Chapter XV

The Evaluation of Adaptive Systems

Judith Masthoff

University of Brighton, UK

This chapter discusses the evaluation of adaptive systems, as it is important to investigate the effectiveness of adaptation and how it may be optimised. It argues that evaluation is often neglected because of difficulties in attributing cause, defining effectiveness, obtaining significant results and getting resources. Using example evaluations of an interactive instruction system, it shows how these issues can be tackled by separating concerns, using learning curves, splitting subjects into groups, and using experts and models instead of subjects. The author hopes that these methods will be more widely adopted for evaluating adaptive systems (not just adaptive instruction systems), producing more insight in experimental data and its causes, delivering more significant results, and making evaluations less costly.

INTRODUCTION

Evaluation of their effectiveness should be an integral part of research in and development of adaptive systems. However, it is often neglected due to many problems associated with the evaluation of adaptive systems. In this chapter, we will

explore the issues and show how we have tackled them in our research on adaptive instruction systems. Evaluations we have conducted in the past will be used as illustrations.

BACKGROUND

Empirical evaluation of adaptive systems has been rather limited. Less than one out of four articles published in *User Modeling and User Adapted Interaction* reports significant empirical evaluations (Chin, 2001). Similar figures seem to hold for other journals and conferences in the field: for instance, less than 25% of the full papers presented at the 2002 Intelligent User Interface Conference (Gill & Leake, 2002) contained an empirical evaluation. Similarly, evaluation of adaptive hypermedia systems has been very limited (Eklund & Brusilovsky, 1998). Researchers are well aware of this problem, and it led to the 2001 Workshop on Empirical Evaluations of Adaptive Systems (Weibelzahl, Chin, & Weber, 2001; after already having been a main focus of the Second Workshop on Adaptive Hypertext and Hypermedia, 1998).

The research reported in the present chapter puts evaluation at center stage. Our work focuses on adaptation in the domain of interactive instruction, producing an artificial teacher that can adapt instruction to the student (Masthoff, 1997). Our teacher is based on a community of autonomous software agents, each representing a role of a teacher, like giving feedback, navigating through course material, tailoring explanations, etc. (see, Masthoff, in press, for the design and architecture of this teacher). Each agent contains a set of behaviours, specifying how the agent should act in certain situations. These behaviors were as much as possible based on literature on human learning. A vital part of the research was to show that the artificial teacher's adaptive behaviour did indeed benefit the students. Additionally, quite often we had to decide which behavior the teacher should adopt from a set of alternatives. We wanted to base these decisions on empirical evidence regarding their relative effectiveness.

ISSUES

There are many reasons why the empirical evaluation of adaptive systems has been so limited. We encountered the following when evaluating our own system.

Difficulty in Attributing Cause

Adaptivity is only one contributing factor to the usability of a system. So, if an adaptive system is evaluated as being "effective" (or usable, satisfying, efficient,

Related Content

Recurrence Indicators for the Estimation of Characteristic Size and Frequency of Spatial Patterns

Chiara Mocenniand Angelo Facchini (2013). *Complexity Science, Living Systems, and Reflexing Interfaces: New Models and Perspectives (pp. 209-217).*

www.irma-international.org/chapter/recurrence-indicators-estimation-characteristic-size/69463

Appropriate Evolutionary Algorithm for Scheduling in FMS

B. B. Choudhury, B. B. Biswal, D. Mishraand R. N. Mahapatra (2011). *International Journal of Applied Evolutionary Computation (pp. 15-26).*

www.irma-international.org/article/appropriate-evolutionary-algorithm-scheduling-fms/55961

Ways to View the World: A Standard Ontology as the Reality Framework and the World Code

Azamat Abdoullaev (2008). *Reality, Universal Ontology and Knowledge Systems: Toward the Intelligent World (pp. 28-57).*

www.irma-international.org/chapter/ways-view-world/28309

Design of a Functional H Filter for Linear Singular Systems With an Additional Unknown Input, Bounded Disturbance, and Variable Time Delay

Hamzaoui Fatma, Khadhraoui Malekand Messaoud Hassani (2020). *International Journal of System Dynamics Applications (pp. 24-54).*

www.irma-international.org/article/design-of-a-functional-h-filter-for-linear-singular-systems-with-an-additional-unknown-input-bounded-disturbance-and-variable-time-delay/266462

Mixing Workflows and Components to Support Evolving Services

Françoise Baude, Virginie Legrand, Ludovic Henrio, Paul Naoumenko, Heiko Pfeffer, Louay Bassboussand David Linner (2012). *Technological Innovations in Adaptive and Dependable Systems: Advancing Models and Concepts (pp. 320-344).*

www.irma-international.org/chapter/mixing-workflows-components-support-evolving/63589