Miguel-Ángel Sicilia; Elena García Barriocanal; Tomasa Calvo

An inquiry-based method for Choquet integral-based aggregation of interface usability parameters

Kybernetika, Vol. 39 (2003), No. 5, [601]--614

Persistent URL: `http://dml.cz/dmlcz/135558`

KYBERNETIKA — VOLUME 39 (2003), NUMBER 5, PAGES 601–614

AN INQUIRY-BASED METHOD FOR CHOQUET INTEGRAL-BASED AGGREGATION OF INTERFACE USABILITY PARAMETERS

MIGUEL–ÁNGEL SICILIA, ELENA GARCÍA BARRIOCANAL AND TOMASA CALVO

The concept of usability of man-machine interfaces is usually judged in terms of a number of aspects or attributes that are known to be subject to some rough correlations, and that are in many cases given different importance, depending on the context of use of the application. In consequence, the automation of judgment processes regarding the overall usability of concrete interfaces requires the design of aggregation operators that are capable of modeling approximate or ill-defined interactions among criteria. In addition, justified expert opinions are given a prominent status in the current practice of usability evaluation, which points to the convenience of including experts as an integral part of the aggregation operator design process. On the basis of these assumptions we review in this paper possible approaches to design a suitable aggregation operation and describe a method for such kind of design process that explicitly models expert-elicited relationships among criteria, enforcing some properties on a Choquet capacity. The method subsequently uses experimental data to fine-tune operator design. A case study is described to illustrate the method, and a comparative study with other common aggregation approaches is also provided.

Keywords: usability, Choquet integral, inquiry methods
AMS Subject Classification: 68U35, 68T37, 03B52, 47S40, 28E10

1. INTRODUCTION

The *usability* of a man-made artifact is a concept which is implicitly understood by most people, as we all know that it is related to finding the artifact more or less easy to use. However usability is a difficult concept to precisely define and measure, as many aspects of the artifact contribute to how it is used and perceived by humans. In the context of the evaluation of man-machine interfaces, usability is often break down into measurable elements which include effectiveness, efficiency, and satisfaction, as defined by ISO [10]. But this characterization of the concept is by no means considered as definitive, and consequently, a considerable amount of distinct "usability attribute lists" have been described and used. In an attempt to avoid such divergences, a layered model of usability has been proposed that provides a common understanding of the concept, useful to allow comparisons form theoretical or practical viewpoints [24]:

— The first layer is an "abstract" level including the three mentioned ISO aspects.

— The second level of "usage indicators" includes aspects that "can actually be observed in practice", including learnability, performance speed, errors/safety, memorability and satisfaction, each of them connected as contributors to one of the aspects in the first level.

— A third level of "means" that "are not goals by themselves" includes elements like consistency, adaptability and feedback, that are connected to one or several indicators has having a potential positive or negative impact on them (for example, consistency may lead to improved learnability).

Nonetheless, even in the case that one of such kind of layered models eventually becomes universally accepted, it still remains open the issues of what are the relationships among usability aspects [6], and how should a number of usability aspects be aggregated to come up with an overall (or global) measure of usability.

Anyway, many usability evaluation processes include an aggregation stage in which partial scores regarding different attributes or measures need to be summarized into a single overall score. This is specially the case in automated usability analysis tools [11], that attempt to provide estimators of usability from measures collected by software modules informed with correlations discovered among usability attributes that are not precisely understood yet [6]. In addition, transparent and self-explanatory aggregation means would be required to come up with usability metrics tailored to benchmarking industrial systems of specific kinds. Examples of that kind of processes are the aggregation of the results of different factors in the `WAMMI` questionnaire [14] and other usability models [3].

In this paper, we approach the design of aggregation operators for usability aspects, in the common case in which a small number of attributes must be summarized in a unique figure of usability, considering both the given usage context and existing general knowledge about the attributes considered. More concretely, we describe how expert-elicited fuzzy measures can be used to model interactions between diverse usability criteria, in an attempt to come up with more realistic summarization processes.

The results described here are based on a previous experiment design reported in [21], and the definition of an expert-based method for the design of Choquet capacities provided in [22]. In addition, a comparison is drawn with a number of other common approaches to design aggregation operators, with the intention of estimating, from a pragmatic point of view, the adequacy of our inquiry-based approach.

The rest of this paper is structured as follows. Section 2 describes how fuzzy measures can be used to model interacting usability aspects or criteria. Then, an inquiry-based method for the design of such measures is described in Section 3. Then, a concrete case study illustrating the method is provided in Section 4, along with a comparison with a number of well-known approaches to model aggregation processes. Finally, conclusions and future research directions are provided in Section 5.

2. FUZZY MEASURES AS MODELS FOR INTERACTING USABILITY CRITERIA

The Choquet integral has been described elsewhere [17] as an aggregation operator enabling the explicit modelling of fuzzy interactions among criteria. This can be accomplished by a careful selection of the mappings included in the fuzzy measure used as the Choquet capacity. A fuzzy measure on a set X is a monotonic (i.e. $v(S) \leq v(T)$ whenever $S \subseteq T$) set function $v : 2^X \rightarrow [0,1], v(\emptyset) = 0, v(X) = 1$. Concretely, several types of interactions between usability attributes [24] can be modelled with 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 by Marichal in [17]. Here, we will deal only with correlation and substitutiveness, since they are the two types of interactions identified in the case study described below.

It must be always taken into account that the requirement for two correlated criteria i and j is that they are subadditive, as showed in (1)

$$v(\{i,j\}) < v(\{i\}) + v(\{j\}) \tag{1}$$

In addition, two substitutive criteria are required to satisfy the relationship expressed in (2), so that the addition of a substitutive criterion has a small effect in the fuzzy measure (having no effect if the criteria are completely interchangeable).

$$v(T) < \left\{ \begin{array}{c} v(T \cup i) \\ v(T \cup j) \end{array} \right\} \approx v(T \cup \{i,j\}) \; ; \; T \subseteq X - \{i,j\} \tag{2}$$

The discrete Choquet integral can be used as a generalization of the weighted arithmetic mean that accounts for interacting criteria [17]. The general expression of the integral given in (3) is a specific case of the general form of a discrete aggregation operator on the real domain: $M_v : R^n \rightarrow R$, which takes as input a vector $x = (x_1, x_2, \ldots, x_n)$ and yields a single real value.

$$C_v(x) = \sum_{i=1}^{n} x_{(i)}[v(\{j | x_j \geq x_{(i)}\}) - v(\{j | x_j \geq x_{(i+1)}\})] \tag{3}$$

In expression (3) we have that $(x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ is a non-decreasing permutation of the input n-tuple x, where $x_{(n+1)} = 0$ by convention. The integral is expressed in terms of a fuzzy measure (or Choquet capacity) v.

It should be noted that the number of usability attributes that must be measured in a evaluation is always small. Even in the case of a large list of them, like the one established in [5], authors themselves suggest an informal aggregation of attributes. For example, Dix proposes 14 usability attributes, but they are aggregated in just three ones: Predictability, synthesability, familiarity, generalizability and consistency are aggregated in *learnability* attribute; dialog initiative, multi-threading, task migrability, substitutivity, customizability are aggregated in *flexibility* attribute and observability, recoverability, responsiveness, task conformance are aggregated in *robustness.* If the number of attributes were greater than usually and there were no interactions between attributes in the sets of cardinality $k+1$, the definition of k*-additive* measures [9] would be required, in order to decrease the complexity of fuzzy measure definition.

3. AN INQUIRY-BASED METHOD FOR THE DESIGN OF USABILITY EVALUATION AGGREGATION OPERATORS

Expert judgments are given a prominent status in current usability evaluation practices, as evidenced by the wide use of inspection methods (e.g. heuristic evaluation or cognitive walk-through) for the purpose of usability evaluation [18]. This suggests that designing usability attribute aggregation operators should follow some sort of inquiry process in which both the specific requirements of the problem at hand and expert knowledge on usability are combined to come up with realistic summarizations.

Continuing research reported in [22], in this section we describe a method that approaches aggregation operator design as an inquiry system regarding an uncertain situation. A related approach is used in `WebQEM` [20], that proposes using nonlinear multicriteria scoring models to aggregate even fine-grained usability metrics, but a specific approach for aggregation operator design is not provided, so that evaluators freely provide concrete hierarchical aggregation parameters through a graphical user interface.

In this section, we first outline an inquiry-based method to design fuzzy measures for the purpose of aggregation, and then a concrete case study in which such approach has been used is discussed.

3.1. Outline of the inquiry-based method

The first phase of the method bears some resemblance with Delphi inquiry techniques, since it shares the same set of general features that are used to characterize Delphi processes – according to Dalkey in [16] – namely: involving a group, having an information goal, being uncertain to some extent, and approaching the inquiry in a structured manner. Figure 1 depicts the phases and artifacts generated in the course of the inquiry process.

The process departs from a body of existing empirical evidence (obtained *ad hoc* or gathered from the literature in the concrete field of study), plus a number of pre-conceptions about the aggregation context and the required criteria, possibly including business rules or other kind of concrete restrictions. Then, two sequential (and iterative if required) phases of a very different nature take place, being the second one dependent on the first and eventually optional:

- **Inquiry Phase.** This phase proceeds as an streamlined Delphi process, involving a sequence of two or more structured rounds in which a group of experts (group A) is asked to provide a fuzzy measure for the given case. In the first round, this is done individually, but subsequent rounds may involve feedback

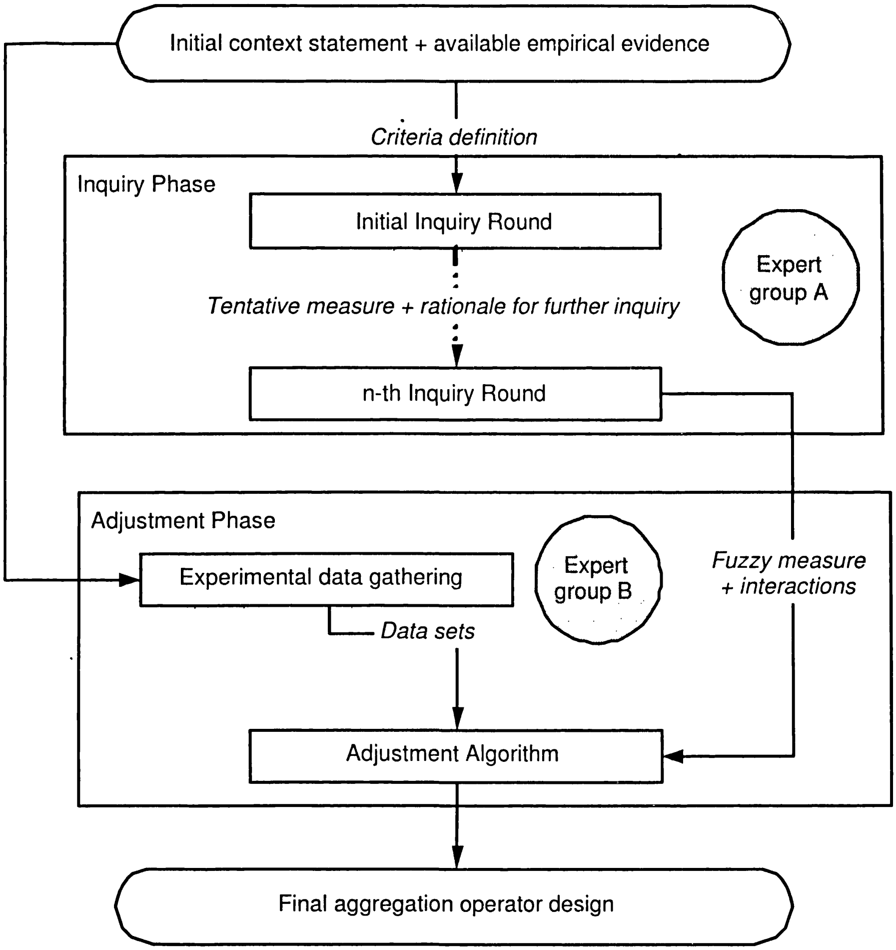

Fig. 1. Overall view of the inquiry-based method.

from other individuals in the group regarding justifications, claims or argues regarding concrete aspects of the measure.

- **Adjustment Phase.** Once the inquiry has resulted in a fuzzy measure, a data collection phase proceeds with a (possibly) different set of experts (group B), as a mean to validate the design. Obtaining of such data may follow a blind approach in which only the departure assumptions are provided, in an attempt to discover possible biases of expert group A. This results in a collection of data involving expert assessments of concrete aggregation cases, that can be used later as the empirical base to adjust the design of the inquiry phase, always preserving the criteria interactions that were established.

The process ends with a fuzzy aggregation operator design including the concrete Choquet capacity, but also the set of conceptual decisions that were obtained as a result of the inquiry, which provide the explanation for the relationships among the concrete mappings in the fuzzy measure.

3.2. Case study description

To illustrate the modelling approach, the following fuzzy measure can be used to describe an example of concrete evaluation case in which four well-known usability attributes are used as criteria:

— efficiency (e): how effectively can a user perform their job using the system?

— memorability (m): can a user, who has used the system before, remember how to use it effectively next time: does the user have to learn everything from the beginning?

— satisfaction (s): how much does the user like using the system?

— and learnability (l): how fast can a novice user learn how to use the user interface sufficiently to accomplish basic tasks?

Two interactions between criteria are considered in our concrete study. First, a positive correlation is assumed between efficiency and satisfaction (cited in [6], although the validity of this interaction depends on the application context). This interaction, according to [17], expresses that the marginal contribution of satisfaction to every combination of criteria that contains efficiency is strictly less than the marginal contribution of satisfaction to the same combination when efficiency is excluded. Second, memorability and learnability are considered as substitutive by the experts (in the specific example context), so that the presence of memorability or learnability produces almost the same effect than the presence of both.

Table 1 shows one of the possible fuzzy measures modelling the just described interactions between criteria. In this example and in the general case we have considered two basic assumptions:

1. The number of usability attributes that must be measured in a evaluation is always small, as previously mentioned.

2. In usability evaluation, the scale to measure attributes is usually small. Often numeric Likert scales that range from 1 to 5 or 7 are used.

Five usability experts took part in the assessment. The process began with an informative session in which the usability evaluation context (but not the concrete Choquet integral-based aggregation approach) was described, including the following explanations:

(i) The specific type of interface they have to evaluate. In this study, we have selected a typical *e-mall* application that enables the acquisition of different kind of goods.

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

$n=1$	$n=2$	$n=3$	$n=4$
$\{e\} \to 0.4$	$\{e,s\} \to 0.5$	$\{e,m,s\} \to 0.95$	$\{e,m,s,l\} \to 1$
$\{m\} \to 0.15$	$\{e,m\} \to 0.8$	$\{e,s,l\} \to 0.95$	
$\{s\} \to 0.3$	$\{e,l\} \to 0.8$	$\{e,l,m\} \to 0.85$	
$\{l\} \to 0.15$	$\{m,s\} \to 0.45$	$\{m,l,s\} \to 0.5$	
	$\{s,l\} \to 0.45$		
	$\{l,m\} \to 0.2$		

(ii) The usability attributes that must be taken into account and their relevance (weights) are justified, according to the nature of the interface. Selected attributes are those described in Section 3.2: efficiency, memorability, satisfaction and learnability, with the weights showed in Table 1.

(iii) The interactions between attributes and their meaning, explaining the experts how these interactions have been previously identified.

Once the experts know these details the second sub-phase of the evaluation takes place. Twenty randomly selected input values are given to each of the experts. To avoid the selection of non-realistic evaluations, input values with three or more attributes scored with zero are discarded. Then, the sample is obtained randomly selected twenty equidistant values from the input set X.

Then, experts are encouraged to elaborate an aggregation value for each of the x_i in the sample, taken into account the weights of the usability attributes and the interactions among them, but without information about the fuzzy measure and the fact that the Choquet integral was used to obtain them. In addition, it was decided not to provide the experts with the result of the previously designed operator, to prevent any form of confirmation biases [15] or selective thinking originated from that results. It should be noted that aggregating usability issues is often done by usability experts, for example, when rating defects in heuristic evaluation or when making multi-attribute design comparisons.

As a result, a data set ds consisting on pairs $(x, o(x))$ are obtained, where $o(x)$ denotes the output value given by the expert. In case that more than one expert evaluates the same input, the average of the given output values is used. Since experts judged the outcome intuitively, without using any formal mean or procedure, divergences in the values of $o(x)$ and $C_v(x)$ for any given input x – if larger enough – can be interpreted as potential imperfections of the designed fuzzy measure.

Concretely, a global average error $E(C_v, ds)$ can be computed as $\sum_{x \in ds} |o(x) - C_v(x)| \cdot \frac{1}{|ds|}$. In consequence, a process of readjustment that strictly respects the interactions embodied in the design of the fuzzy measure may lead to a more realistic aggregation device. This approach of adjustment based on a limited sample of inputs considered by a small number of experts fits in the notion of 'discount usability' [19] that have been proven economically useful in the area of usability evaluation. Obviously, the repeated evaluation of the entire range of possible inputs would provide better adjustments, but in most practical situations, it simply becomes economically prohibitive.

Once the evaluation data is available, and provided that a significant[1] global error has been detected, the following straightforward heuristic greedy algorithm is used for the refinement of the fuzzy measure.

```
error := E(Cv, ds)
while (error > epsilon)
      and (#-iterations < MAX) do begin
    v' := modify-measure-value(Cv, ds)
    if (E(Cv', ds)< error) then begin
        v := v'
        error := E(Cv, ds)
    end
    inc(#-iterations)
end
```

The `modify-measure-value` function is the core of the adjustment process. In our current tentative algorithm, it proceeds by slightly changing one of the non-singleton values of the fuzzy measure v, provided that the resulting updated v' continues to be a fuzzy measure, and still satisfies the design considerations discussed above, i.e. (1) and (2). The following pseudocode illustrates the process, in which one of the mappings of the fuzzy value is changed, and then the resulting mapping is checked to verify if it still complies to the design considerations (`constraints`).

```
modify-measure-value(v:measure; ds:data-set,
           epsilon: real) : measure
begin
    changed := false
    while (not changed)
       v' := try-change(v, ds, epsilon)
       changed := check(v', constraints)
    end
    return v'
end
```

It should be noted that the greedy heuristic proceeds by exploring the spaces of conforming fuzzy measures, and eventually the algorithm may not provide any significant improvement or may fail to find an adequate fuzzy measure for a given tolerance to error ϵ. The obtention of some sort of informed strategy for the search – i.e. an heuristic for the `try-change` procedure – will be addressed in the future, but data provided in the following section support the hypothesis that the greedy algorithm provides a good-enough adjustment. Relevant related work on heuristics is described in [7].

[1]The characterization of significance here is not straightforward, since it embodies also some degree of subjectiveness, and it can be partially attributed to inconsistent human aggregation behaviors.

Table 2 provides the final fuzzy measure obtained after a concrete execution of the algorithm after about five thousand iterations. Nonetheless, the algorithm provided adjustments with only a slightly worse performance within three hundred iterations in some cases.

Table 2. Adjusted Case Study fuzzy measure $v(X)$.

$n=1$	$n=2$	$n=3$	$n=4$
$\{e\} \rightarrow 0.4$	$\{e,s\} \rightarrow 0.574824600299883$	$\{e,m,s\} \rightarrow 0.9001076634152206$	$\{e,m,s,l\} \rightarrow 1$
$\{m\} \rightarrow 0.15$	$\{e,m\} \rightarrow 0.6005739410111545$	$\{e,s,l\} \rightarrow 0.9016744863484101$	
$\{s\} \rightarrow 0.3$	$\{e,l\} \rightarrow 0.6278689487788578$	$\{e,l,m\} \rightarrow 0.6999860286841342$	
$\{l\} \rightarrow 0.15$	$\{m,s\} \rightarrow 0.4538956126629927$	$\{m,l,s\} \rightarrow 0.47933582886918746$	
	$\{s,l\} \rightarrow 0.4793311052382566$		
	$\{l,m\} \rightarrow 0.17457855415698448$		

4. COMPARISON WITH OTHER APPROACHES

The inquiry-based method described above represents an *aprioristic* account for aggregation operator design, that proceeds by explicitly modelling criteria interaction. In this section, we provide a quantitative comparison of a number of other techniques that can be used for the same purpose, but depart from a different design philosophy. Of course the list of techniques applied here do not exhaust the wide range of possibilities for aggregation operator design [4], but provides a comparative view of our inquiry-based method with well-known techniques for which software tools are available. Other methods for adjusting Choquet capacities to empirical data – like Grabisch's methods described in [8] – are targeted to problems with different initial settings, so that we do not discuss them here. Table 4 summarizes the results of the comparative study (adjusted expert inquiry refers to our method after using the greedy heuristic for fine-tuning the fuzzy measure).

The concrete parameters of the methods that have been used in the study are described in Table 3. The second column in Table 4 provides the average global error of each aggregation method for the two data sets used in the study, that are called training (t) and validation (v), consisting in 38 and 40 expert-elicited aggregations obtained as described in the previous section. The training data set is used as such for all the methods that depart from data to design the aggregation operator (all except the inquiry-based and rule aggregation ones), while the validation data set is used to test the *generalizability* of the resulting operator. The third column provides a measure of the predictive value of each method, measured as the percentage of predictions whose accuracy with respect to expert estimations falls in a 15% of error. The fourth column provides the maximum errors, and the last column provides the absolute divergence of each method with respect to the adjusted results of the inquiry-based method. The quantitative results evidence that the inquiry approach provides results comparable to other techniques, both in the case of using the greedy adjustment heuristic or not. Remarkably, the neural network approach is able of providing significantly better adjustment to training data, but such accuracy is dependent on the data used to train the network, as evidenced by the increase on error when querying the network with a different data set.

Table 3. Details of the parameters used for the comparative study.

Method	**Tool**	**Parameters**
Additive Choquet	Beliakov's AOTool	General fuzzy measure
Symmetric Choquet	Beliakov's AOTool	General fuzzy measure
2-additive Choquet	Beliakov's AOTool	General fuzzy measure
3-additive Choquet	Beliakov's AOTool	General fuzzy measure
Neural Network	EasyNN-plus	Backpropagation network with a single hidden layer with 5 nodes. Learning rate of 0.6 and momentum of 0.8.
Least-Square Regression Method (LSM)	EasyNN-plus	Default setting
f-Regression	Furea Tool	Fuzzy points with $m = 2$ and default spread value. Using the arithmetic mean as compensatory aggregation operator for the regression method.
Rule-based aggregation	Fuzzy Java Toolkit (FJT)	Mamdami rule executor and weighted mean deffuzificator, aggregating the output of the chain of rules with fuzzy set union

Beliakov's AOTool[2], EasyNN-plus[3], Furea Tool[4], Fuzzy Java Toolkit (FJT)[5].

The Choquet aggregators obtained with Beliakov's tool [1, 2] provide similar or slightly worse approximations for the problem at hand, specially when considering their predictive P(0.15) accuracy. This points out that an inquiry-based method can be considered worth the effort, provided that the fuzzy measure embodies and satisfies a number of interaction requirements that explain the design rationale for the given situation.

The *f-regression* method [13] provided in the Furea software tool [12] provides a fuzzy regression approach in which fuzziness is considered in the experimental data. An interesting point regarding this method is its high degree of predictive power with P(0.15), outperforming classical LSM regression approaches.

The Java program using FJT that was used as rule-based aggregator included the 16 rules determined by experts that are showed in Table 5, with the terms 'low',

[2] http://www3.cm.deakin.edu.au/ gleb/aotool.html
[3] http://www.easynn.com
[4] http://www.fuzzy.ru/
[5] http://www.iit.nrc.ca/IR_public/fuzzy/fuzzyJToolkit2.html

Table 4. Total average error (E), prediction (P), maximum error and divergence with the adjusted expert inquiry method for each of the methods.

Method	E(t/v)	P(.15)(t/v)	Max.err(t/v)	Div
Expert inquiry	.203/.184	50/57.5	.95/.65	.111/.066
Adjusted expert inquiry	.124/.152	76.31/57.5	.65/.55	-
Additive Choquet	.179/.209	55.3/45	.677/.625	.11/.15
Symmetric Choquet	.264/.253	34.2/30	.754/.6	.259/.226
2-additive Choquet	.291/.256	34.2/37.5	.79/.67	.289/.755
3-additive Choquet	.285/.258	28.9/40	.7/.61	.275/.221
Neural Network	.094/.227	81.6/40	.458/.717	.100/.172
LSM regression	.149/.187	52.6/42.5	.522/.521	.11/.158
f-Regression	.158/.202	79/50	.679/.74	.131/.141
Rule-based aggregation	.279/.35	23.7/22.5	.9/.9	.330/.29

'medium' and 'high' represented respectively by a right-linear fuzzy set RL(0.5, 1.5), a triangular fuzzy set T(1, 1.5, 2), and a left-linear fuzzy set LL(1.5, 2.5). This approach provides the benefit of a declarative, expert-understandable aggregator design, but for the described problem, a high number of rules is required to obtain an accuracy level comparable to the other methods, so that the initial benefits of declarative definition are limited by the effort required to elicit a high number of very similar rules.

Table 5. Rules used for the aggregation.

e	m	s	l	Usability
high				more or less medium
high		medium		somewhat high
high	high	medium		high
high		medium	high	high
high	high	high	high	very high
medium	medium	medium	medium	plus medium
medium	high	high	high	high
low	low	low	low	low
low		low		more or less low
medium	high	low		plus medium
low			high	somewhat medium
low	high			somewhat medium
low		medium	high	plus medium
low		high	high	plus medium
low	high	high		plus medium
high		high		more or less high

5. CONCLUSIONS

Inspired in the importance given to expert opinion in current usability evaluation practices, an inquiry-based method for the design of usability aggregation operators has been described. The method proceeds by eliciting a fuzzy measure to model different interactions between the required usability attributes used as criteria. Then, the Choquet integral is used for the aggregation of partial scores in global ones that represents a summarized usability analysis estimation. The resulting measure can afterwards be fine-tuned with a set of experimental data, using a simple greedy heuristic.

The inquiry-based method has been compared quantitatively with a number of other well-known approaches to aggregation design. The resulting measures point out that the inquiry-based method provides better or similar results than other techniques, with the additional benefit of obtaining an artifact – the fuzzy measure – that explicitly embodies criteria interactions, and the inquiry process provides the supplementary benefit of obtaining a justified rationale for aggregation design that gives it a higher degree of credibility. Nonetheless, the applicability of such an inquiry-based method is restricted to problems with a small amount of input criteria, and for which criteria interactions can be assessed in some uncertain but definite way. The inquiry method can also be applied to other domains in which similar situations of fuzzy interactions among criteria are regularly used to obtain overall scores, for example, a similar approach to that described here has been used recently to aggregate competence levels in human resource selection processes [23].

ACKNOWLEDGEMENTS

The work of the third author has been supported by BFM 2003-05308, and by the LOBFI research group from the University of Balearic Islands.

(Received August 22, 2003.)

REFERENCES

[1] G. Beliakov: How to build aggregation operators from data? Internat. J. Intelligent Systems *18* (2003), 903–923.

[2] G. Beliakov, R. Mesiar, and L'. Valášková: Fitting generated aggregation operators to empirical data. Internat. J. Uncertainty, Fuzziness and Knowledge-based Systems (2003) (to appear).

[3] J. Brajnik: Towards valid quality models for websites. In: Proc. 7th Human Factors and the Web Conference, 2001.

[4] T. Calvo, A. Kolesárová, M. Komorníková, and R. Mesiar: Aggregation operators: Basic concepts, issues and properties. In: Aggregation Operators: New Trends and Applications (T. Calvo, G. Mayor, and R. Mesiar, eds.), Studies in Fuzziness and Soft Computing *97* (2002), 3–106.

[5] A. Dix, G. Abowd, R. Beale, and J. Finlay: Human-Computer Interaction. Prentice Hall, Englewood Cliffs, N.J. 1998.

[6] E. Frøkjær, M. Hertzum, and K. Hornbæk: Measuring usability: are effectiveness, efficiency and satisfaction really correlated? In: Proc. Human Factors in Computing Systems, 2000, pp. 345–352.

[7] M. Grabisch: A new algorithm for identifying fuzzy measures and its application to pattern recognition. In: IEEE Fuzzy Systems Internat. Joint Conference 1995, pp. 145–150.

[8] M. Grabisch: The application of fuzzy integrals in multicriteria decision making. European J. Oper. Res. *89* (1996), 445–456.

[9] M. Grabisch: The interaction and Mobius representations of fuzzy measures on finite spaces, k-additive measures: a survey. In: Fuzzy Measures and Integrals. Theory and Applications (M. Grabisch, T. Murofushi, and M. Sugeno, eds.), Physica–Verlag, Heidelberg 2000, pp. 70–93.

[10] N. Bevan: ISO 9241: Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability. TC 159//SC 4 Technical Committee of International Organization for Standardization 1998.

[11] M. Y. Ivory and M. A. Hearst: The state of the art in automated usability evaluation of user interfaces. ACM Computing Surveys *33* (2001), 4, 1–47.

[12] B. Izyumov, E. Kalinina, and M. Wagenknecht: Software tools for regression analysis of fuzzy data. In: Proc. 9th Zittau Fuzzy Colloquium 2001, pp. 221–229.

[13] E. Kalinina and M. Wagenknecht: Fuzzy regression analysis and application to a crisp model. In: Proc. 8th Zittau Fuzzy Colloquium 2000.

[14] J. Kirakowski and B. Cierlik: Measuring the usability of web sites. In: Human Factors and Ergonomics Society Annual Conference 1998.

[15] J. Klayman and Y.-W. Ha: Confirmation, disconfirmation, and information in hypothesis testing. Psychological Review *94* (1987), 2, 211–228.

[16] H. Linstone and M. Turoff (eds.): The Delphi Method: Techniques and Applications. Addison Wesley, Reading 1975.

[17] J. L. Marichal: An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria. IEEE Trans. Fuzzy Systems *8* (2000), 6, 800–807.

[18] J. Nielsen: Heuristic evaluation. In: Usability Inspection Methods (J. Nielsen and R. L. Mack, eds.), Wiley, New York 1994, pp. 25–61.

[19] J. Nielsen: Guerrila HCI: using discount usability engineering to penetrate the intimidation barrier. In: Cost-Justifying Usability (R. G. Bias and D. J. Mayhew, eds.), Academic Press, New York 1994, pp. 245–272.

[20] L. Olsina and G. Rossi: Measuring Web application quality with WebQEM. IEEE Multimedia Magazine *9* (2002), 4, 20–29.

[21] M. A. Sicilia and E. García: Modelling interacting Web usability criteria through fuzzy measures. In: Proc. 3rd Internat. Conference on Web Engineering (Lecture Notes in Computer Science 2722), Springer–Verlag, Berlin 2003, pp. 182–185.

[22] M. A. Sicilia, E. García, and T. Calvo: On the use of the Choquet integral for the aggregation of usability interface related scores. In: Proc. Summer School on Aggregation Operators (AGOP 2003), University of Alcala de Henares (Spain), 2003, pp. 159–164.

[23] M. A. Sicilia, E. García, and R. Alcalde: Fuzzy specializations and aggregation operator design in competence-based human resource selection. In: Proc. 8th World Federation on Soft Computing Conference 2003 (to appear).

[24] M. Van Welie, G. C. van der Veer, and A. Eliëns: Breaking down usability. Proc. Interact'99 (1999), pp. 613–620.

Prof. Dr. Miguel-Ángel Sicilia, Department of Computer Science, University of Alcalá, Ctra. Barcelona km.33.6 – 28871 - Alcalá de Henares, Madrid. Spain.
e-mail: msicilia@uah.es

Prof. Elena García Barriocanal, Department of Computer Science, University of Alcalá, Ctra. Barcelona km.33.6 – 28871 - Alcalá de Henares, Madrid. Spain.
e-mail: elena.garciab@uah.es

Prof. Dr. Tomasa Calvo, Department of Computer Science, University of Alcalá, Ctra. Barcelona km.33.6 – 28871 - Alcalá de Henares, Madrid. Spain.
e-mail: tomasa.calvo@uah.es