Performance Evaluation MetricsofAdaptive Educational Hypermedia System (AEHS)

Dr. yassir, M., Hajhamad
Faculty of Computer Science and Information Technology
University of karary, Sudan.

Dr. Nafisa, A., Hassan
Faculty of Computer Science and Information Technology
University of West Kordofan, Elnihood, Sudan.

Abstract:

Education is one of the important processes in our life and human can't abandon it. Recently, E-learning and distance learning is the trend of the educational process. Adaptive learning technique is a new field of research in the field of hypermedia and adaptive systems. Performance Evalution Metrics of Adaptive Educational Hypermedia System (AEHS) is a challenging application area for developing and assess the effect of adapting educational materials individualized to student's needs quantitatively,due to the complexity of these systems (AEHS). Actually,an adaptive system refers to a system which tailors its output, using implicit interfaces based on interaction with the user. In this paper we attempt to highlightthe importance of the adaptive educational and the evaluation metrics that use to measure this kind of adaptive educational system.Finally, empirical evaluation requirements were also covered. The finding support the use of descriptive statistics and inferential statistcs to measure student's academic achievements in AEHS.

Keywords: evaluation metrics for (AEHS), Performance, adaptive educational hypermedia system (AEHS).

مستخلص:

التعليم هو احدي العمليات المهمة في حياتنا، ولا يمكن للانسان التخلي عنه. في الاونة الاخيرة، التعليم الالكتروني والتعليم عن بعد اصبحا هم الاتجاه للعملية التعليمية. تعد تقنية (التعلم التكيفي) مجالا جديدا للبحث في مجال الوسائط التشعبية والانظمة التكيفية. مقاييس تقييم الاداء لنظام الوسائط التشعبية التعليمية التكيفية (AEHS) هو مجال صعب التطبيق لتطوير وتقييم وتأثير تكييف المواد التعلمية للاحتياجات الفردية الطلاب من الناحية الكمية، بسبب تعقيد هذه الانظمة ((AEHS. في الواقع يشير النظام التكيفي الي نظام يقوم بتكييف مخرجاته، باستخدام واجهات ضمنية تعتمد علي التفاعل مع المستخدم. نحاول في هذا البحث إبراز أهمية المقاييس التعليمية والتقويمية التكيفية التي تستخدم لقياس هذا النوع من النظام التعليمي التكيفي. وتم اخيرا في هذا البحث ايضا، تغطية متطلبات التقييم التجريبي.تدعم النتيجة استخدام الاحصاء الوصفي والاحصاءات الاستنتاجية لقياس الانجازات الأكاديمية للطالب في نظام الوسائط التشعبية التعليمية التكيفية.

المفتاح الاساسي: مقاييس تقييم الاداء لنظام الوسائط التشعبية التعليمية التكيفية , الاداء, نظام الوسائط التشعبية التعليمية التكيفية.

1. Introduction

A current problem with the research of adaptive systems is the inconsistency of evaluation applied to the adaptive educational hypermedia systems (AEHS). However, evaluating an adaptive system is a difficult task due to the complexity of such systems. Evaluators need to ensure correct evaluation methods and measurement metrics are used [1].

Evaluation is defined as the process of examining the product, system components, or design, to determine its usability, functionality and acceptability which is measured in terms of a number of criteria essential for any softwaredevelopment project. Evaluation of all systems is important. It is important to not only evaluate but also to ensure that the evaluation uses the correct method [2].

Adaptive Educational Hypermedia Systems (AEHS) have been proposed as the underlying facilitator for personalized web-based learning with the general aim of generating and providing personalized learning experiences to an individual learner [1,

2]. According to [4] Adaptive Educational Hypermedia Systems (AEHS) have been proposed as the underlying facilitator for personalized web-based learning with the general aim of generating and providing personalized learning experiences to an individual learner.

In order to the evaluation of AEHS has long been acknowledged as a difficult, complicated and very demanding endeavour due to the complex nature of these systems. Therefore, evaluation is an important tool in software quality assurance.Evaluation of all systems is important, not only to evaluate but also toensure that the evaluation uses the correct methods and metrics[3].

In this paper, we present our performance evaluation metrics for measuring the use of this approaches in AEHS. The main factor to provide adaptivity in the AEHS is the student model that represents relevant aspects of the student such as preferences, knowledge and interests The student model dynamically maintains information for each user such as his/her knowledge, preferences, etc. The system collects this student information by observing the use of the application, by presenting series of questionnaires or feedback forms [12].

The paper is structured as follows: Section 2 presents the related works for AEHS; Section3 provide the metrics for evaluating the performance in AEHS. Finally, the empirical evaluation requirements were also covered and discuss our findings and theconclusions that can be offered.

2. Related work

A number of pioneer AEHS were developed between 1990 and 1996,one of the most interesting works in this area is the ELM-ART tutoring system that supports learning of the programming language LISP [5].

- INTERBOOK is a system for authoring and delivering adaptive electronic textbooks on the Web. All INTERBOOK-

served electronic textbooks have generated table of content, a glossary and a search interface. The online books in the same way as ELM-ART use colored bullet annotation to inform the user about the status of the node behind the link [5].

TANGOW structures Web courses by means of teaching tasks and rules. It differs from ELM-ART in that uses a dynamic tree to restrict the set of teaching tasks to be reviewed. This is achieved by including in each dynamic generated page only those subtasks (fragments), which are considered to be relevant by the system at run-time. In addition, rules are used to analyze prerequisite conditions [5].

KBS Hyperbook is another goal-driven approach that uses a Bayesian network technique for its user model [5].

SmexWeb is a framework that permits the development of teaching applications through the instantiation of a collection of abstract and concrete classes. Similar to TANGOW the authoring process consists mainly of the definition of concepts (tasks) and adaptation rules.

All types of adaptation are supported by SmexWeb: adaptive content, adaptive navigation, adaptive presentation and passive navigation (Albrecht, Koch & Tiller, 1999).

AHA or Adaptive Hypermedia Applications is a generic hypermedia system based on the adaptation of pages using conditional fragments. The structure of the domain is similar to the SmexWeb structure. Concepts are related to other concepts through weighted links [6].

ISIS-Tutor system uses different forms of adaptive navigation, such as direct guidance, hiding and annotation. The goal is to highlight the links corresponding to the student's goal and to hide concepts that belong to future learning targets [7].

The Dynamic Course Generation (DCG) represents a quite different approach. It consists of a concept domain structure rep-

resented as a plan, which relates known concepts for the learner with the goal-concept of the course. The plan is then adapted dynamically according to the student's learning progress. This results in changes to the subtasks and steps the learner has to follow proposed by Vassileva 1997[5].

These systems provide specific navigation aid, Selection of content,metadata cognition, provide results of the learning style tests, change learning explicitly and provide scrutability, reusability, provision of relevant learning materials which are personalized to specific learner, the efficiencyof the AEHS systems which are user specific, Student motivation, Avoidance of information overload, Monitory and temporal and spatial relevance [3].

Evaluation methods of Adaptive Educational Hypermedia Systems

The evaluation processis an important step to any system;it should ensure the correct methods were used. The evaluation of learner and tutor feedback is essential in the production of high quality personalized Technology Enhanced Learning (TEL) services. The evaluation focuses on the technological design and performance of systems without justifying the designs through the lessons learned from evaluations. To evaluate adaptive hypermedia systems, there are many methods that used in traditional software systems such as ISO/IEC 9126 1991.

These measures define many sub factors to be evaluated to measure the system qualities, such as:

Accessibility: express the facility to reach the nodes.

Adaptability: is the facility of an application to be configurable according to a set of decisions taken by the user, which usually define her preferences and/or background.

Adaptivity: denotes the capacity of the application to alter the user model according to the user behaviour during the application run and adapt dynamically to the current state of the user model.

Assistance: measures the amount of help in the form of additional information or link annotations is offered by the application to the user.

Availability: indicates whether the content is updated, and whether information obtained e.g. from a database is always accessible.

Completeness: measures the content for missing information and the structure for missing and dangling links.

Consistency: measures the regularity of the application, i.e. similar treatment of similar aspects (at content, navigation and presentation level) and clear differences for nodes with different content, for different access structures, for different types of navigation or differences in the layout. This is considered to be the most important evaluation criteria, although it is difficult to define what a consistent hypermedia application is.

Functionality: indicates how the application functions satisfy the users.

1. Implement ability: defines the overhead to providing adaptive features.
2. Maintainability: defines the effort needed to make specified modifications.
3. Performance: expresses the system's response time to user interaction as well as the amount of resources used by the system under stated conditions.
4. Predictability: measures how easily the user can guess the reaction of the system to her interaction.
5. Portability: indicates the ability of the software to be transferred from one environment to another.
6. Reliability: measures number of crashes resulting for e.g. from SQL or JavaScript error messages or too many hits during peak periods of Web use.

7. Reuse: defines the percentage of elements that are used for more than one purpose within the same application or in different applications. In hypermedia systems reuse means use of objects in different contexts, use of the same interface objects or navigation elements for different nodes. Reuse promotes consistency, accessibility and predictability.
8. Richness: denotes the amount of information nodes contained in the application.
9. Satisfaction: shows the user's subjective impression of the adaptive system.
10. Self-evidence: expresses how well the user can guess the meaning of the visualised content or the navigation elements.
11. Usability: measures the effort the user needs to use the system and individual assessment of such use.
12. User-retention-overtime: indicates how long the user remains using the application. Studies to measure one or more of these criteria usually compare user's handle an adaptive system and its non-adaptive variant [5].
13. Other criteria are specific to hypermedia systems, such as the criteria related to nodes and links or specific to adaptive systems it can also apply to adaptive tutoring systems (AEHS) which is a wide area of AHS such as Adaptivity, adaptability and Consistency, consistence improves quality in the same way as consistency is responsible for the success of a teaching book.
14. Generally, evaluation of anadaptiveweb systemcan be divided into three types of evaluation; the first one is a Formal evaluation, which looks at predication of performance, complexity, leanability and task analysis, GOMS.

The second one is analytical means that without the users, heuristic evaluation, cognitive walk-through. And the last one is Empirical. It is assessed by observing in experiments, rather thanappraisingthe theoretical validity. Whenever Adaptive systems created for practical use, hence empirical methods should be used for their evaluation. The benefits of Empirical Evaluation; it is best for objective assessment of design and best for the broadest range of usability problems. The following subsection present the most common approach developed to achieve the previous functionalities.

3. Empirical evaluation approach

Weibelzahl[11], acknowledges that empirical research is absolutely necessary for anestimation of the effectiveness, efficiency, and usability of a system that applies artificialintelligent techniques in real-world scenarios. Empirical evaluations (also known ascontrolled experiments) refer to the appraisal of a theory by observation in experiments.These evaluations help to estimate the effectiveness, efficiency and usability of a systemand may uncover certain types of errors in the system that would remain otherwiseundiscovered. The researchers acknowledge that the key to good empirical evaluation isthe proper design and execution of the experiments so that the particular factors to betested can be easily separated from other confounding factors. This method of evaluationis derived from empirical science and cognitive and experimental psychology. Empirical studies are very good at identifying design errors and false assumptionsbut they do not suggest new theories or approaches directly. Evaluators are faced with theproblem of defining control groups for those systems that either cannot switch off theadaptivity, or where a non-adaptive version appears to be absurd because adaptivity is aninherent feature of these systems[1].

4. Empirical evaluation requirements

Careful planning, Careful execution, Users must represent actual user population and Must have expense account – pay users. And the Empirical Methods are; focus group, interviews, questionnaires, Systematic Observation, expert review, prototyping, cooperative evaluation, participative evaluation, contextual inquiry, usability testingand controlled experiments.

The evaluation is focused on System interface, time, etc, is controlled –factor, independent variable or in User satisfaction – measured, observable, dependant variable. These variables are related to user and system, on user-related which describe user characteristics like demographics (gender, age, SES, etc.), observed performance (success rate, number of pages viewed, etc.), typically collected using questionnaires. Also on system-related which describe the system's operation as in the average response time, no. concurrent queries, perceived response delay (etc.) and typically collected using system logs [8].

In order to provide the best support for learners, a user-centered evaluation approach for enhancing and validating the student model of AEHS has been proposed, that combines AH and information retrieval techniques.

User-Centered Evaluation (UCE) can serve three goals: verifying the quality of an AEHS, detecting problems in the system functionality or interface, and supporting adaptivity decisions. These functions make UCE a valuable tool for developers of all kinds of systems, because they can justify their efforts, improve upon a system or help developers to decide which version of a system to release.

The benefits of the user-centered approach is savings in terms of time and cost, ensuring the completeness of system functionality, minimizing required repair efforts, and improving user satisfaction. This may lead to higher adoption of the AEHS, ease of use and a more enjoyable student experience. Student model performance is usually measured in terms of actual and expected accu-

racies, where actual accuracy is a model's probability of a correct response averaged across all users. For example, Corbett and Anderson in 2008 used correlation, mean error and mean absolute error to quantify model validity [3].

5. Pitfalls and problems in evaluation of AEHS

The evaluation of an adaptive system is a difficult task due to the complexity of such systems, as shown by many studies. It is of crucial importance that the adaptive features of the system can be easily distinguished from the general usability of the designed tool. Issues arise in the selection of applicable criteria for the evaluation of adaptivity. The evaluation of adaptive educational hypermedia systems is not easy, and several researchers have pointed out potential pitfalls and challenges when evaluating adaptive systems[1]. These Pitfalls and problems identified in the the following table.

Table 1 show the pitfalls and problems identified in evaluation of AEHS [1]

Pitfalls in evaluation of adaptive systems	
Statistically insignificant results	Adaptivity is typically used when individual users differ.However, differences in approach and preferences are likely to lead to a large variance inperformance results, which makes it more difficult to produce statistically comparableresults. In order to produce significant results, large volumes of queries and users arerequired. There are few general guidelines for the selection of these .measurements

Difficulty in defining the effectiveness of adaptation	It can be difficult to define whatconstitutes a useful or helpful adaptation.
Insufficient resources	To fully evaluate an adaptive system it is often necessary to have a large number of individuals interacting with the system. This is in part due to the expected variance between participants mentioned above.
Too much emphasis on summative rather than formative evaluation	Evaluations oftenmeasure only how good or bad a system is rather than providing information on where theproblems are and how a system can be improved.

Conclusion

The evaluation of learner and tutor feedback is essential in the production of high quality personalized Technology-Enhanced Learning (TEL) services. The evaluation focuses on the technological design and performance of systems without justifying the designs through the lessons learned from evaluations. Therefore, in this paper, we focus on a set of performance evaluation metrics that have been proposed in the literature. Throughout the paper we found that there are several evaluation metrics to validate or measure the educational system. The empirical evaluation requirements were discussed as well,such as; careful planning, careful execution, …, etc. The results of applingempirical evaluation approach will be reported later in a separate paper to assess the impact of incorporating four user's characteristics within AEHS on learning outcomes.

References

1. Catherine Mulwa, Seamus Lawless, Mary Sharp and Vincent Wade, The evaluation of adaptive and personalized information retrieval ystems: a review. Int. J. Knowledge and Web Intelligence, Vol. 2, Nos. 2/3, 2011
2. Brusilovsky, P. (2004) 'Knowledge-Tree: a distributed architecture for adaptive e-learning', *Proceedings of the 13th International World Wide Web Conference on Alternate Track*, pp.104–113.
3. P. De Bra, "Web-based educational hypermedia", Data Mining in E-Learning, C. Romero, and S. Ventura, eds.,WIT Press, pp. 3-17, 2006.
4. P. Brusilovsky, "Adaptive hypermedia. Methods and techniques of adaptive hypermedia," Int. J. User Modeling and User-Adapted Interaction, vol. 11, no. 1/2, pp. 87-110, 2001.
5. C. Mulwa et. al., "A Web-based Framework for User- centred Evaluation of End-User Experience in Adaptive and Personalized E-Learning Systems", IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, 2011.
6. Demetrios G. Sampson, Pythagoras Karampiperis, "Decision Models in the Design of Adaptive Educational Hypermedia Systems" IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited, 2012.
7. N.P. De Koch, "Software Engineering for Adaptive Hypermedia Systems Reference Model, Modeling Techniques and Development Process", PhD Thesis, University Munchen. 2001.
8. P.D. Bra, L Calvi, "AHA! An Open Adaptive Hypermedia Architecture, " New Review of Hypermedia and Multimedia 4 (1),115-139. 1998.
9. Brusilovsky, P. and Pesin, L. "Adaptive Navigation Support in Educational Hypermedia: An Evaluation of the ISIS-Tutor", Journal of Computing and Information Technology, (1998).
10. Brusilovsky, P et. al., "The Value of Adaptive Link Annotation in E-Learning: A Study of a Portal-Based Approach ", ACM ,Toronto, Ontario, Canada, pp. 223-227, 2010.
11. Weibelzahl, S. (2003) Evaluation of Adaptive Systems, PhD Thesis, University of Trier.
12. Herman, DwiSurjono, The Evaluation of a Moodle Based Adaptive e-Learning System. International Journal of Information and Education Technology, Vol. 4, No. 1, February 2014.