Quantifying User Interface Usability

Park, Kyung S. and Lim, Chee H.
Department of Industrial Engineering, KAIST, Taejon, Korea

ABSTRACT

The importance of usability evaluation is increasing in developing a new system and product. The current approaches for usability evaluation are: the comparative evaluation to measure usability, and the iterative user interface design to find usability problems. This paper presents three types of characteristics and a set of criteria for usability evaluation. The methodology for criteria-based quantitative analysis of user interface usability is investigated with a view to measuring usability. The fuzzy weighted-checklist method with linguistic variables is used for quantitative analysis. This analysis provides a quantitative measure, which reflects the degree of excellence of user interface usability during the design and development phases.

Introduction

The objective of this paper is to present the following aspects of usability engineering, to aid in the development of new procedures and measures for user interface usability:

(1) a set of criteria that the product should satisfy some features and attributes required to make a product usable; (2) a quantitative analysis, by the assignment of characteristics and criteria that specify a figure of merit or degree of excellence for product usability.

Usability Characteristics and Criteria

The important characteristics for usability are classified into three parts of user interface: physical, cognitive and affective characteristic[6]. Physical characteristic is a necessary condition for usability that includes such physiological and formal aspects as 'see' and 'grasp'. Cognitive characteristic includes such cognitive features as 'know' and 'easy to learn'. Moreover, affective characteristic includes such human feeling as 'comfortable' and 'pleasant'. The three characteristics for usability are evaluated by using the criteria, which are discussed in the following.

A criterion for usability is a measure of the extent to which a certain characteristic—quality, property or attribute—is exhibited or possessed by a computer product. Thus, the product should satisfy some criteria of usability to be secure usability of interactive computer product. To establish a set of major criteria for 'usable' interfaces, we reviewed a number of the literature in this area, and selected several representatives of them.

High level principles for user interface design are contained in ISO 9241-10 "Dialogue principles"[2]. These principles that should be applied to the design of dialogues between humans and information systems are listed in Table 1. For example, the factor 'self-descriptiveness' could be thought of as being composed of aspects concerned with self-documenting, transparency and decision supporting.

Dzida et al.[3] had presented an analytical approach for identifying essential properties of interactive systems as perceived by their users. In addition, they were reported that user-perceived quality was defined as a multidimensional property of interactive systems comprising at least seven aspects.

Table 1.

To measure user satisfaction, and hence assess user perceived software quality, meanwhile, Kirakowski et al.[5] had developed the Software Usability Measurement Inventory(SUMI) as part of the Metrics for Usability Standards in Computing(MUSiC) Project. SUMI decomposes usability into five different components: efficiency, effect, helpfulness, control and learnability. SUMI can benefit software designers and people involved in the purchasing of software, by indicating how easily the software can be used.

Ravden and Johnson[10] had presented a set of software ergonomics criteria, or 'goals', which a well-designed user interface should aim to meet. The nine criteria are listed in table 1. As the usability factors summarized in Table 1 were compared with each other, we found the common grounds and differences of usability factors. Ravden and Johnson's criteria had more subdivide than the pervious other factors, and could become the appropriate template for our usability criteria.

Specifically, visual clarity among other criteria could be more important factor as the graphic user interface and multimedia technology are introduced. Moreover, consistency is one of the most important factors that are considered in the design of user interface[4]. However, Ravden and Johnson excluded the learnability factor from their criteria set.

The effect factor in SUMI refers to the user feeling good, warm, happy or the opposite as a result of interacting with the product. Therefore, the effect factor can be seen as an element within the affective characteristic of usability. The affective characteristic could be evaluated by the criteria satisfaction, reliability, safety, originality and pleasantness. These are derived from the adjective words of Image Technology(or Kansei Engineering)[7].

A usability checklist is a questionnaire that contains a set of questions. A set of exhaustive checklists is considered for each of the criteria, to evaluate all aspects of usability measured by that criterion. Whenever a checklist question needs to be considered by assessing a number of points, a subchecklist is prepared, which consists of a number of questions. These checklists are mainly meant for anomaly detection.

Qualitative analysis

The current approaches for usability evaluation are: the comparative evaluation to measure usability, and the iterative user interface design to find usability problems[8]. When designers must choose between two or more interface alternatives, the comparative evaluation to measure which alternative is the most usable can be performed. However, the iterative design aims specifically at refinement based on lessons learned from previous iterations. Therefore, the comparative evaluation is usually viewed as constituting a methodology different from iterative design. In this study, we investigate the quantitative analysis to measure usability.

Quantitative analysis to measure user interface usability

Ravden and Johnson regarded the quantitative analysis as an unrealistic goal, because their method aimed at identify the strengths and weakness of an interface[10]. However, the quantifiable outcomes, which allow some comparison between interfaces, could help the decision making during the design and development phases. For example, interface designers typically can choose between several alternative solutions for various design issues and want to know how each alternative affects the resulting system's overall usability.

The method for quantitative analysis

The method used for quantitative analysis is the fuzzy-weighted checklist method proposed by Park and Kim[9]. The method is a fuzzy version of the weighted checklist technique used for evaluating or comparing complex systems(or subjects). The procedure of this method consists of converting the ratios of weights obtained by pairwise comparison to a best combination of normalized weights and obtaining the fuzzy composite score and its linguistic approximation. The method can be of service to design engineers and system analysts for system evaluation with qualitative criteria.

To apply this method to the present application, a hierarchy of four levels has been defined as shown in Fig.1. Each hierarchical level is assigned a weight relative to its importance.

Figure 1.

The highest level is broken down into lower levels of 100% weight. These levels are again split into lower levels of 100% weight and so on, until the lowest level is reached. The highest level in this application is user interface usability characteristics, the next lower level is usability criteria, and the lowest two levels are usability checklists and subchecklists.

To perform reliable usability evaluation of interactive system, several people are usually involved. When multiple evaluators are involved, it is necessary that their judgments are synthesized into a single judgment. Aczel and Saaty[1] proved that synthesizing the individual ratios of weights was effectively performed by the geometric mean. Therefore, the geometric means of the individual ratios of weights are converted to a best combination of normalized weights.

These normalized weights are assigned on the principle of relative importance of the constituents of a level at that level and to that at the higher level. To assign weights to characteristics and criteria the same principle is used. The analysis carried out is a total static analysis. The system designers or users evaluate the lowest two levels, i.e checklists and subchecklists, according to the system, i.e the user interface type. The evaluators grade the system on a rating scale.

Rating scale

The scale used to rate the checklists is a 7-point scale(0 to 6) using 14 fuzzy sets of linguistic values(verygood, good, fair, medium, poor, bad, verybad, unknown). For fuzzier rating, hyphenated ratings such as fair-good can be used. Rating of the checklist is the most important part of the analysis. An individual has to evaluate the question and assign a linguistic value to it.

Evaluation Procedure

The evaluation starts from the lowest level, i.e., checklists or subchecklists. That level is evaluated and the linguistic value for rating is assigned. The assigned values are then multiplied by the normalized weights. It is obtained the composite score by summing these results.

The fuzzy composite score at that level is carried forward to the next higher level. Again, computing the weighted average is performed to obtain the composite score for that level, and this process continues until the highest level is reached.

Because the composite score from the weighted average is a fuzzy set, the median of composite score is used for transforming such a fuzzy set into a representative numerical score[9]. The final composite score(and its linguistic approximation) gives the figure of merit or degree of excellence of user interface usability during the design and development phases.

Quantifying usability improvements in iterative design

A uniform scoring of usability improvements could be use to quantify relative changes in usability and to compare the different iterative design processes. The approach normalizes all usability measures concerning the values measured for the initial design. Thus, the initial design has a normalized usability value of 100, and the approach normalizes the subsequent iterations by dividing their measured values by those measured for the initial design.

Converting measures of user interface usability into improvement scores is more challenging, because the approach used in this paper is measured with questionnaires and rating scales. The underlying problem is that rating scales are not ratio scales. It is theoretically impossible to arrive at a perfect formula to calculate how much better one rating is relative to another.

The method used for this quantitative analysis is the transforming subjective ratings to a ratio scale proposed by Nielsen[8]. The purpose of this method is to compare relative ratings of the various interface iterations.

The method is performed by calculating overall usability in relative terms as the geometric mean of the normalized values of the relative improvements in the individual usability criteria. The geometric mean increases more when all the usability criteria improve a little than when a single criterion improves a lot and the others are stagnant.

Table 2.

For example, if median of fuzzy composite score of learnability criterion is 3.6 while completing a task with version 1 and 4.2 with version 2 of an interface, the normalized usability of interface 2 with respect to learnability is 123 percent of the usability of interface 1, indicating a 23 percent improvement in usability. The relative change of cognitive characteristic in an example of Table 2 is 125.6 percent. This means 25.6 percent improvement in cognitive characteristic of usability.

Conclusions

The methodology considered for usability characteristics and criteria is viewed as a step toward a more rational approach to user interface usability assurance. A set of criteria presented in this study provides a mean for quantitatively specifying the level of usability possessed by an interactive system.

The application of a fuzzy set theory approach for criteria-based analysis of usability is of benefit to quantifying and manipulating qualitative statements, subjectivity of opinion, vagueness of usability concept.

The quantitative analysis methodology described here lends itself to sensitivity analysis. The figure of merit can be used to compare components within the system and to evaluate the overall system status. The quantifiable outcomes that allow some comparison between interfaces or signify the relative improvements in iterative design processes could help the decision making during the design and development phases.

References

[1] Aczel, J., and Saaty, T. L., "Procedures for synthesizing Ratio Judgments", Journal of Mathematical Psychology, 27, pp.93-112, 1983

[2] Billingsley, P. A., " Ergonomic Standards Go Beyond Hardware", IEEE Tr. on Software, pp.82-84, March, 1994

[3] Dzida, W., Herda, S., and Itzfeldt, W. D., " User-Perceived Quality of Interactive Systems", IEEE Tr. on Software Engineering, Vol. SE-4, No. 4, 1978

[4] Grudin, J., "The Case Against User Interface Consistency", Communication of the ACM, 32(10), pp. 1164-1173, 1989

[5] Kiralowski, J., Porteous, M., and Corbett, M., "How to use the software usability measurement inventory(SUMI)", Proceedings of European Conference on Software Quality, Madrid, November, 1992

[6] Kakeyoo, N., "Ergonomic design approach", In Illustrated Ergonomics, Japan Institute of Standards, 1990

[7] Nagamachi, M., Kansei Engineering, Tokyo, Japan: Kaibundo Press,1989

[8] Nielsen, J., "Iterative User-Interface Design", IEEE Computer, November, pp.32-41, 1993

[9] Park, K. S., and Kim, J. S., "Fuzzy weighted-checklist with Linguistic Variables", IEEE Tr. on Reliability, 39(3), pp.1-5, 1990

[10] Ravden, S. J., and Johnson, G. I., "Evaluating Usability of Human-Computer Interfaces", Ellis Horwood, Chichester, 1989

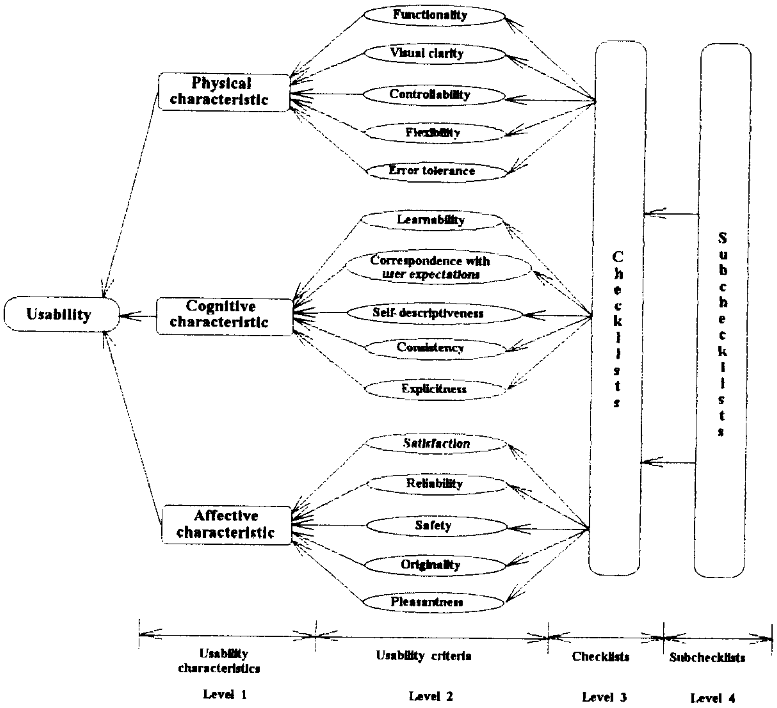

Figure 1. Usability characteristics and criteria tree

Table 1. Comparision of factors of user interface usability with several representatives

ISO 9241 Part-10 Dialogue Principles	User-perceived Quality of Interactive systems	SUMI	Ravden & Johnson's criteria
			✧visual clarity*
suitability for the task			✧appropriate functionality
✧controllability	user control	control	control
suitability for individualization	flexibility in task handling		✧flexibility
✧error tolerance	fault tolerance		error prevention and correction
suitability for learning	ease of learning	✧learnability	
conformity with user expectations	✧correspondence with user expectations		compatibility
✧self-descriptiveness	self-descriptiveness	helpfulness	information feedback user guidence &support
			✧consistency*
	problem adequate usability*	efficiency* ✧effect*	✧explicitness*

✧ A selected criterion for our study
* A criterion that has no common ground

Table 2. An example of transforming rating score and calculating improvement score

cognitive characteristic	composite score in version 1 (V1)		composite score in version 2 (V2)		improvement score
	median	transformed value	median	transformed value	V2 / V1 (%)
learnability	3.6	1.2231	4.2	1.5091	123
correspondence with user expectations	4.7	1.8266	5.1	2.1715	119
self-descriptiveness	4.5	1.6881	4.6	1.7551	104
consistency	4.2	1.5091	5.3	2.3956	159
explicitness	3.8	1.3099	4.5	1.6881	129

geometric mean: 125.6(%)