Chapter #

An approach of computer-aided choice of UI evaluation criteria and methods

ANDRE NENDJO ELLA¹, CHRISTOPHE KOLSKI¹, FABRICE WAWAK², CATHERINE JACQUES³ AND PASCAL YIM⁴

¹ Laboratoire d’Automatique et de Mécanique Industrielles et Humaines
LAMII-UMR CNRS 8530, Le Mont Houy, BP 311, F-59304 Valenciennes Cedex, France
{anendjo, kolski}@univ-valenciennes.fr

² Laboratoire I3D, Bat. P2, University of Lille I
59651 Villeneuve d’Ascq cedex, France

³ Centre Lillois d’Etudes et de Recherches Sociologiques et Economique
CLERSE-URA CNRS 345, USTL, 59655 Villeneuve d’Ascq, and
Institut Fédératif de Recherche sur l’Economie et les Sociétés Industrielles (IFRESI)
2, rue des Canonniers 59800 Lille, France
C.JACQUES@ifresi.univ-lille1.fr

⁴École Centrale de Lille, LAIL URA CNRS 1440 - BP 48
59651 Villeneuve d’Ascq cedex, France
yim@ec-lyon.fr

Abstract Interactive system evaluation is the source of many difficulties for human-machine communication specialists as well as for non-specialists. The quality of the evaluation depends on many parameters which are often contradictory. A major problem which project managers (and even human evaluators) are confronted with is the choice of criteria and methods appropriate for the evaluation. Given the international context linked with the situation of use, social and cultural aspects must also be considered in certain cases. Our research is oriented towards the design of a Decision Support System (DSS) for computer-aided choice of user interface evaluation criteria and methods. It is based on Fuzzy sets. A first prototype is implemented with PrologII.

Keywords: Decision Support System, Evaluation, User Interface, Fuzzy Sets.
1. INTRODUCTION

For the evaluation of interactive systems, two main families of criteria are often distinguished [15]: they concern (1) practical acceptability which comprises several very different criteria, such as compatibility, cost, system reliability but also the usefulness; (2) social acceptability. The usefulness is composed of two main sub-criteria: utility (technical or functional dimension) and usability (ergonomic dimension). For instance, concerning the ergonomic dimension, Bastien and Scapin [2] propose a classification with 18 sub-criteria.

Several dozen evaluation methods are described in the literature (see for instance [3] [6] [9] [15] [18]) and concern one or more of these evaluation criteria. For instance, some of these methods concern essentially human workload evaluation. The choice problem concerning these criteria and methods is very complex for the non-specialists, and even for the specialists.

What is more the social criteria are never really taken into account, and yet in certain countries, the social considerations become as important as the utility and usability constraints. For example, when the interactive part of a system designed by a western company is intended for Developing Countries (in Africa or Asia), it can lead to a freezing situation or rejection phenomenon when some social and/or cultural requirements (specific shapes, gestures, colors, signs...) are not taken into account [14] [17]. In conclusion, in many situations, the evaluation must check not only the practical requirements (utility, usability, induced costs...) which are the source of many difficulties, but also social and cultural criteria [13]. For this reason, a computer-aided choice of evaluation criteria and methods can in many cases be particularly important. Such an approach is described in this paper. It is based on Fuzzy sets and implemented with PrologIII.

2. RELATED WORK CONCERNING COMPUTER-AIDED EVALUATION

Many computer-aided approaches for helping the evaluators of interactive systems are described in the literature. The aid principles are various:
- automatic evaluation of interaction objects using a knowledge base containing ergonomic rules, like with Synop [10], KRI [11] or ERGOVAL [5];
- consideration of a task model for helping the evaluation, such as in EMA [3] or Chimes [8];
- Automatic capture of user actions, such as in EMA [3], Playback [12] or in the system proposed by Hammontree et al. [7];
- Automatic analysis of visual strategies [1];
Advice concerning the evaluation methods to use [4];
and so on.

Most of the tools dedicated to computer-aided evaluation help the evaluators to perform only partial evaluations. The evaluation is partial insofar as the evaluators can only check a part of the criteria set: in other words, at present, any given tool for computer-aided evaluation can help the evaluators to check only few criteria such as "efficient to use" or "subjective pleasing" (the same tool would be for instance ineffective for other criteria such as "easy to learn").

We adhere to the logic of these tools existing for computer-aided evaluation, but in an overall perspective, carrying out a tool for computer-aided evaluation. It is called ADHESION (in French: Aide à la Décision Humaine pour l’Évaluation des Systèmes Interactifs et leur validation; translation: Decision Support System for Interactive System Evaluation and Validation). Its main objective is to propose and explain the appropriate set of criteria and methods, usable for the "complete" evaluation (i.e. not only based on usability criteria) of a human-computer interface. To our knowledge, such a system does not have an equivalent in the literature. The principles on which it is based are described now.

3. FIRST PROTOTYPE OF THE ADHESION DECISION SUPPORT SYSTEM

3.1 General principles

The analysis of a Decision Support System (DSS) requests the specifications of the elements which take place in the system. First, we must define the set of alternatives on which will be based the decisions. Secondly, we should study the criteria which will be taken into account. This information will be sorted out and processed by a predefined decision-making approach in order to deduce the choice of the best alternative [16], figure 1. In our case, the result concerns the adequate HCI evaluation criteria and methods to be used for the given situation. The possible decision criteria are numerous:

- About the situation: quality of the available information, Type of data (qualitative/quantitative), type of application (supervision, electronic commerce, business system...), presence of representative users, stage in the development process, human resources, available financial resources, location (country), product nature (prototype, paper version...), type of human-computer interaction (menus, virtual reality, hypermedia...), temporal re-
sources (number of days, weeks or months), required knowledge (psychology, sociology, computer scientist...)

- About the potential performance of the evaluators: ability, experience, current workload...

The result (i.e. the choice) concerns:

- Evaluation criteria to consider in priority: social acceptability (ethical, social and cultural criteria), practical acceptability (environmental, economical, utility and usability criteria). Each criterion can be decomposed into sub-criteria. For instance, concerning the cultural dimension, in several countries, the system can give the advice to focus the evaluation on shapes or colours. About the usability criteria, sub-criteria are well defined in the literature: see for instance Bastien and Scapin [2].

- Evaluation methods to use in priority: many evaluation methods and variants exist. Each method can be described with several characteristics: name, concerned criteria, necessity of presence of users, and so on.

3.2 Software architecture

The ADHESION DSS is composed of two main modules: a first module automatically selects and proposes evaluation methods and criteria; the Command Module (CM) is used for the control of the system.

The user (which can be a project manager, a student in HCI or an unskilled evaluator) operates on the CM by using a specific human-computer interface (figure 2). The CM integrates the different elements intended to assist the human decision process. The user operates on the CM, in order to draw up a set of modifications necessary to compensate for unsatisfied suggestions. This first prototype is developed with the PrologIII language. The choice of this language is justified by its ability to manipulate objectives and subjectives data, to model constraints, and to reason on them.
3.3 Overall description of the approach based on Fuzzy logic

Our approach integrates Fuzzy rules and the concept of expert’s preferences (figure 3). One possibility could consist of working only with the Fuzzy rules, but the characteristics of our problem force us to use that combination. This way is encouraged when it is matter of a new or complex problem where well defined production rules do not exist [19]. The main stages followed in our approach are resumed in the figure 3, and progressively explained.

The modelling of expert’s preferences (aiming to build the knowledge base of the ADHESION DSS) can be illustrated with the following example. Let us consider the « human resources » criterion: to adjust this criterion, the expert prefers that another criterion (« available financial resources ») should be « rather good ». We know that the action concerned is « to adjust the human resources » (adjusting human resources consists in reducing or increasing the number of evaluators, seen as actors in the evaluation process) in relation with the criterion: « available financial resources ». The linguistic
variable « L » to be used is denoted « expert’s preferences » concerning the action « to adjust the human resources » in relation with the criterion: « available financial resources »; the expression which must be evaluated is « rather good ». In the literature, the term « good » is called a « describing term » and « rather » is called a « modifying term ».

The production rules to be used for the expression are:

1. S → T good
2. T → rather

Concerning the semantic to be associated to this linguistic variable, the variable is defined by a scale by 100, representing the percentage of the evaluators’ workload or the maximum of the budget that one can not exceed.

We must express what the expert means by « good » and « rather ». In this context, it becomes pertinent to use a Fuzzy approach [20], which, for example, can make easier the expression of the following expertise:

Between 60 and 100% of the overall budget allocated to the evaluation, the expert estimates that the criterion « available financial resources » should be « good » to adjust the criterion « human resources ». Thus, from 50 to 60% of the overall budget, the action « to adjust » becomes less possible. Under 50%, the adjustment becomes impossible.

A Fuzzy rule r_m can be written such as:

« IF C_1 is A_{1m} ... AND C_n is A_{nm} THEN a_j is DCF_m » (A_{im} and DCF_m are Fuzzy sets), that is to say that the C_i criteria respecting the A_{im} sets, we want to obtain DCF_m for the a_j action. »

Considering the same kind of example one can write:

IF the C_1 criterion « available financial resources » is 0.8 AND IF the C_2 criterion « presence of an expert in ergonomics » is 0 AND IF the C_3 criterion « dimension to be evaluated » = "usability" THEN the (a_j) action is « to adjust the human resources » is 0.9.

That can be read like that: if the criterion « available financial resources » represents 80% of the overall budget allocated to the evaluation and if there is no expert of ergonomics (0) in the evaluation team and if the criterion « dimension to be evaluated » is the usability, then the action « to adjust the human resources » is recommended at 90%. That is to say, in practice that the system will suggest to adjust the « human resources ». One can also note that the evaluation criteria and methods suggested by the DSS must be adapted to the types of evaluators present in the evaluation team. For instance, if the evaluators are computer scientists, it would be not pertinent to offer them evaluation methods requiring a strong ergonomic practice.
4. CASE STUDY: ADVICE ABOUT EVALUATIONS FOR THE SAME HCI TO BE USED IN TWO DIFFERENT COUNTRIES

To illustrate our approach, it is important to show (1) the impact of taking into account socio-cultural problems in certain countries, (2) the influence of situational constraints. Note that the system takes into account certain social and cultural differences concerning 212 countries and islands. This knowledge is issued from a collaboration with sociologists (IFRESI Laboratory).

For instance, in two different projects, we suppose that the same situational constraints characterize these projects concerning the same type of HCI to be used in different countries (Pakistan and France). In each case, the role of the user is first to describe the situation. The figure 4 is relating to the HCI to be used in Pakistan; we suppose that, for the HCI used in France, the only one difference is the selection of this country (left corner, at the top).

![Figure 4. Data describing the constraints of the evaluation situation for an HCI planned to be used in Pakistan](image)

On the screen (figure 4), one can distinguish a list of questions (written in French in this first version). The answers are not binary (yes/no): by using three « descriptors » (Bon/Oui – Moyen – Mauvais/Non; translation: Good/Yes – Middle – Bad/No) and three « modificators » (Très – Plutôt –
Modérément; translation: Very – rather – Moderately), the ADHESION DSS is able to propose solutions taking into account the uncertainty or the imprecision of the user. The figures 5 and 6 show parts of propositions (solutions) provided by the ADHESION DSS.

Figure 5. Propositions provided by the ADHESION DSS about the HCI planned to be used in France
In the two cases, it has proposed several evaluation criteria and methods. For the evaluation of the HCI to be used in Pakistan, figure 6, one can distinguish (left corner, in the bottom) that the DSS proposes to take into account some criteria concerning usability, but also criteria concerning social aspects. For the evaluation of the HCI to be used in France, the propositions are more focussed on usability aspects, figure 5. The evaluation methods proposed by the system can be different in the two cases, according to the evaluation criteria to consider. Note that one to five complementary evaluation methods are automatically proposed. Each criterion or method is also described; for instance, about a proposed method: (1) name, (2) development stage in which the method will be useful, (3) approximative duration required to perform the evaluation, (4) its availability (a method can be validated on very simple tasks only, and not directly usable in more complex situations), (5) references, (6) existence of a software version, and so on.

5. CONCLUSION AND PERSPECTIVES

Our approach can bring an original contribution to computer-aided evaluation of interactive systems. Indeed it becomes possible to assist automatically the choice of HCI evaluation criteria and methods adapted in each situation. A DSS called ADHESION is currently developed with the Prolog III language. The first results are encouraging in terms of expertise modelling and the solutions that it provides. The perspectives are numerous, according to: (1) the development of the software modules, (2) the expertise to integrate in its knowledge base, (3) the tests of this DSS, according to evaluation scenarios during real or simulated projects.

REFERENCES

20. Zadeh, L.A.: The concept of a linguistic variable and its application to approximative reasoning. *Information Sciences*, n° 8 (pp. 301-357) and n° 9 (pp. 43-80) (1975)