Expert Systems with Applications 38 (2011) 2110–2118

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A decision support system for usability evaluation of web-based information systems

Asil Oztekin *

School of Industrial Engineering and Management, Oklahoma State University, 322 Engineering North, Stillwater, OK 74078, USA

ARTICLE INFO

Keywords:
Web-based information system
Usability engineering
Machine learning
Sensitivity analysis
Criticality index

ABSTRACT

In this study, a decision support system (DSS) for usability assessment and design of web-based information systems (WIS) is proposed. It employs three machine learning methods (support vector machines, neural networks, and decision trees) and a statistical technique (multiple linear regression) to reveal the underlying relationships between the overall WIS usability and its determinative factors. A sensitivity analysis on the predictive models is performed and a new metric, *criticality index*, is devised to identify the importance ranking of the determinative factors. Checklist items with the highest and the lowest contribution to the usability performance of the WIS are specified by means of the criticality index. The most important usability problems for the WIS are determined with the help of a pseudo-Pareto analysis. A case study through a student information system at Fatih University is carried out to validate the proposed DSS. The proposed DSS can be used to decide which usability problems to focus on so as to improve the usability and quality of WIS.

1. Introduction

The web services provided by web-based information systems (WIS) have gained increasing importance in contemporary society. The users of WIS would like to find information in a fast and convenient way. Yet unfortunately, many WIS are still too slow to be usable and cannot satisfy many of their users. Experts from computer science/information science, usability/human–computer interaction, and requirements engineering areas try to solve web-based information system design problems (Yang & Tang, 2003). For measuring their service *quality*, the ServQual model (Parasuraman, Zeithmal, & Berry, 1988) and its modification for web-based information systems (Li, Tan, & Xɪe, 2002) are still the most widely used approaches. ServQual presents a survey instrument which claims to assess the service quality in any type of service organization (Parasuraman et al., 1988). The service quality is determined as the discrepancy between customers' expectations and perceptions for identifying dimensions that represent the evaluative criteria which customers use to assess service quality (Zeithaml, Parasuraman, & Berry, 1990). ServQual is used by a wide range of users including academicians and practitioners (Mei, Dean, & White, 1999). However, ServQual has also been criticized in some studies (e.g. Babakus & Boller, 1992; Buttle, 1996; Carman, 1990; Cronin & Taylor, 1992, 1994; Teas, 1993) because the development of good quality websites requires more sophisticated methods for design and assessment, and this development is fundamentally achieved by *usability* assessment studies (Frokjaer, Hertzum, & Hornbaek, 2000; Hornbaek, 2006; Li et al., 2002; Liu, Tucker, Koh, & Kappelman, 2003; Nikov, Vassileva, Anguelova, Tzvetanova, & Stoeva, 2003; Oztekin, Kong, & Uysal, 2010; Sauro & Kindlund, 2005). For measuring service *quality* the ServQual approach with 5-point distance semantic scale (or alternatively 7-point Likert scale) is the primary tool used (Parasuraman et al., 1988). To assess WIS quality, an enhanced version of ServQual, namely the web-based ServQual, with six dimensions measured by 28 checklist questions was developed by Li et al. (2002). However, neither of these approaches proposes a quantitative model to assess WIS quality. Based on ServQual, the WebQual approach evaluates the user perceptions of the quality of WIS (Barnes & Vidgen, 2003). It turns qualitative customer assessments into quantitative metrics for supporting management decision-making.

On the other hand, there are many other usability questionnaires/checklists such as Quis (Norman & Shneiderman, 1989), Sumi (Kirakowski & Corbett, 1993), PutQ (Lin, Choong, & Salvendy, 1997), PSSUQ (Lewis, 2002), and UseLearn (Oztekin et al., 2010). *Usability* refers to the extent to which a product can be used by specified users to achieve specified goals with efficiency, effectiveness, and satisfaction in a specified context of use (ISO 9241-11, 1998). Usability stands for the capability to be used by humans easily and effectively; how easy it is to find, understand and use the information displayed on a website (Keevil, 1998); and *quality in use* (Bevan, 1999).

* Tel.: +1 405 744 4664; fax: +1 405 744 4654.
E-mail address: asil.oztekin@okstate.edu

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.07.151

These definitions reveal that there has been an emerging need for a comprehensive methodology for measuring the usability of web-based information systems by integrating *quality*- and *usability*-related measures. It is anticipated that usability and quality do affect each other (Bevan, 1995, 1999). WebQual significantly tries to include usability dimensions in the assessment process. However, it does not mention the quality and usability measures in detail, and the names for the stated measures seem to be confusing. For example, it calls one dimension *usability*, but in fact this dimension is a combination of the several dimensions of other checklist approaches. Similarly, the *service interaction* dimension of WebQual is obviously a mixture of *integration of communication* (from ServQual) and *suitability for individualization* (from ISO 9241-10, 1996). Another modified approach based on ServQual is E-S-Qual, which assesses the quality of the websites in terms of *profitability* for the company (Parasuraman, Zeithmal, & Malhotra, 2005). Considering the fact that most of the usability and quality assessment approaches have many overlapping items in their checklists, recently Oztekin, Nikov, and Zaim (2009) proposed a methodology (UWIS) which includes both quality and usability dimensions. UWIS methodology is an extended form of ServQual that measures the usability of web-based information systems by including the dialog principles for user interface design according to the standard ISO 9241-10 (ISO 9241-10, 1996) and usability heuristics (Nielsen, 1994). UWIS methodology proposes a broader approach applicable to both non-profit and profit-oriented web-based information systems. Due to its broad applicability, the UWIS checklist will also be used in this study.

Discussions on how to measure the quality of information systems have gone on for several decades, first in the areas of ergonomics, ease-of-use, and human–computer interaction and later in the area of usability. However, recently discussions recur on which measures of usability are suitable and on how to understand the relation between different measures of usability (Hornbaek, 2006). To increase the meaningfulness and strategic influence of usability data, the entire construct of usability can be presented as a single dependent variable (*usability index*) without sacrificing precision (Sauro & Kindlund, 2005). The usability index is a measure of how closely the features of a website match generally accepted usability guidelines (Keevil, 1998). This dependent variable can be explained by checklist items, namely the independent variables, and hence their cause-and-effect relationship can be explained. For example: if the button to change the password in a system is not visible enough (i.e. not located at a visible place on the webpage), it would take a long time for a user to find it, select it, and change his/her password. This would decrease the efficiency of the system in terms of speed; hence indirectly the *overall usability* of the system will also be decreased.

2. The proposed decision support system

In Section 1, the usability evaluation studies for web-based information systems are summarized. The fundamental limitation they have is that they merely administer Likert-type checklists to the end-users (usability testing) or alternatively adopt a heuristic evaluation by means of usability experts. They do not provide the usability experts with an analytical foundation or numerical evidence which would rank the emerging usability items in terms of their importance for further improvement and remedy. These approaches strongly rely on the evaluation results provided by a representative sample pool of intended end-users or domain experts and rank their evaluation results based on the *mean scores* received for each checklist dimension. What they mainly ignore is that whether or not improving a particular usability problem would eventually affect the usability perception of the end-users a great deal. In other words, they reveal the problematic usability aspects of the WIS system by suggesting that the smaller the survey-based evaluation average score, the more important the checklist item is, and hence the priority should be given to it for improvement. Yet, they ignore the effect of one unit change of this particular checklist item on the final usability perception, namely overall usability. Although one usability checklist item is fairly low (evaluated badly by the end-users or usability experts), changing/improving it may or may not make a big improvement on the final usability of the WIS analyzed. Since time and efforts should be focused only on *worthy* usability problems that would make a significant difference on the usability at the end, both criteria should be considered in the usability evaluation process. Therefore, in this study I define a combined metric, *criticality index*, which takes both the measures into account and can be calculated as in Eq. (1),

$$\text{Criticality Index} = \text{Sensitivity Score} * \frac{1}{\text{Average of checklist evaluation scores}} \tag{1}$$

where sensitivity score refers to the importance of the checklist item in explaining the cause-and-effect relationship between the dependent variable (overall usability) and independent variables (checklist items) and will be explained in Section 2.2.3 in detail. The average of checklist evaluation scores, on the other hand, is straightforward, but the reason to take the reciprocal of it requires further explanation and is given as follows. The smaller this value, the more important the checklist item is. However, the bigger the sensitivity score, the more important the checklist item is. To make these two contradictory values comparable and combined in one metric (namely in the *criticality index*) it is proposed to use the inverse of the average of checklist evaluation score. The decision support system (DSS) along with the use of UWIS checklist is illustrated in Fig. 1.

The first step in the proposed methodology is to collect sample data from representative end-users to be able to apply the usability testing process. The data is collected through UWIS checklist developed by Oztekin et al. (2009). Table 1 briefly presents the UWIS checklist items and their corresponding questions. For further usage throughout the paper, I refer to the checklist questions/items as the input variables used to predict the overall usability of the WIS by using their abbreviated symbols such as RL1, RL2, and RL3 for the first, second, and third questions related to the "reliability" dimension of the UWIS checklist.

The output variable (overall usability of the WIS system) is represented as the last question in this checklist. I used a 5-point Likert scale, with anchors ranging from strongly disagree (1) to strongly agree (5). Therefore, when deciding whether or not the usability score is satisfactory, I adopted the Likert score of 4 (equivalent to "agree") as a threshold. In other words, the checklist question measuring the overall usability of the WIS system should be rated 4 or greater in average to be able to deem the overall WIS usability "satisfactory". If so, the end-users are assumed to find the system usable enough and hence, no further analysis is required. Otherwise, the evaluation process proceeds as follows. Since the complex relationship between the output variable and the input variables is not known apriori, the next step is to find the best predictive model that explains this potentially complex relationship considering various performance measures (i.e. error rates and correlation). Then, the *criticality index* is calculated for each input variable, namely each checklist item, to rank them in descending order. This ranking would help determine which variables to improve first with scarce time and money constraints. A pseudo-Pareto chart would be useful to pick only the most effective checklist items and to determine the corresponding usability problems of the WIS. The Pareto rule (Pareto, 1971) basically