

DYNAMIC SIMULATION OF A STUDENT E-LEARNING PERFORMANCE PROFILE THROUGH WEB-BASED DATA-LOG MINING

Nikolaos V. Karadimas¹, Vassilis Nikolopoulos¹, Georgios Mpardis¹, Alessandra Orsoni²

¹National Technical University of Athens, Multimedia Technology Laboratory
9 Heroon Polytechniou, Zografou Campus, 157 80 Athens

²Kingston University – Faculty of Business and Law

nkaradim@central.ntua.gr (Nikolaos V. Karadimas)

Abstract

The purpose of this paper is to present a web-based method for simulating the educational performance and the overall educational track-path of a student who takes part in e-learning courses. Various e-learning courses that were organized at the Multimedia Technology Laboratory of NTUA, were used, including over 200 student records. The data logs from the LMS platform (Moodle MySQL database) were being used as a primary data mining source and by using an advanced web-based information system with data mining algorithms and relevant information about the course modules with some difficulty indexes, the educational simulated profile of the students is extracted. Data are correlated and clustered over specific central numerical values, representing course difficulty or other educational parameters and some statistical graphs are generated. The simulation results can be really helpful for e-learning teachers and professors, since by comparing the resulted curves and the difficulty metrics from the simulation, they can estimate and visualize the overall performance of the student during a time-dependending e-learning procedure and can create a complete educational simulation of a class and its effectiveness. By comparing an optimal fused traversal path, generated by the professor with the one obtained by the evaluation tool, learning efficiency and comprehension of the student can be extracted and simulated Weak educational points throughout the course, difficulty parameters and causes that are invisible to the professor, can be detected and focused additional help and support can be realised.

Keywords: e-Learning Simulation, Student Paths, Web-Based Simulation, Educational Visualization, Data Fusion.

Presenting Author's biography

Nikolaos V. Karadimas graduated from Glasgow Caledonian University, Scotland with a BEng (Hons) and a MSc in 1997 & 1998, respectively. He also received a MSc in Distributed and Multimedia Information Systems from Heriot-Watt University, Scotland in 1999 and his PhD from National Technical University of Athens in 2007. He is teaching in Hellenic Army Academy, TEI of Piraeus, TEI of Chalkida and in Technical NCO Academy. He is a member of the Greek Chamber of Engineers, member of IEEE and member of IEE.



1 Introduction

The emergence of e-learning has brought up big changes in the way courses are taught, the role of the teacher and the interaction between teachers and learners. The structure of knowledge transfer to the learners has been totally altered: it is based on various electronic channels (such as Internet, intranets, satellites, interactive TV, CDs, etc.), and therefore has become more impersonal than in the past. The role of the teacher has dramatically changed: since the e-learning system is delivering the course, the role of the teacher now is to guide, support and motivate the learners via a non face-to-face interaction with them through electronic channels, which demands considerable effort and necessitates the use of new techniques that did not exist in the past and for which teachers are not sufficiently trained. At the same time huge investments for the development of e-learning are made. So, as it happens with all investments, there is a growing need for methods to evaluate e-learning, both at the formative level, in order to diagnose weaknesses and make improvements, and at the summative level, in order to measure the value these investments create and conclude whether they have fulfilled the objectives they have been made for.

In recent years, various web based learning systems have been developed and greatly acknowledged, due to the need of asynchronous learning. Although much effort has been made towards the development of e-learning material and platforms, not much has been done in the field of e-learning evaluation. In the present paper, a web-based method for the evaluation, of the degree of student comprehension and anticipated difficulty, is presented. Based on an open source e-learning platform, namely Moodle, and utilizing the student's log-files stored in its data base, we managed to derive an educational statistical index which describes the student's anticipated difficulty for a specific course. The metric can be used to evaluate the effectiveness of an e-learning procedure. Thus, by analyzing the average students' on-line stay to the course and using it as a statistical metric we managed to represent in graphs the difficulty of a random student for the specific course taken. A combination of J2EE n-tier architecture and Matlab was used, in order to create an effective web-based information system, acting as the main evaluation tool. Through data mining procedures and based on a Case study from our e-learning courses at Multimedia Technology Laboratory, various graphs were generated, and useful conclusions were drawn for the importance of the above model for e-courses evaluation.

2 Relevant Work

Most of the research that has been conducted in the area of e-learning evaluation and simulation-modelling, concerns mainly the formative level, and

aims at the identification of the factors affecting either the extent of use of e-learning by learners (user acceptance), which is usually regarded as the basic surrogate measure of the value that e-learning generates, or the learners' intention to use e-learning. In this direction Selim [5] used the Technology Acceptance Model (TAM) [4] in order to investigate empirically the acceptance of course web-sites by students and identify its main determinants and educational paths. For this purpose he developed — the Course Website Acceptance Model“ (CWAM) consisting of the three constructs of the TAM (Perceived Usefulness, Perceived Ease of Use, and Use). By creating a questionnaire based on literature review, he validated the above model (using structural equation modelling techniques) and revealed the most important critical success factors of web-site acceptance. Saade and Bahli [6] conducted an empirical study aiming at understanding and explaining the acceptance and modelling (intention to use) of Internet-based learning systems. They have based on an extension of the TAM, which includes the concept of Cognitive Absorption as antecedent of Perceived Usefulness and Perceived Ease of Use. The results of this study, stemming from data collected from students, provided support for this model as explaining the acceptance of the Internet-based learning system and for Cognitive Absorption as an important variable affecting the TAM variables. In the same direction other researchers focus on the e-learning continuance decision, regarding it as being strongly associated with the value created by e-learning. Chiu et al [7], aiming at supporting the formative evaluation of e-learning, examine the factors affecting the e-learning continuance decision, using a research model combining the Expectancy Disconfirmation Theory (EDT) with conclusions of studies on customer satisfaction and IS quality; their results suggest that continuance intention is determined by satisfaction, which in turn is jointly determined by perceived usability, perceived quality, perceived value and usability disconfirmation.

Wang [8] has a totally different approach: in order to support (mainly) summative evaluation of e-learning and -to a lower extent-formative evaluation, he developed a global instrument for measuring the total e-learner satisfaction with asynchronous e-learning systems (i.e. a global satisfaction index). This index was calculated as the sum of 17 relevant variables, which have been determined through a review of the literature, and then on data analysis of a sample of e-learners. His final model theorizes that e-learner satisfaction is determined by four major constructs: content, learner interface, learning community, and personalization.

The conclusion drawn from reviewing the e-learning evaluation and simulation literature is that a complete modelling, simulation and evaluation methodology is missing: A methodology which supports both the

summative and the formative level, and includes measures of the impact (value created, appropriately measured), capabilities and resources offered to the e-learners (e.g. content, support by instructor, educational feedback etc).

Moreover, most of the studies that have been conducted in this area focus on the formative level and aim at explaining and identifying the main factors affecting either the extent of use or the continuance intention, regarding these two constructs as main surrogate measures of e-learning value, clearly influenced by a marketing or product placement perspective. However, the use or intention to use of e-learning is often just better than the other existing options, or even in some cases mandatory (i.e. there are no other options), so they do not always reflect the magnitude of the value created by e-learning (we can have e-learning systems with equal use but offering very different levels of value). For this reason a much better direct measure of the value created by e-learning is the extent of accomplishment of the various kinds of educational objectives.

3 Learning Data Fusion and Base Model

We have developed a learning data fusion architecture that can be applied for network-based dynamic distributed systems or conventional stand alone centralized LMS systems. The following characteristics can co-exist in both systems: educational components of the system can be inserted or excluded without stopping the entire system; the environment changes with time, as do the educational measures from it; an application's necessity changes according to different states of the learning environment, and as a consequence, it can use or reuse educational components in the different states (quiz, forum, etc);

All of these problems should be addressed from the application perspective to make the LMS system robust to the dynamic environment. We think that some of these problems can be addressed at the learning data fusion module of the application.

Data fusion is a formal framework used to express the convergence of data from different sources in which is expressed the means and tools for the alliance of data that originated from different sources.

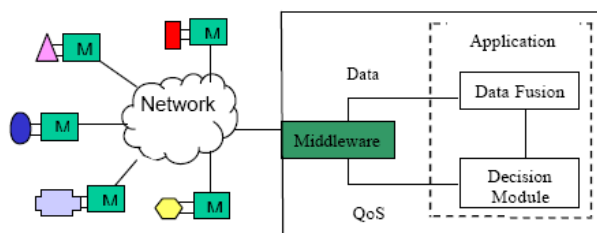


Fig. 1 Learning Data Fusion Topology

Our learning evaluation method is based on the above Data Fusion factor (fused Time, grades, performance, background, experience from different students, etc). This is the students' fused on-line stay to the e-learning course. As stated above, the learning material and student performance, which we will simulate and evaluate, concerns an e-learning course divided into 8 learning units, each of which has duration of one week. We will refer to the time span of each student i in a learning unit j as T_{ij} . Adding, the sum of the time spans of a student, in every learning unit i and dividing the sum to the total number of the learning units, we end up with the average time (\bar{T}_i) of the student i in learning units. The above metric can be shown in equation (1) below:

$$\bar{T}_i = \frac{T_{i1} + T_{i2} + \dots + T_{ik}}{k} = \frac{\sum_{j=1}^k T_{ij}}{k} \quad (1)$$

We will then define as W_{ij} , the statistical deviation of the student i from the average attendance time (\bar{T}_i), in a learning unit j by using the relation $W_{ij} = T_{ij} - \bar{T}_i$.

Adding the deviation W_{ij} of all students in the learning unit j , and dividing the sum to the total amount of students n we can extract vital deductions about the level of difficulty B_j of each learning unit j . Thus we have the final equation (2) below:

$$\begin{aligned} B_j &= \frac{(T_{1j} - \bar{T}_1) + (T_{2j} - \bar{T}_2) + \dots + (T_{nj} - \bar{T}_n)}{n} \\ &= \frac{W_{1j} + W_{2j} + W_{3j} + \dots + W_{nj}}{n} \\ &\Rightarrow B_j = \frac{\sum_{i=1}^n W_{ij}}{n} \quad (2) \end{aligned}$$

The above mathematical expression will be used in a Case study, as a metric, in order to generate a statistical fused diagram representing and simulating the difficulty of a random student for the e-course taken, by taking into account fused data (time, grades, performance). The above fused metric will be also represented in a statistical simulation graph.

4 Proposed Solution and Web-System

The methodology used to analyse and apply the above described evaluation model, comprised of four distinct functional steps

- Definition of evaluation model & metrics
- Logs analysis and fused extraction
- Logs processing and Data Fusion
- Metrics evaluation and graphs

A J2EE web-based architectural model was produced, with the addition of Matlab server. The log analysis was performed on-line and the output results were presented through a simple Internet explorer.

After defining the metrics, the variables and the final mathematical model, a specific database mapping must take place, in order to identify the appropriate attributes that need to be extracted. The modeling and mapping procedure of the database is based on the model that will be used to generate evaluation indicators. Attributes and Tables selection is performed according to the derived model (eg. average student times). The third step was to extract log information from the MySQL database, through ODBC using specific SQL queries. The extracted data were dynamically processed on the Matlab application server in order to produce graphically and numerically the output metric, the average difficulty B_j of the student to the specific course.

Two different environments were involved in the development of the web-based evaluation tool interface (Fig.3): HyperText Markup Language (HTML) standards with the addition of some Java Server Pages (JSP) under J2EE specifications, and advanced Matlab code [11], [10], which permits the dynamic mathematical development of the evaluation model described above. The Matlab web server toolbox [11] was effectively used by combining Java language, to produce dynamic Matlab Server Pages (MSP). Matlab Server Pages, permit the direct execution of Matlab functions directly from the JSP source code, by using specific tags, identical to the JSP Tags. The communication was preserved, through internal Java objects, which are incorporated in the Apache Tomcat server. Remote Matlab Method Invocation (RMMI) was achieved by this way and by combining CGI-based POST methods to the Matlab Web Server; a very powerful Matlab Application Server (MAS) was produced. The developed system allowed the creation of compact programs, called Matlab Beans (MB) and the execution of those beans (interconnected m-files) under HTML and JSP pages, located inside the Apache Tomcat Server. The application, using web GUI facilities sends data, through java calls and post methods to the matlab beans and vice versa. Student log data received from the MySQL database through Open Data Base Connectivity (ODBC), are processed inside the core matlab engine located on a clustered 2nd tier application server. The results are returned and presented on the Apache Tomcat web browser window at the client computer (evaluation statistical graphs and tables). With this configuration the client computer only needs to run the web Apache browser to access to the HTML and JSP document, since the matlab server and the web server run on different remote server stations. Therefore the process is transparent to the user who does not need to know matlab scripting to generate and have access to the

available student logs and the statistical results. Thus, it can be said that the overall produced evaluation tool is based on the classical n-tier philosophy and uses a back-end database (MySQL), a central application server (Matlab server) and a front-end web container (Apache Tomcat). Concerning the front 'client tier', it can be accessed through several client types, such as standard web browser exchanging content in standardized ways. The front-tier is based on the Apache-Tomcat web server with the addition of the MS Internet Information Server (IIS). Java Server Pages and ASP scripts, through dynamic forms, data fields and drop-down menus, gather user input and send the results to Matlab server, using RMMI.

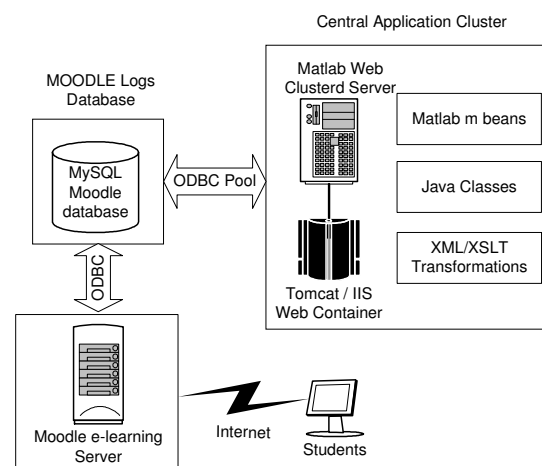


Fig. 2 Technical topology used for Data Fusion

5 Case Study

In order to test and evaluate the efficiency and the results of the above mathematical model, real student log data were used from past e-learning seminars, organised by the Multimedia Technology Laboratory [10]. It was an introductory e-course to Computer Networks and Internet Technologies that was used, as a Case Study and the student log files for this seminar were analysed.

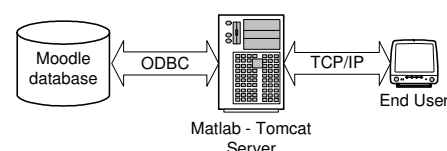


Fig. 3 Case Study Architecture

The seminar lasted for 10 weeks with a participation of approximately 21 students from all over Greece. The course was structured over 8 educational weeks, comprising of lecture notes in PDF format and handouts, presentation slides, flash animation and continuous on-line assessment using multiple choice

quizzes. There were also 5 graded assignments together with a final project that was prepared by students during the last 2 weeks of the course. The platform used was Moodle [10], a very popular and powerful PHP-based platform for the organisation and running of e-learning courses. The platform has a very effective log mechanism that logs every action of the student, by the time he logs into the system as a user. The logs are stored in the back-end MySQL database in specific tables. The basic Case Study architecture that was used, in order to access MySQL student log data, can be seen in Fig.2.

6 Results

The tool was tested on-line, from our Tomcat server at Multimedia Technology Laboratory. The target was the back-end MySQL server that Moodle uses. The tool can also work on-line and in parallel with any e-learning procedure. Initially, one random student was selected, in order to create the appropriate indexes and generate the graphs. The generated graphs can be shown below:

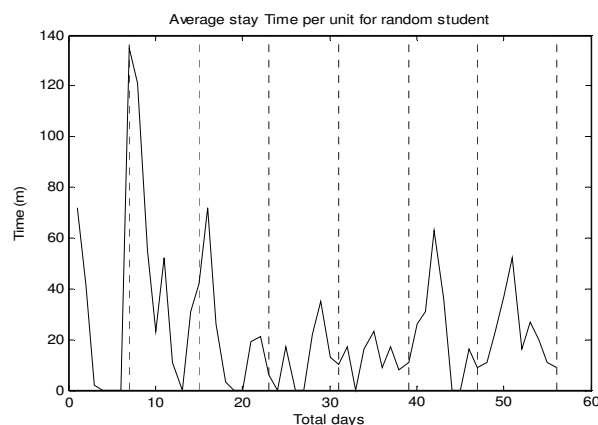


Fig. 4 Simulated stay time per unit for a student

As it can be seen from Fig. 4, the student follows a random time span over 56 days of course duration. The course is divided into 8 weekly units. As we can observe, the student spent more time during the first 3 units (difficult ones) and during the end (final assignment). The above graph in Fig. 4 is called the Simulated Traversal Path of the student.

Next we produce the appropriate graph for the same student, using the W_{ij} metric coefficient.

As it can be seen from Fig. 5, the coefficient follows the average stay time. At the beginning where the student was spending enough time for the course, we have a positive value. During the intermediate units the average time drops (negative weight coefficient) and towards the end we have some positive values.

Finally, by repeating the same procedure for all 21 students who took part in the introductory e-learning course, we generate the final B_j metric coefficient, indicating the level of difficulty per unit taught.

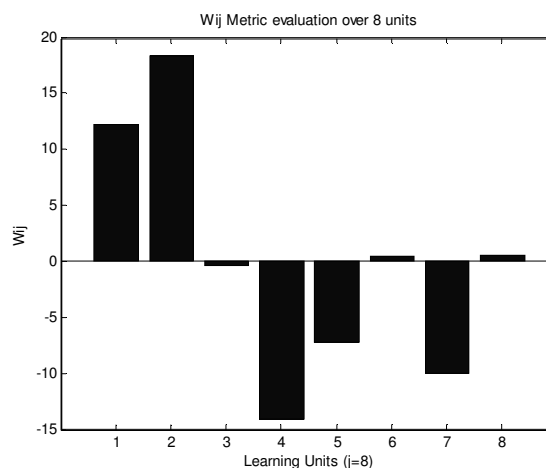


Fig. 5 W_{ij} metric evaluation for a student

The final graph generated can be shown below in Fig.6

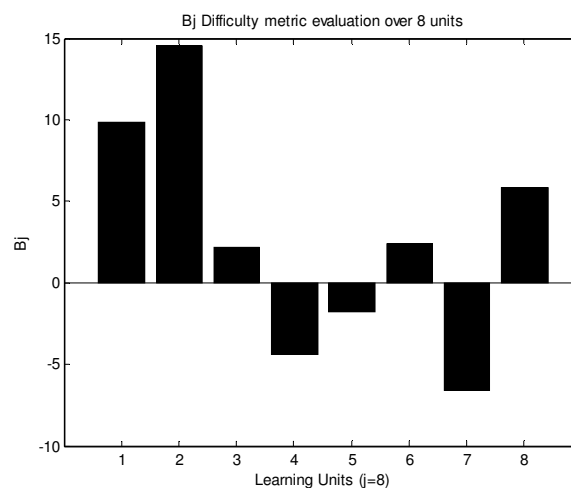


Fig. 6 B_j metric evaluation for a student

We can observe that the bar diagram follows slightly the average time span diagram (Fig. 4). The difficulty regions can be easily identified and these are the first two units (sub netting theory chapters) and the final unit.

7 Discussion and Future Work

The presented simulation model generates the student's fused traversal path for a specific e-course. The derived fused metrics and statistical graphs describe the overall difficulty of a student group who follow an e-learning program. The tool is web-based and can be used in parallel with any e-learning procedure, by evaluating continuously the educational efficiency. Some cross-correlations can be made, based on the combined metrics and graphs of T_i , W_{ij} and B_j indexes. By comparing an optimal fused traversal path, generated by the professor with the one obtained by the evaluation tool, very useful conclusions can be drawn regarding e-learning efficiency and comprehension of the student. Weak

educational points throughout the course, that are invisible to the professor, can be detected and focused additional help and support can be realised. Future work will include the generalization of the above on-line web-based evaluation and simulation tool for different evaluation models and the combination of different heterogeneous educational and statistical metrics, in order to produce combined student traversal paths.

8 References

- [1] Marsh, H. W. (1982) SEEQ: A reliable, valid and useful instrument for collecting students' evaluations of university teaching, *British Journal of Educational Psychology*, 52
- [2] Hoyt, D. P., Cashin, W. E. (1977) IDEA Technical Report No 1: Development of the IDEA System, Kansas State University, Center for Faculty Evaluation and Development.
- [3] Cashin, W. E. Downey R. G. (1992) Using Global Student Rating Items for Summative Evaluation, *Journal of Educational Psychology*, 84(4), pp. 563-572.
- [4] Davis, F. D. (1989) Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology, *MIS Quarterly*, 13(3), pp. 319-339.
- [5] Selim H. M. (2003) An Empirical Investigation of Student Acceptance of Course Websites, *Computers and Education*, 40, pp. 343-360.
- [6] Saade, R., Bahli B. (2005) The Impact of Cognitive Absorption on Perceived Usefulness and Perceived Ease of Use in on-line Learning: an Extension of the Technology Acceptance Model, *Information and Management*, 42, pp. 317-327.
- [7] Chiu, C. M., Hsu, M. H. Sun, S. Y. Lin, T. C. Sun P. C. (2005) Usability, Quality, Value and E-Learning Continuance Decisions, *Computers and Education*, 45, pp. 399-416.
- [8] Wang Y. S. (2003) Assessment of Learner Satisfaction with Asynchronous Electronic Learning Systems, *Information and Management*, 41, pp. 75-86.
- [9] Bloom B. S. (1956) *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*. New York: David McKay Co Inc.
- [10] <http://elearn.medialab.ntua.gr>, Medialab Elearn Server
- [11] The Mathworks, *MATLAB 7 Documentation*. Natick, MA: The Mathworks Inc., 1996.
- [12] V. Nikolopoulos, Ioanna Lykourantzou, Giorgos Mpardis, Vassili Loumos, Eleftherios Kayafas A web-based statistical evaluation method for e-learning course using student logs analysis", *ICIE 2nd International Conference on Interdisciplinarity in Education*, Athens 2006