Evaluation of Adaptive Systems

Stephan Weibelzahl

Pedagogical University Freiburg, Germany
weibelza@ph-freiburg.de

Abstract. Unambiguously, adaptive systems have to be evaluated empirically to guarantee that the adaptivity really works. Nevertheless, only few of the existing adaptive systems have been evaluated. One of the most important reasons for this lack is, that measures for adaptivity success have not been investigated systematically up to now.
The aim of this PhD thesis is to explore a methodology for the empirical evaluation of adaptive systems, including validated criteria, experimental designs and procedures. It will be demonstrated that cognitive and behavioral factors provide important evidence for adaptivity success.

1 Introduction

Empirical evaluations of adaptive systems are hard to find. Several reasons have been identified to be responsible for this lack (e.g., Höök [1]). Besides some structural problems (e.g., short development cycle) one of the major issues is methodological: What has to be done to guarantee the success of adaptation? Straightforward approaches (e.g., asking the users whether they enjoyed the interaction) frequently fail to proof an advantage of adaptive systems or suffer from low test quality.

2 Aim

The aim of this PhD thesis is to explore a methodology for the empirical evaluation of adaptive systems. Such a methodology consists of at least two components: First, a bunch of criteria that are proofed to be reliable and valid to measure adaptivity success. Probably, only a combination of several criteria will be adequately meaningful. Secondly, a specification of experimental designs and procedures is needed to apply those criteria.

Both, criteria and experimental designs, should be independent of the domain and the inference mechanism, because this would allow for a comparison of different approaches. We certainly do not ignore the fact, that there are domain specific differences between systems.

That is, there are criteria that evaluate system specific goals. However, our approach explores those goals that all (or most) adaptive systems have in common: to improve and to simplify interaction.

We claim that such a general approach yields a methodology that is transferable to many systems and would enable researchers to

– find system deficits and failures
– show that adaptivity in their system is useful and successful
– justify the enormous efforts spent on making systems adaptive
– point out deficits of non-adaptive systems.

3 Approach

A synopsis on published evaluations (http://www.softwareevaluation.de) affirmed our claim that only few studies are based on proper experimental designs and statistical methods. We found that evaluation issues can be categorized in accordance with an information processing model of adaptive systems. This model serves as evaluation framework for further considerations.

3.1 Model of Adaptivity

The basic principle of adaptive systems is, that they acquire data about the user, which are used to infer abstract characteristics. Based on these characteristics (sometimes called user properties) the system decides how to adapt, which has impact on both the system behavior and the user behavior [2]. Thus, we distinguish six evaluation steps [6]:

1. Evaluation of reliability and external validity of input data acquisition
2. Evaluation of the inference mechanism and accuracy of user properties
3. Appropriateness of adaptation decisions
4. Change of system behavior when the system adapts
5. Change of user behavior when system adapts
6. Change and quality of total interaction

The point is that this last evaluation step can only be interpreted correctly if all the previous steps have been completed. Especially in the case of finding no difference between an adaptive and a non-adaptive system the previous steps provide hints at shortcomings. The results of such a layered evaluation [3] are much better to interpret and give more exact hints for failures and false inferences than a simple global evaluation of the usability.

3.2 New Criterion: Behavioral Complexity

An evaluation methodology also requires criteria for adaptation success, however, our systematic overview made obvious that most current criteria are either not reliable (or at least not tested to be reliable) or are low in external validity (e.g, duration of interaction as a criterion for satisfaction with a learning environment).

Our approach is based on the assumption that adaptivity aims at reducing the complexity of interaction. E.g., an adaptive help system infers the user's current goals and selects adequate help texts. Thus, the user does not need an overview of all possibly presented themes and then to select the most suitable section. The same argumentation holds for other adaptive systems, e.g., product presentation systems or learning environments.

According to Kieras and Polson [4] the concrete behavior of a user can be described as a state transition network. According to this idea, a system is defined by disjoint states and the transitions between these states. A user's input equals a transition to an other state.

A single network is derived for each user. The complexity of these networks can be computed in several ways. In a series of laboratory experiments Weibelzahl, Klein, and Weber [5] compared four measures of complexity in two different systems (which implemented different inference mechanisms in different domains). Users that had been supported by an adaptive system produced behavior of reduced complexity compared to users that completed the same task ("Find a suitable vacation home in this electronic catalog"; "Complete this chapter in the online-learning-course") with a non-adaptive version.

4 Future Perspectives

As the experimental results with behavioral complexity are very encouraging, a more detailed evaluation of an adaptive learning environment was designed to validate this criterion.

The online-course HTML-Tutor will be tested both in laboratory and in the field which will explore the external validity of behavioral complexity and the scope of our framework.

References

1. K. Höök. Steps to take before intelligent user interfaces become real. *Interacting With Computers*, 12(4):409–426, 2000.
2. A. Jameson. *Systems That Adapt to Their Users: An Integrative Perspective*. Saarland University, Saarbrücken, 2001.
3. C. Karagiannidis and D. G. Sampson. Layered evaluation of adaptive applications and services. In P. Brusilovsky and C. S. Oliviero Stock, editors, *Proceedings on International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, AH2000, Trento, Italy*, pages 343–346, Berlin, 2000. Springer.
4. D. E. Kieras and P. Polson. An approach to the formal analysis of user complexity. *International Journal of Human-Computer Studies*, 51(2):405–434, 1999.
5. S. Weibelzahl, B. Klein, and G. Weber. Criteria for the evaluation of adaptive systems. *User Modeling and User Adapted Interaction*, submitted.
6. S. Weibelzahl and C. U. Lauer. Framework for the evaluation of adaptive CBR-systems. In I. Vollrath, S. Schmitt, and U. Reimer, editors, *Experience Management as Reuse of Knowledge. Proceedings of the 9th German Workshop on Case Based Reasoning, GWCBR2001*, pages 254–263, Baden-Baden, Germany, 2001.