Validating the Evaluation of Adaptive Systems by User Profile Simulation

Javier Bravo and Alvaro Ortigosa

Escuela Politécnica Superior, Universidad Autónoma de Madrid, Spain
{javier.bravo, alvaro.ortigosa}@uam.es **

Abstract. Availability of automated tools and methods to evaluate adaptive systems is a fundamental requirement to promote a wider adoption of these systems. An obstacle in this direction is the difficulty of validating any automatic tool created to help on the evaluation of adaptive systems. In this work a simulation-based technique is proposed as an economic way for testing evaluation tools based on log analysis. Simulog, a tool that implements this simulation technique, is also presented.

1 Introduction

Evaluation of adaptive hypermedia systems is a difficult task, as it is shown by many studies (Weibelzahl et al. [1], Missier et al. [2], Markham et al. [3], and Lavie et al. [4], among others). This is also true when considering Adaptive Educational Hypermedia (AEH). Particularly, evaluating if an AEH system is making good adaptation decisions for a given student profile is a complex task, which can be improved by using automatic tool support. For example, an evaluation tool can help to identify possible failures in the adaptation rules for a course, and even it can suggest corrections oriented to improve these rules [5]. On this way, an automatic tool would allow the course designer to save time, since it does not require a manual analysis of the data. Besides, this tool may support a continuous evaluation of the adaptation effectiveness, not limited to a fixed frame of time [6].

We are specially interested in evaluation tools based on analysis of user log files. These tools, by using techniques and methods from Data Mining, can search through the records of users actions, looking for unexpected behavioral patterns. In this way unusual patterns, called **anomalies** on this work, can hint about possible failures in the adaptation rules. For example, a simple anomaly can be a number of students failing to pass the test associated with a given educational activity. This situation can suggest that there is a problem with the adaptation rules applied to this group of students.

A possible drawback of using this type of evaluation tools is that they need to be validated. In this sense, the main problem is to check if the tool is able to find the anomalies contained in a set of log files. If the tool is based on statistical

** This project has been partially funded by the MEC, Spain (TIN2004-03140).

methods for searching anomalies, a significant number of students exhibiting the unusual behavior should be processed before detecting a potential problem. This produces two difficulties: a large set of student logs is needed to test the tool and, still worse, the logs should be manually analyzed in order to contrast the results achieved by the tool.

A more practical alternative would be using simulation methods. Specifically, we propose to use a tool to simulate the behavior of students by generating logs which are similar to the logs produced by real students using the AEH system. For example, it is possible to generate data containing an anomaly, consequence of a hypothetical problem in the AEH system decisions. Then, the evaluation tool can be fed with the generated data and to test if it is able to find the anomaly in the data and suggest possible causes. In this way, the evaluator has control about the data provided to the evaluation tool and can anticipate the expected output from it. In this work we present a tool, named *Simulog* (Simulation of User Logs), designed to simulate user behaviors represented as log files.

2 Tool Proposal: Simulog

Figure 1 shows the context of *Simulog*. The tool was designed with the intention that it could be used to test any evaluation tool based on log analysis. In this way, both its inputs and output were designed for maximum generality.

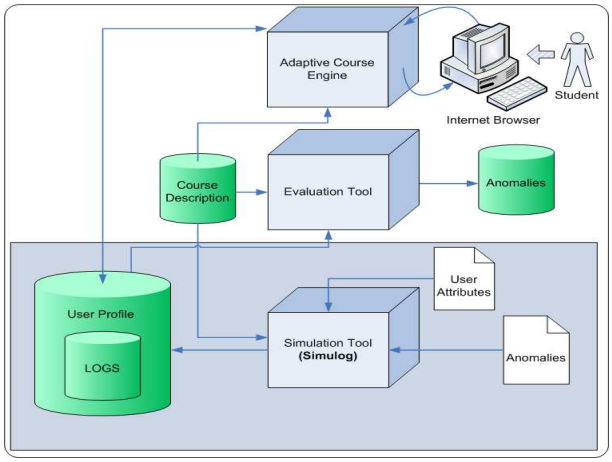

Fig. 1. Architectural context for the evaluation and simulation tools

2.1 Inputs

Simulog simulates student behavior according to certain parameters. It receives the course description, the user attributes, and the anomalies. The user at-

tributes define the profile of the simulated students. For example, it can be specified that 20% of the students represented through the logs are textual learners, while 80% are visual learners, and that the average knowledge level of the students is 7 out of 10. Next subsection explains how statistical methods are used to generate user profiles according to these distributions.

The structure of course is described using the **SCORM** standard. Using the XML-based SCORM specification, the course description is independent of the target evaluation tool and adaptive system.

An important issue is to be able to specify the anomalies using an abstract notation. With this goal, the adopted solution is to use ontologies. Ontology is a formal concept of the Semantic Web field [7,8] that provides a shared common knowledge with the possibility of making inference about this knowledge. In that way, an *ontology of anomalies* was created in order to support the specification of anomalies, that is, the unusual patterns that will included in the log files and that are supposed to be detected by the evaluation tool.

Anomalies are classified in two groups (see figure 2): quantitative and qualitative. Quantitative anomalies reference to measurable situations by means of the criteria of measurement presented by Whiteside (1988) [9]. The second group, qualitative anomalies, makes reference to non-measurable criteria, as for example the situation when *student get lost navigating through the course.*

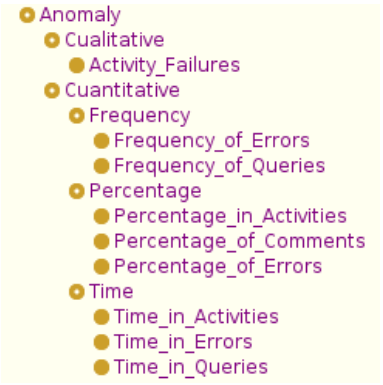

Fig. 2. Classes of the ontology of anomalies

2.2 Simulation engine

The engine has a double function: generating the user models (that is, the set of attributes describing the features of a given student) and generating the log files. Together both of them constitute the user profile. The engine generates the sample with the proportions of attributes (described in the previous section) using the Normal distribution. Two conditions are fulfilled in this distribution: representativeness and randomness; without these conditions the simulation would

not generate reliable data. Besides, the Normal distribution is used in techniques of people simulation, because it produces results next to the real ones. In fact, many parameters of the nature are distributed following a Normal distribution, like for example the length of an arm, height of an individual, etc. [10].

In order to generate the correct proportion of users, some calculations have to be done. For example, if regarding the visual-verbal dimension the goal is to generate 75% of students with visual preferences and 25% with textual preferences, the random numbers must be generated following a Normal distribution[1] with $\mu = 3.66$ and $\sigma = 2$. Applying this method for each user model dimension it is possible to generate a set of profiles containing the distribution of students features specified by the *Simulog* user.

The simulated actions are also generated using a probabilistic model. For example, the probability of success on a question, of leaving an activity before completion, etc., is conditional to the criteria selected by the tool user.

2.3 Generation of Log Files

The structure of log files is composed by a user model, logs of activities, and logs of updates. Figure 3 shows an example of a partial log file. The user model stores the information about the student, in the form <attribute,value>. The logs of updates store, in similar format, every actualization of the user model.

```
<log-root>
  <user-model>
        . . .
  </user-model>
  <log-activity>
       <log activity="activity 1" score="0.0"
            action="START-ACTIVITY" timestamp="2006-03-26T09:52:00"/>
       <log activity="activity 1-1" score="0.3"
            action="START-ACTIVITY" timestamp="2006-03-26T09:54:15"/>
       <log activity="activity 1-1" score="0.5"
            action="END-ACTIVITY" timestamp="2006-03-26T10:02:55"/>
  </log-activity>
  <log-update>
        . . .
  </log-update>
</log-root>
```

Fig. 3. Example of a partial log file

3 Related Work

This work presents a novel approach for simulating user behavior by generating log files. This approach and the resulting tool can be used to validate tools

[1] The statistical explanation of this result can be found in [10, p.253-255].

and methods reported in other works as, for example, Börner (2001) [11] and Brusilovsky et al. (2001) [12]. Regarding similar uses of ontologies, Stetson et al. used them for describing anomalies in the Medicine area [13].

4 Conclusions and Future Work

This paper proposes a practical way for testing evaluation tools designed for evaluating AEH systems. *Simulog*, a simulation tool developed to implement this approach, is also presented. However, there still remains much work to be done, mainly regarding the use of Simulog for testing real evaluation tools. Planed future work also includes adding new features to the simulation tool, such as the capability of generating profiles using other types of well known distributions, like T-Student or F-Snedecor. Besides, it would be desirable to define the ontology using OWL (Web Ontology Language) or F-logic (the current version is in RDF). Finally, we plan to create an ontology for adaptive course description.

References

1. S. Weibelzahl and G. Weber. Advantages, opportunities, and limits of empirical evaluations: Evaluating adaptive systems. *Künstliche Intelligenz*, 3/02:17–20, 2002.
2. F. Del Missier and F. Ricci. Understanding recommender systems: Experimental evaluation challenges. *User Modeling: Workshop of EEAS*, pages 31–40, June 2003.
3. S. Markham, J. Ceddia, J. Sheard, C. Burvill, J. Weir, B. Field, L. Sterling, and L. Stern. Applying agent technology to evaluation tasks in e-learning environments. *Exploring Educational Technologies - from Strategy to Implementation*, 2003.
4. T. Lavie, J. Meyer, K. Beugler, and J. F. Coughlin. The evaluation of in-vehicle adaptive systems. *User Modeling: Workshop on the EAS*, pages 9–18, July 2005.
5. A. Ortigosa and R. M. Carro. The continuous empirical evaluation approach: Evaluating adaptive web-based courses. *User Modeling 2003 (Lecture Notes in Computer Science, Volume 2702 / 2003)*, pages 163–167, June 2003.
6. A. Ortigosa and R. M. Carro. Continuous evaluation of adaptive web/based courses. volume 4. IEEE Computer Society Learning Technology Task Force, 2002.
7. N. Noy and D. McGuinness. Ontology development 101: A guide to creating your first ontology. March 2001.
8. T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web: A new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, May 2001.
9. J. Whiteside. *Usability Engineering: Our Experience and Evolution.* M. Helander (Ed.) New York: North-Holland, 1988.
10. M.H. DeGroot. *Probability and Statistics.* Addison-Wesley Ed., 1986.
11. K. Börner. Adaptation and evaluation of 3-dimensional collaborative information visualizations. *User Modeling: Workshop Proc. of EEAS*, pages 33–40, July 2001.
12. P. Brusilovsky, C. Karagiannidis, and D. Sampson. The benefits of layered evaluation of adaptive applications and services. *User Modeling: Workshop Proc. of EEAS*, pages 1–8, July 2001.
13. P. Stetson, L. McKnight, S. Bakken, C. Curran, T. Kubose, and J. Cimino. Development of an ontology to model medical errors, information needs, and the clinical communication space. *Proc AMIA Symp*, 2001.