Course ranking and automated suggestions through web mining

Stavros Valsamidis
Department of Electrical and Computer Engineering
Democritus University Of Thrace Xanthi, Greece
svalsam@ee.duth.gr

Ioannis Kazanidis
Department of Accountancy
TEI of Kavala
Kavala, Greece
kazanidis@teikav.edu.gr

Sotirios Kontogiannis
Department of Electrical and Computer Engineering
Democritus University Of Thrace Xanthi, Greece
skontog@ee.duth.gr

Alexandros Karakos
Department of Electrical and Computer Engineering
Democritus University Of Thrace Xanthi, Greece
karakos@ee.duth.gr

Abstract*—*This paper introduces new metrics for course evaluation. It is also proposes a ranking algorithm that classifies courses based on the previous course evaluation metrics and suggests appropriate actions for course content improvement. The algorithm was tested and verified successfully in data originated from the eClass platform of TEI Kavala.

Keywords E-learning; Data analysis; Suggestion rules;

I. Introduction

Learning Management Systems (LMS) are extensively used nowadays. One of the main problems of the LMSs is the lack of exploitation of the acquired information due to its volume. Most of the times, these systems produce certain reports with statistic data, which however, don't help instructors to draw out useful conclusions. Server log files store information containing the page requests of each individual user [1]. In this work, an algorithm is proposed for students' usage analysis. The main goal of the proposed algorithm is to seek for possible course's weaknesses through classification so as to assist instructors to improve their courses and offer better educational content. These automated suggestions will be generated according to the values of specific metrics, extracted through data analysis.

II. Related Work

There are several specialized web usage mining tools that are used in the e-learning platforms like CourseVis [2], and MATEP [3]. An interesting methodology for the maintenance of web-based courses was also proposed by [4] that incorporates a specific data mining step.

Previous works have applied data mining in the e-learning [5]. However, according to the best of our knowledge, none of the above tools, uses classification techniques for the generation of automated suggestions. Thus, we propose an algorithm for the evaluation of e-learning platforms based on learners' usage analysis with the aid of new educational metrics.

III. Methodology

The adopted methodology consists of three steps, namely i) logging step, ii) pre-processing and iii) application of the proposed algorithm. The first two steps of the methodology are based on the framework described in [6] and facilitate the extraction of useful information from the data logged by an LMS, while the third step applies the proposed algorithm.

More specifically the first step involves the logging of specific data from e-learning platforms with the use of a data recording module, which is embedded in the web server of the e-learning platform, and records specific e-learning platform fields. Second step involves outlier and missing values detection through statistical methods. After the refinement of the data, the new metrics are being calculated so as to return useful feedback about the course quality. The last step incorporates the application of an algorithm that will enable an automated system to send appropriate suggestions to the authors for the course improvement according to the values of the proposed metrics.

A. *Proposed Metrics*

The outcome of the logging step is the log file which contains among others the following items: i) courseID, which is the identification string of each course; ii) sessionID, which is the identification string of each session; iii) page Uniform Resource Identifier (URI).

In pre-processing step the aforementioned items are used in order to calculate per course metric values. Existing and proposed course metrics are proposed in Table I. We also propose a new ranking algorithm in order to achieve course classification based on course content and user interest.

Enrichment is a new metric that tries to express course content quality. It is a degree of student appreciation towards course instructor maintained information. This metric was first presented in [6] and is defined as the division of total course pages over unique course pages. However, in order to

TABLE I. Indexes name and description

Index name	Description of the index
Sessions	The number of sessions per course viewed by users
Pages	The number of pages per course viewed by users
Unique pages	The number of unique pages per course viewed by users
UPCS	The number of unique pages per course viewed by users per session
Enrichment	It is the one 's complement to the ratio of Pages per Unique Pages
Disappointment	The disappointment of users as expressed by the ratio of number of sessions divided by the number of course pages
Interest	It is the one 's complement to the disappointment

provide meaningful results, we redefine enrichment as the one's complement to the division of the unique pages over total course pages:

$$\text{Enrichment}=1\text{-Unique Pages/Total Pages}, \quad (1)$$

where: Unique Pages<=Total Pages.

Enrichment values are between [0, 1), where a course with minimum unique pages is close to 1 and 0 is a course where students follow only unique server paths. Enrichment offers a measure of how many times the unique pages were viewed by the users, representing rich course educational material. Low Enrichment means that users don't return to a visited page, since they have nothing new to learn in the course.

We propose another metric called *Disappointment* that reflects how quickly the users discontinue viewing course pages. It is defined by the average value of the division user session time sections over course pages viewed by each section users:

$$\text{Disappointment} = 1/Ns*\text{Sum}(\text{Sessions/Pages})_i, \quad (2)$$

where Ns is the number of time sections and for each time section $\text{Sessions}_i <= \text{Pages}_i$.

Due to the negative nature of the metric we replace Disappointment with Interest metric:

$$\text{Interest} = 1\text{- Disappointment}, \quad (3)$$

where Disappointment <=1.

Low Interest in a course means that there were not many viewed pages per session, while high Interest shows that users are interested in course content and continue more with their study. Educational material which does not fulfil user requirements, leads to user log out of the course.

We present a new metric proposed in [6] called *Unique Pages per Course per Session* (UPCS). UPCS is quantitative metric which expresses viewed by user total pages metric without counting duplicate ones. UPCS calculates course activity more objectively since it eliminates back-forward user effects. For example, some novice users may navigate in a course by visiting a page more than once. Another reason for this is because of course disorganization.

B. *Proposed algorithm*

The goal of the proposed ranking algorithm is to let an automated system classify courses and make suggestions for course improvement.

The first step of the proposed algorithm is course ranking in descending order by UPCS. This metric was succesfully used in [6] for course ranking. A course placed in the first ranking positions is a popular one, either because of exclusive quality of its educational content or quantity of course material.

The first suggestion rule (Fig.1) of the algorithm compares Interest metric of each course with the *a**Average(Interest) of all LMS courses, where *a* is a coefficient parameter. If Interest value is lower than *a**Average(Interest), it means that this course either does not have adequate educational content or its content quality does not satisfy user requirements. In order to distinguish these two cases a new condition is applied that checks whether course Unique Pages value is greater than the Average(Unique Pages) value for all LMS courses. If this condition is fullfiled, then course content quality is in need of amelioration, while if not, course content is of fine quality and new content additions need to be made due to course interest expressed by users.

```
IF Interest<a*Average(Interest) THEN
    IF UniquePages>Average(UniquePages)
    THEN "Improve Quality"
      ELSE IF Interest>b THEN "Add Content"
          ELSE "Add Content and Improve Quality"
ELSE  IF UniquePages<AVG(UniquePages)
     THEN  "Light need for new content"
    ELSE  "No further Improvements required"
```

Figure 1. First Suggestion rule

Additional course quality improvments are suggested if Interest is less than *b**Average(Interest), where *b* is a coefficient parameter. Finally, if Interest is more than the *a**Average(Interest), and the number of the Unique Pages is less than the Average(Unique Pages), then there is a light need for new content addition.

The next suggestion rule (Fig.2) applies the Enrichment metric. A low Enrichment value means that users do not visit course pages due to the lack of course content updates. If Enrichment value of a course is less than *c**Average(Enrichment), where *c* is a coefficient parameter, then the algorithm suggests that it would be a good practice for the author to update course content, so as to motivate users to re-visit his/her course.

```
IF Enrichment<c*Average(Enrichment)  THEN
    "Update the course content".
```

Figure 2. Second Suggestion rule

Algorithm's *a, b* and *c* coefficient parameters range between 0 and 1. In order to accurately calibrate the coefficient parameters, we first applied the algorithm to a reduced set of LMS courses. Course selection was performed based on better and worst case LMS courses, using UPCS ranking. Then a value was calculated by using best to worse course Interest deviation value. b coefficient value is the average best to worse course Interest value and c was calculated as the median value of the first k LMS lessons based on UPCS ranging, where k=5:

$$c=\text{median}(\text{Interest}_i)-k*Nc*0.0001, \quad (4)$$

where Nc is the number of LMS courses and k the number of selected courses with maximum UPCS ranking values.

IV. RESULTS

This experiment had two goals. First, to determine suitable values for a, b, c parameters and to test the two suggestion rules of the algorithm with respect to their impact on improving the course quality.

In detail the dataset presented in table II, was collected from the Technological Education Institute (TEI) of Kavala that uses the Open eClass e-learning platform [7]. The data involves 1199 students and 39 courses.

The first step of the algorithm was the ranking of the courses. Table II, due to space limitation, displays the results for the courses in ranking positions 1-4, 36-39, using the UPCS metric. The ranking of the courses is based on: first UPCS, then Enrichment and then Interest values.

Based on the aforementioned metrics we classify LMS platform courses using the following three steps:

1. Course ranking step: we consider course evaluation primarily, using the UPCS value and we rank LMS courses in descending order.
2. First suggestion rule step: The first suggestion rule is used in order to evaluate course content in terms of interest as it is expressed by course users and provide the appropriate suggestions to the instructors, related to the quantity and the quality of their course educational content.
3. Second suggestion rule step: Course content is examined in depth in order to express whether users are satisfied from what they see or course content seems confusing or complex to the end user.

In order perform the final two steps of the algorithm we first calculate coefficient parameters values. According to the experiment outcome, the values for the*, b, c* parameters were 0.9, 0.6 and 0.95 respectively.

The second goal of the experiment was to test the suggestion rules by showing them to the course instructors and receive verification feedback on the suggestion accuracy. When instructors applied the proposed suggestions, their courses improved in UPCS ranking position.

V. DISCUSSION & COCNLUSION

The results of the case study presented above were confirmed since the courses with a high number of UPCS are quite popular among the students. The association of the Interest metric and the number of Unique Pages, with the quality and the quantity of the educational content seems to be confirmed, since instructors of the experiment who followed the proposed suggestions, improved their courses ranking position. The Enrichment metric was associated with the course need for update. We found out that in some minor cases students visited more times the course pages because they couldn't understand their content. However more students act differently in this case and abandon the course pages. Therefore the usefulness of Enrichment is also confirmed.

Taking advantage of these metrics and indexes we go further by building classification rules which will benefit instructors to find courses' weaknesses and improve their educational contents. Consequently students will gain optimal educational material from the study of LMS courses.

The results seem to confirm the proposed algorithm. However more experiments have to be done in other platforms and courses. At present the suggestions as well as the calculation of the metrics and indexes, are being generated manually. Therefore some future work is needed to overcome such limitation.

REFERENCES

[1] M. Ueno, "Learning-Log Database and Data Mining system for e-Learning", Proc. of International Conference on Advanced Learning Technologies 2002 (ICALT2002), 2002, pp.436-438,

[2] R. Mazza, V. Dimitrova, "CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses," Int. Journal of Human-Computer Studies Vol. 65, No. 2, 2007, pp. 125–139.

[3] M. E. Zorrilla, and E. Alvarez, "MATEP: Monitoring and Analysis Tool for e-Learning Platforms," Proc. Eighth IEEE International Conference on Advanced Learning Technologies (ICALT08), 2008, pp. 611-613.

[4] E. García, C. Romero, S. Ventura, C. de Castro, "An architecture for making recommendations to courseware authors using association rule mining and collaborative filtering," Journal of User Modeling and User-Adapted Interaction, Vol. 19, No 1-2, 2008, pp 99-132.

[5] C. Romero, S. Ventura, "Educational Data Mining: a Survey from 1995 to 2005," Elsevier Journal of Expert Systems with Applications, Vol. 33. No.1, 2007, pp. 135-146.

[6] I. Kazanidis, S. Valsamidis, T. Theodosiou, S. Kontogiannis,, "Proposed framework for data mining in e-learning: The case of Open e-Class". Proc. IADIS Applied Computing, Nov. 2009, pp. 254-258.

[7] GUNet. "OPEN eClass", 2009. Retrieved January 05, 2010 from http://eclass.gunet.gr/.

TABLE II. PROCESSED E-LEARNING DATA FOR SELECTED COURSES

Position	Course ID	Sessions	Pages	Unique pages	UPCS	Interest	Enrichment	Automated Suggestions		
								Add new content	Improve content quality	Update
1	IMD105	91	297	11	216	0,694	0,963			
2	IMD132	152	230	7	184	0,339	0,970	*	*	
3	IMD35	87	338	8	179	0,743	0,976			
4	IMD36	72	217	7	134	0,668	0,968	-		
36	IMD130	12	30	5	22	0,600	0,833	-		*
37	IMD67	18	23	4	22	0,217	0,826	*	*	*
38	IMD49	14	23	5	21	0,391	0,783	*		*
39	IMD15	11	24	7	20	0,542	0,708	-		*
Average of all courses				7,25	68,41	0.55	0.892			

Appendix: * Strong suggestion, - Light suggestion