Modelling Interacting Web Usability Criteria through Fuzzy Measures

Miguel-Angel Sicilia[1] and Elena García[2]

[1] DEI Laboratory, Computer Science Department, Carlos III University
Avda. de la Universidad, 30, 28911 Leganés, Madrid, Spain
`msicilia@inf.uc3m.es`
[2] Computer Science Department, Alcalá University
Ctra. Barcelona km. 33.600, 28871 Alcalá de Henares, Madrid, Spain
`elena.garciab@uah.es`

Abstract. Usability is a multifaceted concept that is usually evaluated in terms of a number of aspects or attributes that are not independent but interacting, due to their correlation or to concrete, context-specific design interests. In consequence, the aggregation of partial scores should be carried out by devices that are able of taking into account these interactions in the resulting overall score. In this paper, we describe how fuzzy measures can be used and evaluated as models of a number of the interactions between usability aspects in evaluation processes. It is also described how they can be subsequently used to obtain more realistic final scores by using aggregation operators that are based on that measures, like the Choquet integral.

1 Introduction

Usability is a multifaceted concept [11] that encompasses several attributes regarding the interaction of humans with software systems, which are in many cases summarized in three generic interdependent aspects named efficiency, effectiveness and satisfaction. Many usability evaluation processes include an aggregation stage in which partial scores regarding different attributes or measures need to be summarized in an overall score (which is specially important for automated tools [6]). Examples include the aggregation of the results of different factors in the `WAMMI` questionnaire [7] and other models [3] and analysis tools. These specific aggregation processes are usually carried out by using (weighted or not) arithmetic means — like in `WAMMI` and some usability metrics [1] —, which are not appropriate for cases in which the aggregated criteria interact in some way. Thus, other forms of aggregation are required if meaningful final scores are desired. In addition, it should be noted that the results of evaluation processes are inherently imprecise or uncertain due to the subjectiveness involved in the process — e.g., in heuristic evaluation [9] — or to the limitations of the measurement instrument — e.g., in the case of questionnaires — so that aggregation operators that are based on fuzziness [4] are good candidates for the task.

In this paper, we describe how fuzzy measures can be used to model interactions between diverse usability criteria, resulting in more realistic summarization

processes (a related approach is used in `WebQM` [10]). These measures are able of considering empirical research results about the correlations existent between usability aspects, but also expert or designer considerations about their relative importance in each specific evaluation setting.

2   Fuzzy Measures as Models for Interacting Usability Criteria

Several types of interactions between usability attributes [11] can be modelled by fuzzy measures. Among them, we have known relationships in the form of correlations and designer-established interactions. The latter include substitutiveness and complementarity between criteria and preferential dependencies as described in [8]. A fuzzy measure on a set X is a monotonic (i.e. $v(S) \leq v(T)$ whenever $S \subseteq T$) set function $v : 2^X \to [0 \dots 1]$.

To illustrate the modelling approach, the following fuzzy measure can be used to describe a concrete evaluation case in which four well-known usability attributes are used as criteria: efficiency (e), memorability (m), satisfaction (s) and learnability (l). Two interactions between criteria are considered. First, a positive correlation is assumed between efficiency and satisfaction (cited in [5], although it's validity depends on the application context). Second, memorability and learnability are considered as substitutive by the experts, so that the presence of both of them is not strictly additive. Table 2 shows one of the possible fuzzy measures modelling the just described interactions between criteria.

Table 1. Case Study fuzzy measure $v(X)$

n = 1	n = 2	n = 3	n = 4
$\{e\} \to 0.4$	$\{e,s\} \to 0.5$	$\{e,m,s\} \to 0.6$	$\{e,m,s,l\} \to 1$
$\{m\} \to 0.15$	$\{e,m\} \to 0.55$	$\{e,s,l\} \to 0.65$	
$\{s\} \to 0.3$	$\{e,l\} \to 0.55$	$\{e,l,m\} \to 0.6$	
$\{l\} \to 0.15$	$\{m,s\} \to 0.45$	$\{m,l,s\} \to 0.5$	
	$\{s,l\} \to 0.45$		
	$\{l,m\} \to 0.2$		

Note that in Table 2 the values are heuristically selected to model the interaction between criteria. For example, the weight of $\{e,m\}$ is only slightly augmented when the substitutive criterion l is added to the set, and the set of correlated criteria $\{e,s\}$ has a weight of 0.5, which represents a difference of 0.2 from absolute additivity. In more complex cases, mathematical functions are required to model those relationships. The requirement for two correlated criteria i and j is that they are not completely additive, i.e., that $v(\{i,j\}) < v(\{i\}) + v(\{i\})$. Two substitutive criteria are required to satisfy the relationship expressed in 1, so that the addition of a substitutive criterion have a small effect in the fuzzy measure (having no effect if the criterion are completely interchangeable).

$$v(T) < \begin{Bmatrix} v(T \cup i) \\ v(T \cup j) \end{Bmatrix} \approx v(T \bigcup \{i,j\}) \quad T \subseteq X - \{i,j\} \tag{1}$$

The discrete Choquet integral introduced can be used as a generalization of the weighted arithmetic mean that accounts for interacting criteria [8]. The general expression of the integral given in 2 is a specific instance of the general form of a discrete aggregation operator on the real domain: $M_v : \mathbb{R}^n \to \mathbb{R}$, which takes as input a vector $x = (x_1, x_2, \ldots, x_n)$ and yields a single real value.

$$\mathcal{C}_v(x) = \sum_{i=1}^n x_{(i)}[v(\{j|x_j \geq x_{(i)}\}) - v(\{j|x_j \geq x_{(i+1)}\})] \tag{2}$$

In expression 2 we have that $x' = (x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ is a non-decreasing permutation of the input n-tuple x, where $x'_{n+1} = \emptyset$ by convention. The integral is expressed in terms of a fuzzy measure (or Choquet capacity) v.

The results of the Choquet integral for the measure in Table 2 can be compared with those of a simple weighted mean ($\mathcal{W}$) to appreciate the differences (both expressed in a zero to three scale). For example, $\mathcal{C}(3,3,0,2) = 1.45$, while $\mathcal{W}(3,3,0,2) = 2.4$, showing how the second values overweights the correlated criteria (in addition, $\mathcal{C}(3,2,0,2) = 1.4$, so having a consistent small increment of 0.05, while the weighted mean delta is 0.3). In cases in which the participating criteria are independent, like $(0,2,2,0)$, $\mathcal{C}$ and $\mathcal{W}$ yield the same value. A thorough analysis lead to concluding that weighted means are not always appropriate overall usability scores.

3   On Evaluating Fuzzy Measure Design

Designer-dependant decision on the form of fuzzy measures must be empirically evaluated for adjustment. This entails asking about the degree to which the resulting aggregations are considered as (approximately) realistic to domain experts (in our case, usability experts) with regards to the established correlations and criteria. A procedure to accomplish this task can be that of eliciting a membership function describing the *appropriateness* of the aggregation. Existing membership evaluation techniques can be used for this purpose [2]. A modified pooling technique can be used to validate the aggregation. The technique proceeds by taking a evenly distributed sample of input vectors, and asking each evaluation participant, for each input value, the question "is the output value appropriate for the design at hand?" — after informing him/her about concrete the design considerations embodied in the fuzzy measure —, and whenever the answer is negative, a gradation of appropriateness in the [0..100] interval is asked for. A weighted mean using the weights $v(x)$ for each individual criterion is used as a clue to help experts in assessing the quality of the aggregation with respect to it. The resulting n-dimensional function can be analyzed to help in detecting potential pitfalls in the fuzzy measure (this graphical inspection approach will be subject of future analysis in the search of heuristics for the task itself).

4   Conclusions and Future Work

The use of fuzzy measures to model different interactions between usability attributes used as criteria has been described. The Choquet integral can be used to use these measures for the aggregation of partial scores in global ones that reflect a summarized usability analysis estimation. Techniques for membership elicitation have been proposed as evaluation procedures for the adequacy of the design of those fuzzy measures.

Future work should address a systematic analysis of the properties and validity of the approach with regards to current usability definitions and empirical evidence.

References

1. Babiker, E. M., Fujihara, H. and Boyle, C.D.B.: A metric for hypertext usability. A metric for hypertext usability. In: Proceedings of the 9th annual international conference on Systems documentation (1991):95–104
2. Bilgiç, T. and Türksen, T.: Measurement of Membership Functions: Theoretical and Empirical Work. In: D. Dubois and H. Prade (eds.) Handbook of Fuzzy Sets and Systems Vol. 1, Chapter 3, Fundamentals of Fuzzy Sets, Kluwer (1999):195–232
3. Brajnik, J.: Towards valid quality models for websites. In Proc. of 7th Human Factors and the Web Conference, Madison, Wisconsin June 2001.
4. Calvo, T., Kolesárová, A., Komorniková, M. and Mesiar, R. : Aggregation Operators: Basic Concepts, Issues and Properties. In Calvo, T., Mayor, G. and Mesiar, R. (eds). Aggregation Operators: New Trends and Applications. Studies in Fuzziness and Soft Computing, 97 (2002):3–106
5. Frøkjær, E., Hertzum, M. and Hornbæk, K.: Measuring Usability: are effectiveness, efficiency and satisfaction really correlated?. In: Proc. of Human Factors in Computing Systems (2000):345–352
6. Ivory, M.Y. and Hearst, M.A.: The State of the Art in Automated Usability Evaluation of User Interfaces. ACM Computing Surveys, 33(4) (2001):1–47
7. Kirakowski, J., Cierlik, B. Measuring the usability of web sites. Human Factors and Ergonomics Society Annual Conference (1998)
8. Marichal, J.L.: An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria, IEEE Transactions on Fuzzy Systems 8 (6) (2000):800–807
9. Nielsen, J.: Heuristic evaluation. In Nielsen, J., and Mack, R.L. (eds.): Usability Inspection Methods, John Wiley and Sons, New York (1994):25–61
10. Olsina, L. and Rossi, G.: Measuring Web Application Quality with WebQEM. IEEE Multimedia Magazine 9(4) (2002):20–29
11. Van Welie, M., van der Veer, G.C. and Eliëns, A. (1999), Breaking down usability. In: Proc. of Interact'99 (1999):613–620