



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,

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/evaluation-adaptive-systems/4226

Related Content

Efficient and Fine-Grained VMM-Level Packet Filtering for Self-Protection

Kenichi Kourai, Takeshi Azumi and Shigeru Chiba (2014). *International Journal of Adaptive, Resilient and Autonomic Systems* (pp. 83-100).

www.irma-international.org/article/efficient-and-fine-grained-vmm-level-packet-filtering-for-self-protection/118287

Impact of Cross-Layer Adaptations of Mobile IP on IEEE 802.11 Networks on Video Streaming

P. De Cleyn and C. Blondia (2012). *Technological Innovations in Adaptive and Dependable Systems: Advancing Models and Concepts* (pp. 261-281).

www.irma-international.org/chapter/impact-cross-layer-adaptations-mobile/63586

An Architecture-Based Adaptation Framework for Soft Real-Time Applications

Ning Gui, Hong Sun and Chris Blondia (2010). *International Journal of Adaptive, Resilient and Autonomic Systems* (pp. 12-25).

www.irma-international.org/article/architecture-based-adaptation-framework-soft/48260

Mobile Computing and Mixed-Initiative Support for Writing Competence

Vive Kumar, Maiga Chang and Tracey L. Leacock (2012). *Intelligent and Adaptive Learning Systems: Technology Enhanced Support for Learners and Teachers* (pp. 327-341).

www.irma-international.org/chapter/mobile-computing-mixed-initiative-support/56089

Processing Combinatorial Thinking: Innovators Marketplace as Role-Based Game Plus Action Planning

Teruaki Hayashi and Yukio Ohsawa (2013). *International Journal of Knowledge and Systems Science* (pp. 14-38).

www.irma-international.org/article/processing-combinatorial-thinking/101153