Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Economics and Finance 00 (2014) 000–000

Procedia
Economics and Finance

www.elsevier.com/locate/procedia

The Economies of Balkan and Eastern Europe Countries in the Changed World (EBEEC 2013)

E-LearningActivityAnalysis

First Author[a], Second Author[b], Third Author[a,b,]*

Stavros Valsamidis[a], Ioannis Kazanidis[a], Ioannis Petasakis[a], Sotirios Kontogiannis[b], Eleftheria Kolokitha[a]

[a]*Accountancy Department, Kavala Institute of Technology*
Agios Loucas, 65404,Kavala
[b]*Department of Electrical and Computer Engineering,*
Democritus University of Thrace, University campus
Kimeria, 67100, Xanthi

Abstract

Nowadays, web-based educational systems are being installed more and more by universities, schools, businesses, and even individual instructors in order to add web technology to their courses and to supplement traditional face-to-face courses. These systems accumulate a vast amount of data which is very valuable for analyzing the contentof the courses and their usage from the learners.

A new method is proposed in order to analyze these datain three layers based on an innovative organization. Seven variables are used.They are combined properly to quantify quality characteristics of the courses and offer useful insights based on their material and usage by the learners.

All the variables were put to test in a case scenario that took place in a Greek university web based e-Learning data. The evaluation of the results has shown that the proposed system managed to track both educational content and course usage successfully. The results confirmed the validity of the approach and showed a relationship among the components of the proposed 3-layer organization.

Keywords: E-Learning; course content and usage assessment;3-layerorganization.

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
*E-mail address:*author@institute.xxx

1. Introduction

Nowadays, the variety of different kinds of E-learning systems is very large.There are systems which support individual learning, collaborative learning, learning content management, learning activity management, formal learning, informal learning, and workplace learning. E-learning is the delivery and management of teaching, training and learning by electronic means (Wentling et al., 2000). According to (Normark, and Cetindamar, 2005) e-learning describes the ability to electronically transfer, manage, support, and supervise learning and learning materials. E-learning systems are used in both training and education resulted in their adoption by academia, as well as industry.

An important goal of e-learning is that it should be equivalent to or better than the learning provided through other delivery modes, such as the traditional face-to-face and classroom-based methods of instruction. A significant benefit of e-learning is that it allows learners' access to learning material at their convenience (DeLima, 1999) without the necessity for a physical classroom, because learners can learn anywhere where there is access to the Internet. E-learning enables the instructor to monitor the learners' progress continuously.

Christner (2003) cautions, however, that both critics and supporters have identified some weaknesses associated with e-learning. These include the lack of social presence usually associated with physical classrooms, as learners miss the real-life interaction with their colleagues and the instructor. This feeling of loneliness could be a serious stumbling block to learning; in adult education in particular there is much that learners can learn from each other. Christner further observes that it takes a long time for trust to develop among online learners. The fact that there is not physical contact between the educator and the learner is an obstacle to the quality and the quantity of the information which is usually provided to the educators about learners' progress. They do not track and assess all the activities performed by learners and they do not fully evaluate the effectiveness of the learning process.

Campbell et al. (2007) used academic analytics by improving student learning outcomes through "actionable intelligence" especially if academic organizations want to remain competitive in the global economy. If academic organizations want to remain competitive in the global economy have to improve student learning outcomes through actionable intelligence Campbell et al. (2007).

Regression analysis can be used to explore the relationship among different characteristics of each performance metrics and to predict the upcoming evaluation o both courses-educators and students based on past knowledge. (Kotsiantis and Pintelas, 2005; Feng et al., 2005; Myller et al., 2002; Mcdonald, 2004). Some regression techniques have also been used to predict student's performance from log and test scores in web-based instruction (using a multivariable regression model) (Yu et al., 1999) to identify variables that could predict success in colleges courses (using multiple regression) (Golding and Donalson, 2006).

2. Method

The method presents a new organization based on three layers. Each layer is described by different variables. A regression analysis is proposed in order to examine how these variables are related. A similar approach was presented by the authors (Kazanidis et al., 2012; Valsamidis et al., 2012a; Valsamidis et al., 2012b) in previous works but from a different perspective.

2.1. 3-Layer Organization

The e-Learning procedure is considered to be consisted of three layers: Course primitive, Course content and Course usage, as it is depicted in Fig. 1.

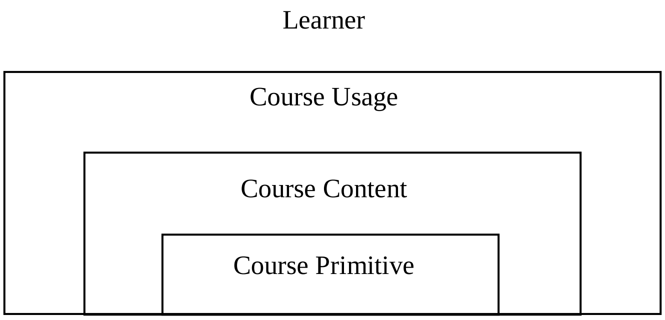

Fig.1 Three Layer organization

2.2. Variables

For each course we define seven variables, which are presented in Table 1.

Table 1.Variable name, descriptionand corresponding layer.

Variable	Description	Layer
Course ID	Unique identification code of the course	All
Year	The year in which the course is taught	Primitives
Category	If it is compulsory or optional	Primitives
Type	If it is theory or lab or mixed	Primitives
Files	Number of available files	Content
Size	Total size of the existing files	Content
Visits	Number of visits by all users	Usage
Duration	Duration of visits by all users	Usage

The above variables are selected due to certain educational goals. The main goal is to examine the factors that may affect learners' activity and are divided to three main categories (i) course primitive such as Year, Category, Type, (ii) course content such as Files and Sizeand (iii) course usage by learners such as Visits and Duration.

The first category of variables is related to the course. In particular the year that course is taught gives an estimation of attendees' experience in Higher Education. It would be interesting to examine whether novice learners of first years have similar performance with their fellows of higher years who are more expert in the skills of academia. The course category whether could affect learners' activity since compulsory courses are typically more challenging than optional courses. The type of the course is selected in order to examine if learners perform better in practical lab courses or theory courses.

The second category of variables is related to the courses' online educational content. More specificallythe number of pages, the number of files and their corresponding sizes give an estimation of the content quantity, which is a crucial factor of online educational content. On one hand, if the number of files and their size are small, this might be due to the weakness of the educator to upload enough educational content into the online platform. On the other hand if the course has a lot of files with big sizes this could lead learners to face the cognitive overload problem and not study the course effectively.

The third category of variables helps researchers to discover learners' activity and follow up in a course. The variable number of visits and duration show if learners find course useful and like to visit its pages. If learners of a specific course visit more pages for a long time, this means that course content is interesting and useful for the learners. This could reflect the course quality. Consequently a good course in terms of quality may help learners at their study.

2.3. Regression Analysis

The proposed method is a procedure consisting of a series of distinctive steps that analyze the e-learning data at the three layers. This procedure exploits the available data in order to find, through regression analysis, course parameters that affect learner learning outcome. Initially, collected data from E-Learning log files and the aforementioned variables are computed. Then a regression analysis is performed on the variables of the three layers.

Some of the goals of the present research uncover how the learners might improve their performance and in which way they will visit more often the educational content in the platform, as more visits means greater interest and more updated content for each course. It is investigated whether the attributes such as the number of visits and the duration are affected by the educational content. Educational content is expressed as the number of files and their corresponding sizes. Other attributes such as the academic year in which each course corresponds, the category (compulsory/non-compulsory) and the type (theory, laboratory, mixed) of each course are also investigated to validate whether they affect the output attributes.

3. Results

In this section we present the results of applying the proposed method to the data collected from the Open eClassE-Learning platform used by [Authors University] during the autumn and spring semester of academic year 2011-12. The data refer to 73 different courses.

Table 2.View of analyzed data.

Semester	Courses	Type	Kind	Total Number Of Files	Total Size	Number Of Visits	Duration Of Visits
WinterSemester	38	34 Required courses	23 T*	432	211666,32	323508	6288619
		4 elective courses	11 L*				
			4 M*				
SpringSemester	35	31 Required courses	19 T*	376	222786,83	276201	405701
		4 elective courses	9 L*				
			7 M*				
Total	73			808	434453,15	599709	6694320

Since 12 courses have minimal educational content and consequently minimal activity we examined the rest of the courses, namely 61 courses.

3.1. Course primitive

For each course we have data about the year in which it taught, if it is theory, lab or mixed, its type (required or elective course), Tables 3, 4 and 5.

Table 3.Courses allocation according to Year.

Year	Frequency
1st	17
2nd	17
3rd	19
4th	8

Table 4.Courses allocation according to Kind.

Kind	Frequency
Laboratory	19
Mixed	22
Theory	20

Table 5.Courses allocation according to Type.

Type	Frequency
Required course	51
Elective course	10

3.2. Course content

For each course, we have data about the number of files and their mean sizes (Table 6).

Table 6. Mean values for course content.

	Mean Number of Files		Mean Size	
Year	Mean	St. deviation	Mean	St. deviation
1st	22.8	5.47	10700	4116
2nd	5.29	1.65	4413.10	2337
3rd	7.63	2.40	5325	3040
4th	19	14.03	13338	11762

3.3. Course usage

For each course we have data about the number of visits and their duration (Table 7).

Table 7. Mean values for course content.

	Mean duration		Mean Number of visits	
Year	Mean	St. deviation	Year	Mean
1st	10.02	8,57	1st	10.02
2nd	1.87	0.07	2nd	1.87
3rd	1.92	0.11	3rd	1.92
4th	2.09	0.23	4th	2.09

3.4. Variables values

The values of the variables of the 3 layers for a sample of 12courses are represented in Table 8.

Table 8.Tracked data and variables.

Course (id)	Semester	Type(O/ S)	Kind(T/L/P)	Total number of files	Total Size	Number of visits	Duration of visits
2103	2	Required courses	Theory	25	14796.00	22005	26373
2104	2	Required courses	Lab	27	1834.50	13132	11146
3101	2	Required courses	Theory	11	50305.27	14840	26621
5103	4	Required courses	Theory	14	37905.40	11963	16122
5104	4	Required courses	Mixed	1	362.50	10153	16151
4108	4	Required courses	Theory	10	769.50	7854	13949
2108	4	Required courses	Theory	0	0.00	956	2153
1111	4	Required courses	Mixed	0	0.00	513	1416
6105	6	Elective courses	Theory	12	2943.79	6879	13710
6106	6	Elective courses	Mixed	6	2943.79	754	1459
6111	6	Elective courses	Theory	12	1793.13	3796	7016
6112	6	Elective courses	Lab	8	409.63	873	1524

Foreach coursewe have data aboutthe year in which it taught, if it is theory, labor mixed, its type (required or elective course), the total number offiles in theeclass for this course,the number of visitsin theeclass for this course and the averageduration of eachvisit. For all categorical attributes we used dummy variables.

3.5. Regression Analysis

Initially we tried to investigate whether the number of visits depends on other data which were recorded, using multiple linear regression.

The graph of standardized residual versus the standardized predicted values explain that satisfied the assumptions of multiple linear regression (a. for each value of the independent variable, the distribution of the dependent variable must be normal, b. the variance of the distribution of the dependent variable should be constant for all values of the independent variable c. the relationship between the dependent variable and each independent variable should be linear and d. all observations should be independent). Moreover satisfied the assumption of linearity/multicolinearity (all VIF test are between 1.04 and 1.65) and this of lack of autocorrelation (Durbin Watson = 1.8). Finally F testis 28.2 with five degrees of freedom, p-value<0.001 and adjusted R^2=69.4%, so the following model can be used to predict the mean number of visits.

The values of the variables of the coefficient, VIF and p-valueare represented in Table 9.

Table 9.Results of regression analysis.

	coefficient	VIF	p-value
constant	11286.8		<0.001
Number of files	184.1	1.159	<0.001
Mean duration of each visit	161.2	1.040	<0.001
year2*	-4437.6	1.656	0.007
year3*	-6808.7	1.640	<0.001
year4*	-9621.8	1.304	<0.001

* year2, year3 and year4 are dummy variables for the year

The above table explains that number of visits depends on number of files, mean duration and year of the courses.

The equation of linear regression is:

(number of visits) = 11286.8 + 184.1* (number of files) + 161.2* (mean duration) - 4437.6 * (year2) - 6808.7 * (year3) - 9621.8 * (year4)

It means that courses of second, third and fourth year have, at average, from 4437 to 9621 lower value of number of visits than these of first year.

For example the average number of visits for a lesson of the second year with number of files =10 and mean duration = 5 is: 11286.8 + 184.1 * 10 + 161.2 * 5 - 4437.6.

4. Discussion and Conclusions

This study is a method for discovering dependencies based on e-Learning data. The method presents a new organization based on three layers. Each layer is described by different variables. A regression analysis is performed in order to examine how these variables are related. It examines whether the courses usage by the learners is affected by the educational content exposed by the educators. It uses some attributes such as the number of files and their sizes that each course has on the e-learning platform, the number of visits and the duration of each visit. These attributes are used to assess the courses usage.

All these variables were put to test in an experimental case scenario that took place in a Greek university e-Learning platform. The findings of this first-time system experimentation are also worthwhile. There is evidence that the frequently updated or increased, in terms of content size educational material may potentially increase course usage results.

The originality of the method lies in the different use of existing techniques. It builds on existing work, but also extends it in a different way for the e-learning field. It has the following advantages: (1) It is independent of a specific E-Learning platform, since it is not based on the platform itself. Thus, it can be easily implemented for every E-Learning platform. (2) It uses variables in order to facilitate the evaluation of each course in the E-Learning platform and the instructors to make proper adjustments to their course educational material.

The method has to be tested in more experiments in other platforms, so as to check for conformant with the current results. At present, the calculation of the variables is being generated manually. Therefore, some future work is needed to overcome such limitation. Thus, a plug-in tool has to be developed to automate the whole procedure. This tool will run in periodically (each month) and will e-mail to the instructors the results and the suggestions for their courses. Similar policy was also applied by Feng and Heffernan (2006), where after long term vision the instructors were informed automatically by email about the quality of the content of their courses.

References

Campbell, J., Deblois, P., Oblinger, D., 2007. Academic analytics: A new tool for a new era. EDUCAUSE Review, 42, 4, 40−57.

Christner, T.2003. A classroom of one (book).Library Journal, 128, 1.

Delima, F., 1999. Web-based learning more cost effective. Computing Canada, 25, 27.

Feng, M., Heffernan, N., 2006. Informing teachers live about student learning: Reporting in the assistment system. Technology Instruction Cognition and Learning, 3, 1–8.

Feng, M., Heffernan, N., Koedinger, K., 2005.Looking for sources of error in predicting student's knowledge.AAAI Workshop Educational Data Mining, pp. 1–8.

Golding, P.,Donalson, O., 2006.Predicting academic performance.ASEE/IEEE Frontiers in Education Conference., IEEE Press, San Diego, CA, pp. 21–26.

Kazanidis, I., Valsamidis, S., Kontogiannis, S., Karakos, A., 2012. Measuring and mining LMS data, Panhellenic Conference on Informatics (PCI) 2012, October 5-7, 2012,University of Piraeus, Greece.

Kotsiantis, S. B., Pintelas, P. E., 2005. Predicting students' marks in Hellenic Open University.IEEE International Conference on Advanced Learning Technologies (ICALT 05), IEEE Press, Washington, DC, pp. 664–668.

Mcdonald B., 2004.Predicting student success."International Journal forMathematics Teaching and Learning, 1, 1–14.

Myller, N., Suhonen, J., Sutinen E., 2002. Using data mining for improving web-based course design.International Conference on Computers in Education, Washington, DC, 2002, pp. 959–964.

Normark, O. R.,Cetindamar, D., 2005.E-learning in a competitive firm setting.Innovations in Education & Teaching International, 42, 4, 325-335.

Valsamidis, S., Kazanidis, I., Kontogiannis, S., Karakos, A., 2012a. Measures for Usage Assessment in E-Learning.4th International Conference on Education and New Learning Technologies, Barcelona, Spain.

Valsamidis, S., Kontogiannis, S., Kazanidis, I., Karakos, A., 2012b.An approach for LMS assessment. International Journal of Technology Enhanced Learning IJTEL, 4, ¾, 265-283.

Yu, C. H., Jannasch-Pennell, A., Digangi, S., Wasson, B., 1999.Using online interactive statistics for evaluating web-based instruction.Journal of Educational Media International, 35, 157–161.

Wentling, T., Waight, C., Gallaber, J., Fleur, J., Wang, C., and Kanfer, A. 2000.E-learning-a review of literature. (Report No. ED-RR-89-3). Urbana-Champaign, Il: Knowledge and Learning Systems Group. (ERIC Document Reproduction Service No. ED450394).