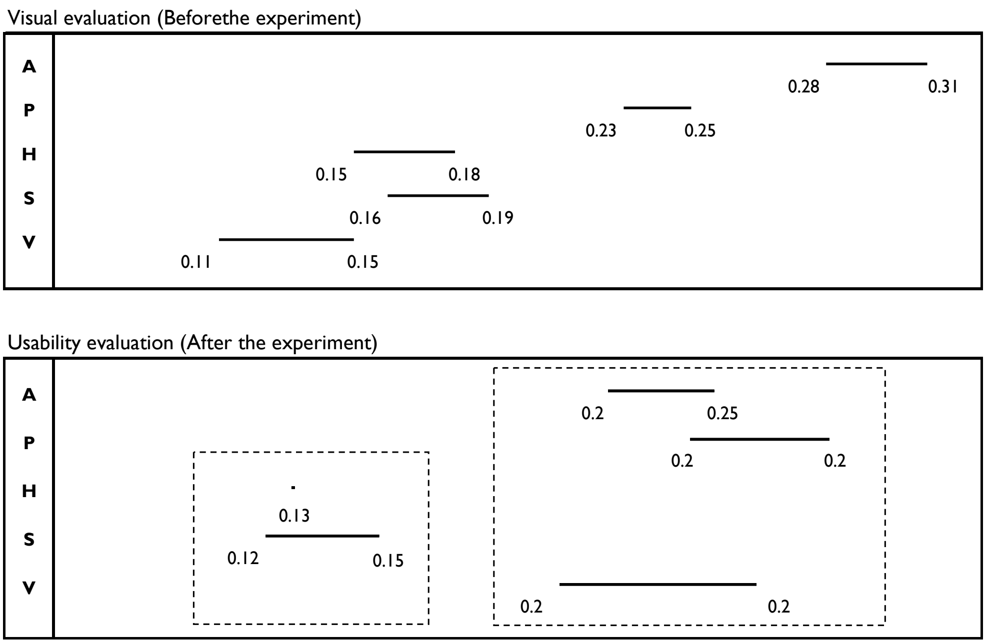

Fig. 1: Result of AHP based on Rough Approximations.

On the other hand, in the synthetic usability evaluation after the experiment, the lowest sample, Sample V, shifts to a higher rank, and the interval is the largest, which shows that the evaluation of Sample V by ten subjects is dispersive. Moreover, Sample H, which was ranked in the middle before the experiment, was ranked lower after the experiment, and the evaluation was barely dispersive. Next, the rankings of Samples A and P were reversed after the experiment, as compared with the rankings before the experiment. The interval between Samples A and P also becomes wider. In the following chapters, we present a discussion and analysis of these factors.

3. 2. DISCUSSION OF TASK ANALYSIS

After each subject evaluated 31 items for the three tasks shown in Table 2, we calculated the average value of each item for the ten subjects and rearranged the items in ascending order with respect to each sample, as shown in Fig. 2. The value “1 ~ 3” in the figure indicates the number of tasks. We also added a contour line at the average position (a) “2.5”, “3.0”, and “4.0”.

In addition, except for Sample A, the averages were omitted. The values (b) shown at the bottom of Fig. 2 are the averages of the individual average ranks (a) for the 31 items. Each task corresponds to Beginner-class level, Middle-class level, or Upper-class level at the top of Table 1. In other words, each task corresponds with degree of ease to use.

	A	Average (a)		V		P		S		H	
	22)	1.7	**3**	1)	1	29)	**3**	22)	**3**	32)	**3**
2.5	5)	2.5	1	16)	2	1)	1	16)	2	4)	1
	32)	2.9	**3**	22)	**3**	22)	**3**	23)	**3**	3)	1
3	16)	3.1	2	19)	**3**	25)	**3**	29)	**3**	21)	**3**
	15)	3.2	1	24)	**3**	23)	**3**	24)	**3**	15)	1
	7)	3.3	1	2)	1	15)	1	30)	**3**	24)	**3**
	19)	3.3	**3**	11)	1	24)	**3**	1)	1	25)	**3**
	1)	3.5	1	8)	1	27)	**3**	3)	1	12)	1
	10)	3.5	1	5)	1	30)	**3**	15)	1	17)	2
	13)	3.6	1	10)	1	10)	1	25)	**3**	26)	**3**
	23)	3.6	**3**	13)	1	26)	**3**	32)	**3**	7)	1
	6)	3.7	1	17)	2	32)	**3**	4)	1	29)	**3**
	29)	3.7	**3**	14)	1	16)	2	26)	**3**	27)	**3**
	18)	3.8	**3**	18)	**3**	28)	**3**	10)	1	30)	**3**
	24)	3.8	**3**	23)	**3**	21)	**3**	27)	**3**	16)	2
	4)	3.9	1	26)	**3**	6)	1	28)	**3**	10)	1
	14)	3.9	1	27)	**3**	4)	1	18)	**3**	20)	**3**
4	3)	4	1	32)	**3**	2)	1	2)	1	8)	1
	17)	4	2	9)	1	3)	1	20)	**3**	14)	1
	28)	4	**3**	15)	1	11)	1	31)	**3**	1)	1
	30)	4	**3**	20)	**3**	17)	2	8)	1	5)	1
	2)	4.1	1	12)	1	18)	**3**	19)	**3**	6)	1
	9)	4.1	1	28)	**3**	19)	**3**	17)	2	19)	**3**
	11)	4.1	1	25)	**3**	20)	**3**	9)	1	22)	**3**
	26)	4.2	**3**	30)	**3**	31)	**3**	13)	1	2)	1
	12)	4.3	1	31)	**3**	7)	1	14)	1	18)	**3**
	20)	4.3	**3**	21)	**3**	12)	1	12)	1	28)	**3**
	27)	4.3	**3**	3)	1	9)	1	21)	**3**	31)	**3**
	21)	4.4	**3**	6)	1	13)	1	7)	1	13)	1
	25)	4.4	**3**	7)	1	8)	1	11)	1	11)	1
	8)	4.5	1	29)	**3**	14)	1	6)	1	23)	**3**
	31)	4.6	**3**	4)	1	5)	1	5)	1	9)	1
	Average (b)	3.76		3.73		3.59		3.2		2.73	

Fig. 2: Ranking of Evaluation Criteria of Task Analysis.

From the results of Fig. 2, the ranks for Task 3, which is the most difficult, are higher than those of the other tasks for Samples A and V. In contrast, the ranks for Task 3 are low and those for Task 1, which is the easiest task, are higher for Sample P. For Samples H and S, the dispersion of the evaluation scores is too wide to recognize a clear tendency. In other words, the interface design concepts of Samples H and S are less clear than the other samples.

Furthermore, the average ranks (b) for each of the 31 items for Samples H and S are much lower than those for the other samples. As indicated by the third boundary line from the left-hand side of Fig. 2, for example, Sample H has no item with an average rank of more than “4.0”, and the average of “2.73” for all of the items for Sample H is the lowest among the samples. As a result, we considered Sample P to be ranked highly after the experiment because it was designed based on the basic concept of usability. In contrast, Samples A and V were ranked highly because for these samples the difficult tasks were comparatively easy to understand.

3. 3. DISCUSSION OF PROTOCOL ANALYSIS

We extracted the advantages and disadvantages reported by subjects in the recorded video of the experiment. The contents were arranged as shown in Table 2. We marked a sample corresponding to an item with “*” and a sample only corresponding to the item with “***” especially. In other words, the “***” symbols indicated as advantages are selling points, and the “***” symbols indicated as disadvantages denote poorly evaluated items.

First of all, Samples A and V have more selling points because there are a number of advantages and fewer disadvantages compared to the other samples, as suggested by the high evaluation after the experiment, as shown in Fig. 1. In contrast, Sample H did not have any advantages, but had several disadvantages, as indicated by the low evaluation in Fig. 1. Despite having several advantages, Sample S had the greatest number of disadvantages. This result suggested the low evaluation and wide dispersion after the experiment, as shown in Fig. 1.

Finally, since Sample P had no “***” symbols by evaluation items and its some marks in disadvantages were common in the other samples, its some faults was not conspicuous. Therefore, Sample P is considered to be evaluated highly.

Table 2: Advantages and Disadvantages.

Advantages		A	P	H	S	V
Screen	Large screen size	*				*
	Large volume of information of the screen	*				*
	Color liquid crystal	*			*	*
Display	GUI					***
	Icon is comprehensible	*	*		*	*
Button	Buttons are easy to use because they are few	*	*			
	Large decision button					***
Power supply	Power supply is comprehensible	***				
	Playback the tune at the same time as turning on the power supply				***	
Function	Two or more kinds of song can be shuffled.	*			*	
	Use a shortcut during playback.				***	
Others	Easy to use after become accustomed to the unit	***				
	Easy to select songs	*	*			*
	Can be used without reading the manual					***
	Usage is similar to that of a cell phone					***

Disadvantages		A	P	H	S	V
Screen	Small screen		*	*	*	
	Small volume of information on the screen		*	*	*	
	Monochrome screen		*	*		
Display	Characters are small			*	*	
	Character is English				***	
Button	Method of turning the power on is not clear		*			*
	Method of turning the power off is not clear	*	*		*	*
	Location of the selection button is not clear				***	
	The rotation button is difficult to use	***				
	The button is of the slide type			***		
	No consistency in button use			*	*	
	Buttons are difficult to use			***		
	No feedback for button use		*	*		
Menu	Location of menu panel is not clear				***	
	Method of going menu panel is not clear				***	
Function	Method of searching selection is not clear				***	
	Precision fast forwarding is difficult				***	
	Only one song can be selected at a time		*		*	
Others	Location of operating now is not clear			*	*	
	The screen structure is not clear			*	*	
	The power supply shuts down quickly				***	

4. ANALYSIS AND CONSIDERATION OF DESIGN FACTORS

4. 1. CONSIDERATION OF VISUAL EASE TO USE

We analyze the design factor of visual ease of use for each sample shown in Fig. 1 by rough sets. First, in preparation for calculating rough sets, we extracted the cognitive parts listed in the left-hand side of Table 3 by referring to the protocol (utterance) data in the usability evaluation experiment.

In addition, we created a decision table of the decision class, in which five samples are arranged in order of evaluation (sample A = 1, sample P = 2, sample S = 3, sample H = 4, sample V = 5).

Table 3: Two Types of Cognitive Parts.

Visual evaluation (Before the experiment)

Whole shape	Main body shape	Portrait square	A1
		Oblong square	A2
		Barrel type	A3
	Size	Standard	B1
		Small	B2
	Thickness	Thick	C1
		Standard	C2
		Thin	C3
Screen	Screen size	Standard	D1
		Small	D2
	Direction of screen	Length	E1
		Side	E2
	Screen color	Color	F1
		Black and white	F2
Main button	Main button position	Center	G1
		Right side	G2
	Main button shape	Round	H1
		Square	H2
		Special	H3
	Operation of main button	Push-button	I1
		Push-button + rotation	I2
		Slide	I3
		Stick	I4
Button	Number of buttons	Many	J1
		Few	J2
	Size of decision button	Standard	K1
		Small	K2
	Font size of button	Standard	L1
		Small	L2

(Table above: Visual ease of use)

Usability evaluation (After the experiment)

Character	Character representation on menu	GUI + character	A1
		Only character	A2
	Size of GUI	Large	B1
		Small	B2
		None	B3
	Character mark	Japanese	C1
		English	C2
	Font size on GUI	Large	D1
		Standard	D2
Volume of information	Volume of information on display	Large	E1
		Normal	E2
		Small	E3
	Volume of information when selecting	One line	F1
		Two lines	F2
		Four lines	F3
		Six lines	F4
Operation sound	Operation electronic sound at each operation	Yes	G1
		No	G2
	Operation sound of button	Yes	H1
		None	H2
Function	Automatic playback	Yes	I1
		No	I2
	Short cut	Yes	J1
		No	J2
	Number of search functions	Many	K1
		Few	K2
Others	Scroll bar	Yes	L1
		No	L2
	Time to turn off power automatically	Long	M1
		Short	M2
	"Hold" display on screen	Yes (all buttons)	N1
		Yes(some buttons)	N2
		No	N3

After the preparation, the decision rules of each sample, which are shown in the left-hand side of Fig. 3, were obtained by calculation of rough sets. In addition, since the CI value of each decision rule of each sample is "1", the CI value has been omitted from the left-hand side of Fig. 3. The CI value is an index that shows the degree of contribution to the decision class. The decision rules shown in the left-hand side of Fig. 3 can be used to determine the cognitive parts of the core. We therefore surrounded these parts by square frames. For example, the core of Sample A consists of “C3” and “J2”.

We used the decision rule analysis method to extract this important core. The result calculated using this analysis method for each sample is on the right-hand side of Fig. 4. The right-hand side of Fig. 3 shows that the CI value of “C3” as a core is high in the standardization column score of Sample A.

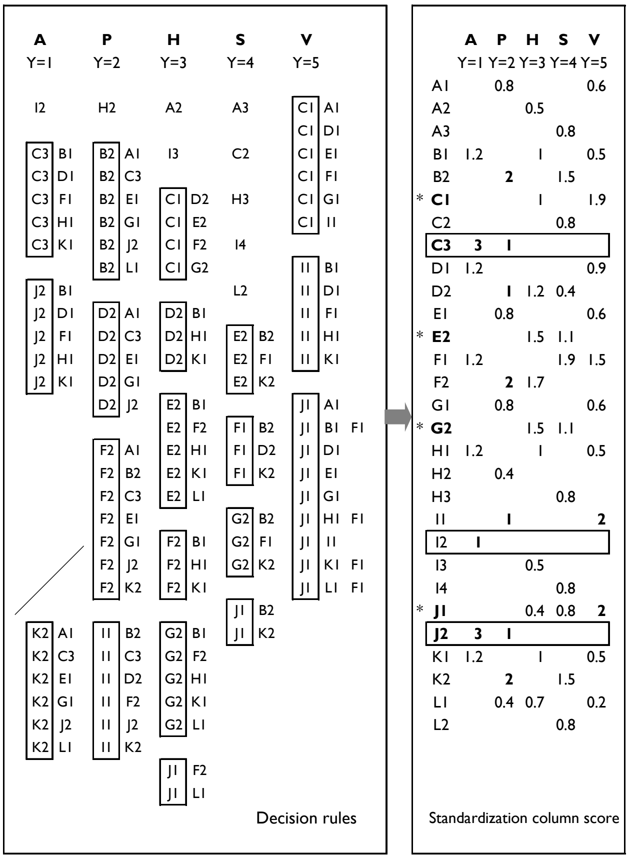

Fig. 3: Result of Visual Ease of Use.

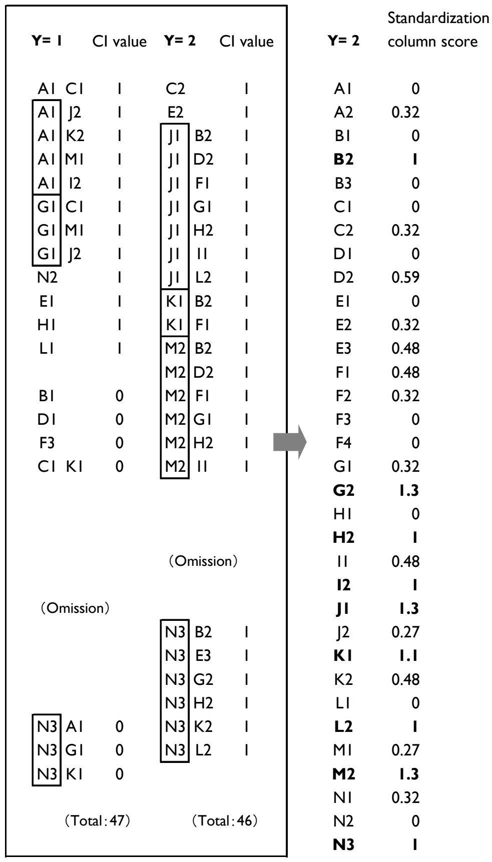

Fig. 4: Result of Display Screen Design.

First, we extracted the cognitive core parts of Samples A and P having high evaluation rankings in Fig. 1, so that the cognitive core part consisted of “C3 (thin whole shape)”, “I2 (push button + rotation)”, and “J2 (few buttons)”. These three cognitive core parts were the only feature for these two samples. Therefore, these parts were considered to be features related to high ranking.

On the other hand, we considered the cognitive core part of samples with low ranking from the same view. “J1 (many buttons)” was a characteristic cognitive core part in Samples H, S, and V because it was not found in samples with high ranking. In particular, since Sample V has the highest value of “J1”, this feature worsens the first impression, and we considered that this worse impression made ranking the lowest.

Next, although, unlike “J1”, “C1 (thick whole shape)”, “E2 (width type screen)”, and “G2 (main button located on the right-hand side)” were not cognitive core parts common to all three samples, these parts were considered to facilitate the evaluation of three samples. Since “G2” of Samples H and S have higher values, we considered that “G” is the factor which lowered the evaluation. Moreover, since the value of “C1” of Sample V is higher, “J1” and “C1” are considered to have made the evaluation the lowest.

Furthermore, “A3 (body form with a width type and slack type)”, “C2 (standard whole-shaped thickness)”, “H3 (special main button form)”, “I4 (main button with stick type)”, and “L2 (small character size)” were features unique to Sample S. From the lower ranking of Sample S, these cognitive core parts did not contribute to the visual ease of use.

Figure 3 shows that the number of decision rules in Sample A was extremely low. The small number of features indicates that the features of Sample A are clear. Sample A was found to be different form other samples, which may explain the why the subjects felt that the visual ease of use of Sample A was best from this point.

4. 2. CONSIDERATION OF EASE OF USE OF THE DISPLAY SCREEN

Next, we analyzed and considered the display screen after the experiment. We extracted cognitive parts that contributed to ease of use of the display screen from the video record. The usability of the cognitive part is shown on the right-hand side of Table 3. We separated the samples into highly ranked samples (Samples P, V, and A) and poorly ranked samples (Samples S and H) at the bottom of Fig. 1 and created a decision table with the decision class (Y) of these two groups. The results calculated from the rough sets in the table are shown in Fig. 4.

The decision rule for which the CI value of the high evaluation group (Y = 1) was a maximum (CI = 1) is indicated by the combination of “A1 (GUI (Graphical User Interface) + character)” and “C1 (Japanese character)” in Fig. 4. In decision rules in which the CI value was “0.67”, “A1” was a core of the combination, as illustrated by the square frames around the values.

Similarly, the cognitive part of “G1 (with operation electronic sound)” also serves as a core. In the independent cognitive part, “N2 (“hold” display shown on the screen)”, “E1 (large amount of information on the display)", "H1 (operation sound of a button)", and "L1 (with a scroll bar)" were calculated. The results indicate that the display screen design with a large amount of information and a GUI with both Japanese characters and various kinds of electronic sounds and scroll bars were easy to use.

On the other hand, for the group (Y = 2) of poorly evaluated samples in Fig. 4, we could not focus on high CI values because all of the CI values were the same. Therefore, we examined the group using a decision rule analysis method. The analysis results are shown on the right-hand side of Fig. 4. The cognitive core parts with values greater than "0.96" for the standardization column score, which is bordered by the square frame on the left-hand side of Fig. 4, were "B2, G2, H2, I2, J1, K1, L2, M2, and N3".

These results indicate that the display screen design both with a small GUI and without sounds and other functions was difficult to use. In addition, the independent cognitive part of "A2, C2, E2, F2, and N1" indicates that the display screen design with a menu using English characters and in which information presented at song selection required two lines was difficult to use.

5. CONCLUSIONS

The proposed method differs from the conventional usability evaluation method in that we can extract both the advantages and disadvantages of a user interface design based on the valuation basis of ranking. Although the causal relationship (inverse problem) among a few samples, as in the case study of the present study, was not able to be determined by the conventional mathematical technique, the types of contents of the operation that contributed to ease of use were clarified through factor analysis by rough sets.

REFERENCES:

Mori, Tanaka and Inoue (2004) Rough Sets and Kansei Engineering, Kaibundou Press, Japan.

Tanaka and Sugihara (2001) Dual Mathematical Models based on Rough Approximations, Japan Society for Fuzzy Theory Vol.13, No.6, Japan, 592-599.

Tanaka, Entani and Sugihara (2005) Approach to Interval Evaluation in Decision Making, Japan Society for Fuzzy Theory and Intelligent Informatics Vol.17, No.4, Japan, 406-412.