Multilevel clustering and association rule mining for learners' profiles analysis

Nawal Sael[1], Abdelaziz Marzak[2] and Hicham Behja[3]

[1] Laboratory of Information Technology and Modelization, Faculty of Science Ben M'sik Casablanca, 20800, Morocco

[2] Laboratory of Information Technology and Modelization, Faculty of Science Ben M'sik Casablanca, 20800, Morocco

[3] Laboratory of Command and Control and Production Systems, National High School of Art and Craft Meknes, 50000, Morocco

Abstract

Educational Data Mining [1] is concerned with developing methods for exploring data that come from educational domains, and using those methods to better understand learner, and how they interact with those environments.

In this research, we benefit from a new preprocessing approach applied to Moodle platform [1] [2] in order to apply clustering and association rule mining techniques to analyze learners' behaviors, to help in learning evaluation, and to enhance the structure of a given SCORM content.

We adopted the feature selection process and multilevel clustering that allowed us to confirm the importance of these new data preprocessing methods and to validate the usefulness of the attributes describing the learners' interactions with the SCORM content pertaining to learners' profiles detection. We also benefited from this approach as we sought to find possible relationships between the different parts of the relevant content and to help the teacher/ tutor to evaluate the structure of such content.

Keywords: Educational data mining; Moodle; preprocessing; clustering; association rule mining; learning profiles.

1. Introduction

Over the last few decades, e-learning has become an increasingly significant part of the overall learning process. By proposing an interesting alternative to traditional education based on new forms of access to pedagogical contents, it allows the development of the educational environments and. This is in addition to the variety of online courses it involves and the resources and activities it offers. Furthermore, virtual learning environments allow students to be the most important stakeholders in the learning process and facilitate their interaction with teachers, teaching resources and the overall virtual environment.

Despite their flexibility and adaptability, e-learning environments lack the direct interaction between teachers and learners [3] [4]. Because teachers are unable to analyze students' learning and behavior or detect which students require support, they will have no feedback regarding the instruction process. Consequently, they will not be able to adequately evaluate the structure of course content and its effectiveness in the learning activity.

Fortunately, e-learning systems accumulate a vast amount of information, known as log files, recorded in the server log or in the database. The latter seems to be very valuable for analyzing students' behavior and offering more information about their interaction with courses; it could also create a gold mine of educational data [5]. However, this is not the case since the log file doesn't only require a convenient tool for analysis, but an advanced approach for mining and preprocessing data [6], as well.

Some e-learning environments provide summary information on user access, such as the most visited pages, the most accessed activity or resource, and relevant other statistics [5]. This information is not sufficient to analyze learners' behavior and monitor their progress, even more so when the learners' interactions become greater. Accordingly, teachers face serious problems when trying to assist learners and understand their interaction with the e-learning environment.

Web Usage Mining (WUM) concerns the application of data mining techniques on web data in order to identify and characterize users' navigation behavior and tendencies. It is an emerging discipline, notably in e-learning domains (educational data mining) [7].

The goals of this research were: a) to offer an analysis of e-learning environment in order to provide interesting

[1] www.educationaldatamining.org

Copyright (c) 2013 International Journal of Computer Science Issues. All Rights Reserved.

insights into the learning process, b) to assist teachers in their monitoring, and c) to aid in making decisions about the effectiveness of SCORM content structure and organization. The research was constructed according to Web Usage mining process. In [1] we proposed our new data preprocessing method applied to Moodle logs in order to structure the data and allow the analysis of SCORM content in a new way that gives more detail about learners' interaction with this content. Next, in [2] we developed our preprocessing tool and applied statistics and visualization techniques in order to give interesting information about learners' use of SCORM course content in general and about each part of this content. In this study, we suggested applying clustering and association rule mining techniques to our preprocessed data in order to gain more knowledge about this content use. The aim was to also illustrate the potential of this preprocessing approach, to provide important information about the learners' profiles and to offer a very interesting way for grouping learners. In addition, an attempt will be made to decipher the more relationship between different parts of this pedagogical content in order to help teachers to have useful, informed knowledge regarding their pedagogical content structure.

This paper is organized as follows: Section 2 provides background information regarding educational data mining. Section 3 presents the context of this study. Section 4 describes our clustering approach and provides the results obtained via multilevel clustering. Section 5 exposes our association rules mining process and the results gained from it. Finally, some potential future lines of research and conclusions are suggested in Section 5.

2. Background

Educational Data Mining (EDM) concerns developing methods for exploring unique types of data that come from educational settings and using those methods to better understand learners and how they interact with those environments. It can be applied mainly to improve learning styles, evaluate the learning environment itself, study the efficiency of the educational support provided by the learning tools, and develop scientific research vis-à-vis the learning process and the learners [8] [9].

Moodle is one of the most popular e-learning environments all over the globe. The majority of Moroccan universities, for example, adopt it as an e-learning environment. Courses in this platform are structured on a set of tools, such as course content, course materials, tests (quizzes, assignments, and surveys), and collaborative tools (forum and chat)[2].

[2] www.moodle.org

In the past few years, a lot of research has been conducted to analyze learners' profiles and to provide more information about learners' interactions with all the courses or with particular or many activities (tools) [9].

The activities generally targeted by such an analysis include, for example, the analysis of collaborative interaction in chat [10] [11]; statistical indicators of learner's interactions in forums [12][13]; predicting final marks based on student participation in forums [14]; visualizing and clustering in discussion forums to measure the cohesion of small groups in collaborative distance learning[15]; analysis of learners' access to resources, overview of discussions and results on assignments and quizzes, [16]; and a blog-based recommendation generating mechanism [17].

Hung and al (2012) reviewed the literature on the existing EDM model in order to improve online teaching and learning to predict students' academic performances. They analysed learners' behaviours according to different activities, such as wiki, glossary, assignments, course materials, surveys, forum, and chat.

[19] [20] confirmed that the learning content and teaching strategies operate for a successful e-learning process more than tools and technologies. We can notice that pedagogical SCORM content is a very important activity for an effective learning process; it is generally used to generate an automatic learning path and to propose recommended itinerary in SCORM [21]. It is also used to in selecting different Learning Objects (LOs) for different learners based on learners' profiles [22], for recommending related LOs to learners [23], and for adapting learning resources to learners [24].

3. Context of the Study

3.1 Goal

As is indicated above, we have developed a preprocessing tool in order to implement a new data preprocessing method applied to Moodle logs based on SCORM content [1]. This new way gives more detail about learners' interaction with this content and allows us to apply a statistics and visualization technique in order to give interesting information about learners' interactions with SCORM content in general and about each part of this content (multilevel analysis)[2]. In this study, we propose the use of clustering, which allows us to study multiple attributes simultaneously, to find a group of students that can mainly work together in the project according to their similar uses of content and to study the relationship between content use and learner scores. We also suggest applying association rule mining in order to find a

Copyright (c) 2013 International Journal of Computer Science Issues. All Rights Reserved.

potential relationship between the different parts of SCORM content according to our multilevel data preprocessing methods.

3.2 The data

In this study, we used FOAD-ENSAM logs collected from the Moodle database management system. The experiments were conducted with Java/C++ students in a UML course proposed in this platform. This course complies with the course structure presented in [1]:

Chapter → Section → Sequence → Tasks

The course content (4 chapters, 14 sections, 20 sequences, 42 tasks) contains three evaluations dispatched at the end of the second and third chapters with a case study at the end of the course. The number of entries was 44507 and the number of final tracking entries was 3000.

4. Clustering Approach

In Sael and al. (2013) [25], we pointed out that statistical and visualization techniques permit the analysis of the profile of a group of learners globally or project on a particular learner. However, even if the information obtained is very rich and allows the teacher to have a large vision on the learning process, these results are in some way limited and hence do not allow the multidimensional analysis considering several attributes simultaneously. Clustering techniques allow us to overcome this limitation and identify similar profiles in an unsupervised manner. In addition, they offer the possibility to identify groups of learners with similar profiles and analyze the structure of each group.

4.1 Clustering approach

The case study described here is a multiple identification of learning profiles through clustering.

1. The first alternative is to provide a description of these profiles by analyzing learners' interaction with the e-learning environment according to the first level of data (session) [see Sael and al (2012)]. We used EM (Expectation-Maximization) and K-Means as clustering algorithms. EM algorithm was used as a similarity measure to characterize data and to know the number of clusters with which the K-Means algorithm will be executed. We present the results obtained with K-Means because it is easier to understand graphically and statistically.

2. In our second variant, we analyzed these profiles using information describing the interactions of learners according to the second level of access to educational SCORM content.

a. The first step was to add to the attributes used in the first variant the new ones that describe the interactions with the different chapters of this content.

b. Next, we applied a set of feature selection algorithms, in order to find the most significant attributes for clustering analysis. A process of feature selection was used to identify the attributes that could be more descriptive of a learner profile.

We used RapidMiner[3] tools that offer multitude of feature selection algorithms and incorporate several Weka [4] algorithms. To study the weights of attributes and their importance in determining the learner profiles we chose ten algorithms. We then chose the most selected attributes by those algorithms.

c. Finally, we applied the K-means clustering using the attributes obtained from the feature selection process described above.

In this paper, we propose the use of the results obtained by our preprocessing methodology. "Fig. 1" shows the overall process adopted.

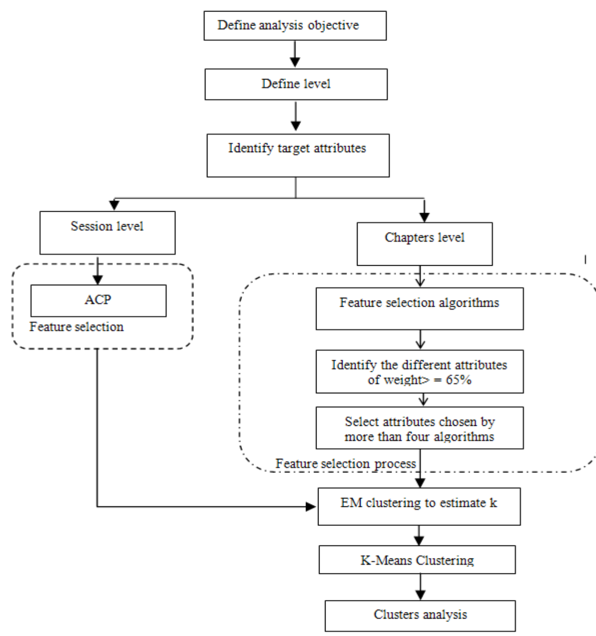

Fig. 1 clustering approach proposed

[3] http://rapid-i.com/
[4] http://www.cs.waikato.ac.nz/ml/weka/

Copyright (c) 2013 International Journal of Computer Science Issues. All Rights Reserved.

4.2 Learners' profiles clustering according to sessions levels

K-Means clustering

Initially, we chose a set of eight digital attributes representing learners through their interactions with educational content. We started with an analysis of the session level "Tab. 1"

Table 1: Session level clustering attributes

Attribut	Type	Signification
Duree_ses	datetime	Total session duration
Nbre_act_ses	Nombre	Number of action per session
Nbre_session	Nombre	Nombre de session apprenant
Heure_My_db_ss	Heure	Average time of the beginning of the sessions of the learner
Datefirstsession	Dateheure	Date of the first session of the learner
Datelastsession	Dateheure	Date of the last session of the learner
Nbrejoursession	Nombre	Number of days between the first and last login
score_UML	Nombre	Final score

In the first experiment, we analyzed these attributes using Expectation Maximization algorithm. The application of this algorithm allowed us to estimate K to three. In addition to this experiment, we confirmed this proposal (k = 3) through a series of discussions with the learner's teacher, he estimates that there are three groups of different learner profiles. Furthermore, the final marks of the learners in the UML module confirm that students can be divided into three groups according to their scores. We first started k-means with k = 3 cluster.
We also experimented with k = 2 and k = 4 and the results were more significant for k = 3.

Attribute	cluster_0	cluster_1	cluster_2
nbr_ss	4.100	7.143	5.800
nbr_act	77	160.429	111
score_UML	14.500	14.571	14.800
nbrejoursession	11.264	12.454	11.322
tt_dt_ss	17	10.714	13.600
datefirstsession	18.900	21.143	27.300
datelastsession	13.600	14.857	16
h_my_db_ss	10.500	4.571	13.500

	average	cluster 0	cluster 1	cluster 2
tt_dt_ss	14.111	17,000	10,714	13,600
score_UML	14.630	14,500	14,571	14,800
nbrejoursession	11.594	11,264	12,454	11,322
nbr_ss	5.519	4,100	7,143	5,800
nbr_act	111.222	77,000	160,429	111,000
h_my_db_ss	10.074	12:00:00	13:42:51	12:00:00
datelastsession	14.815	13	14	16
datefirstsession	22.593	18	21	27

Fig. 2 K-means clustering centroïd table (session level)

We can conclude from "Fig. 2" that our learners are distributed into three groups (clusters). The first group (10 students) of learners spent more time on the course (total duration of sessions is greater). However, they performed a minimum number of sessions (average of 4 sessions) and a minimum number of web pages visited (77 pages). The group spent a longer time on different parts of the course (max sessions duration). Hence we can say that this group is more engaged in the learning process (working time could be a measure of attention or engagement of learners) [26], even though their average score is not very important.
The second group of learners (cluster 1, 7 learners) had a minimum length of session and a maximum number of sessions and interactions. We can say that this group had a high browsing speed and its average UML score is significant. In their diagnosis, Vicente and Pain (2002) [27] estimated that if the speed of interactions is high and the performance is proved (average score), then motivation is high. Thus, we can conclude that this group was very motivated. This motivation allows the group to have good results.
The third group of learners (cluster 2, 10 learners) has a medium, total session duration, a number of sessions and average actions, with a moderate frequency of interactions and an average score which is greater than the overall average. We can say that this group is quite organized in the follow-up and subsequently may have the best result.

K-Means clustering using PCA factors

The principal component analysis (PCA) is a mathematical technique that uses an orthogonal transformation to transform a set of possibly correlated variables into a set of uncorrelated variables called principal components. Ding and He (2003) demonstrate that the use of K-Means via the PCA factors gives better results than simple k-Means uses. In the next step, we applied PCA and K-Means to our data shown in "Table 1".
Compared with the results of the application of simple K-Means, we obtained approximately similar clusters. The first cluster contains the same ten learners previously obtained, in addition to learner N° 8. The second cluster keeps its same learners and subsequently its characteristics. Finally, the last cluster has lost learner N° 8.

4.3 Learners' profiles Clustering according to chapter levels

The analysis of learners' interactions, according to session levels, allows us to have general information about learners' interactions with the whole content; it does not in any way allow us to have more information on the structure and internal organization of this content. In this section, we propose using the results of our preprocessing methodology in order to analyze learners' interactions, taking into account the attributes and variables that

Copyright (c) 2013 International Journal of Computer Science Issues. All Rights Reserved.

describe these interactions with the different parts of the content. In this case study, we focus on the analysis of learners' profiles according to chapter levels.

Select Attributes

To study the interactions of learners according to chapter levels, we propose to add to the attributes used in the previous analysis "Table 1" the following attributes: nbr_act_ch1, tt_dt_ch1, nbr_act_ch2, tt_dt_ch2, nbr_act_ch3, tt_dt_ch3, nbr_act_ec, tt_dt_ec. They describe the total duration and the number of learners' interactions in each chapter of our content, and thus describe the engagement [26] and the motivation of learners [27] depending on those chapters.
However, assuming that some attributes are not very relevant for the identification of learner profiles and in order to confirm the importance of the attributes describing the interaction of learners with chapter levels, we conducted a feature selection process so as to find the attributes which are the most discriminative of learning profile and the most significant for the clustering.

1. To study the weights of attributes and their importance in the description of a learner profiles, we chose ten algorithms (Feature Selection): ChiSquaredAttributeEval, FilteredAttributeEval SVMAttributeEval, Weight by Correlation, Weight Deviation by W-ReliefFAttributeEval, Weight by Uncertainty, W InfoGainAttributeEval-W-Principalcomponents, W-SymmetricalUncertAttribute Eval.
2. We adopted 65% as a feature selection attribute weight.
3. Finally, we chose the attributes which were selected by at least four of these algorithms.

The following "Table 2" shows the different selected attributes and how many times they were selected by these different algorithms.

Table 2 : Attributes are ranked by the process and frequency of appearance

Attribute	Number of times the attribute was selected
nbr_ss	4
tt_dt_ch2	4
nbr_act_ch3	5
h_my_db_ss	6
nbr_act_ch2	6
score_UML	6
tt_dt_ss	6
datelastsession	7
nbr_act	9

Analysis of the results allows us to say that the attributes initially selected in the first clustering are not important discriminators of the learning profiles (in comparison with other). Some attributes that describe the learners' interactions with the various chapters are more interesting (selected in previous Table). The session duration in the second chapter, the number of learners' interactions with the second and the third chapter are more important.
We can therefore conclude that multilevel preprocessing and the description of learners' interactions according to different levels of access to educational content in SCORM allows us to better describe the learners' profiles.

K-Means clustering

In what follows "Fig. 3" we present the results of the application of K-Means clustering on data that describe learners across different selected attributes in the feature selection process

	Average	Cluster 0	Cluster 1	Cluster 2
nbr_ss	5,519	5,000	7,143	4,923
nbr_act	111,222	71,429	160,429	106,154
score_UML	14,630	14,643	14,571	14,654
nbr_act_ch2	52,963	33,000	80,000	49,154
nbr_act_ch3	24,407	9,571	42,000	22,923
tt_dt_ss	14,111	15,286	10,714	15,308
h_my_db_ss	10,074	9,286	4,571	13,462
tt_dt_ch2	3,259	1,857	6,143	2,462
datelastsession	15	14	15	15

Fig. 3 K-means clustering centroïd table (chapters level)

We conclude that the second cluster (cluster 1) keeps the same structure as that which was found previously. However, the attributes describing the centroid of the cluster can better characterize learners' profiles. In addition to an overall minimum length of session and the maximum number of sessions and interactions, we can say that those students spent a significant part of their time in Chapter 2 with a significant number of interactions in this Chapter and in Chapter 3, as well.
The first cluster contains seven learners obtained from the previous results cluster. However, in addition to learner N° 8, which was removed during the PCA application to K-Means, two other learners (N° 24 and N° 25) were also removed. Only seven learners had common properties. In the previous section, we concluded that this group spent maximum session duration and a minimum number of sessions and interaction. However, we note that Chapter 2 and Chapter 3 were not among the priorities of this group (minimum duration and number of interactions), furthermore their UML score is under the average. Compared to the second cluster, we can assume that

Copyright (c) 2013 International Journal of Computer Science Issues. All Rights Reserved.

Chapters 2 and 3 are important in order to have a good score.
The third group (cluster 2, 13 learners) is characterized by a quite organized usage with very significant interactions on Chapter 2 and 3.
Overall, we could say that the multilevel analysis of learners' interactions describes in more detail the learners' profiles.

5. Association Rule Mining Approach

E-learning environments have a large number of webpages describing the various resources available. In addition, educational content can in reality be seen as a web tutorial that is available to learners and that contains a multitude of webpages offering a diverse educational content. Therefore, the application of standard algorithms for mining association rules from learners' interactions, with that content online, is not an easy task. Even if it can be applied, the results would be not very interesting.

5.1 Multilevel analysis of a pedagogical SCORM content.

In this analysis we propose using the hierarchical structure of educational content in SCORM format to extract possible relationships between its different parts according to a particular level of content organization [2].
The basic principle of our approach is to analyze the different relationships that may have different parts of a particular level of the SCORM content. For example, we can analyze the relationships that arise during learners' interaction with the different chapters of SCORM (chapter level), or entering different sections of a given chapter. In this case, the diversity of web pages and elements analyzed will be limited to a given level of analysis.
In this case study, we sought to analyze the relations between various chapters of the course. Taking advantage of our new preprocessing approach [1] [2], it was possible to know each learner's session and whether or not the learner has accessed the different chapters.

5.2 Experiment results

In this case study, we analyzed the different relationships that could occur between different parts of the first hierarchical level of educational content in SCORM (chapter).
RapidMiner tool offers several scenarios to apply association rule mining. In our case, we chose to apply the operator "FP-Growth" which calculates effectively all frequent patterns using the FP-tree structure. Then we ran the "Create Association Rules" operator which generates a set of association rules from the given set of frequent patterns.
In this case study, since our goal was to find sets of frequent patterns, we ran the algorithm with minimum support 0.2 and a minimum confidence 0.8 as parameters.

No.	Premises	Conclusion	Support	Confidence
1	chp3	chp1	0.309	0.885
2	chp2, chp3	chp1	0.221	0.917
3	chp2	chp1	0.544	0.953
4	etc	chp1	0.201	1

Fig. 1. The most common rules (chapter level)

The results in "Fig. 4" show that, generally, any interaction with a particular chapter is succeeded by a return to Chapter 1. By studying the structure of Chapter 1 we see that it proposes generalities on UML. However, statistical analysis over the duration spent in learning this first chapter shows that learners didn't give much time to this part of the content. Therefore, these generalities don't seem well assimilated by learners and they need to be reminded every time.
In order to facilitate learners' interactions with the content, we can recommend that the teacher first put some definition reminder whenever a chapter of the course deals with a particular concept. On the other hand, we propose a test or assessment at the end of the first chapter (even if it offers generalities) to enable learners to know their degree of assimilation of different concepts proposed in this chapter; we also urge them to give more importance to these concepts that constitute the basis of all other concepts studied latter.

6. Conclusion

In this paper, we have exploited our preprocessing methodology of the Moodle platform in order to provide recommendations for teachers. The techniques we used are clustering techniques to analyze the structure of the group and association rules mining to find eventual relationships between the different parts of content offered.
The application of the feature selection process and multilevel clustering allows us to stress the importance of our new data preprocessing methods and validate the usefulness of the attributes describing the learners' interactions with that content on learners' profiles detection. We also benefit from this preprocessing approach so as to find possible relationships between the different parts of such content and to help teacher/tutor to evaluate the structure of this content.
In the future, Both our data preprocessing method and the data mining techniques can be extended and enriched. For

Copyright (c) 2013 International Journal of Computer Science Issues. All Rights Reserved.

example, we can extend our preprocessing method to more activities like forum and quiz. Data mining techniques can be improved using greater data set and other techniques can be applied using our preprocessed data.

References

[1] N. Sael, A. Marzak, and H. Behja, "Prétraitement avancé des fichiers logs Pour une plate forme d'enseignement à distance" in NGN2010. 244-248.

[2] N. Sael, A. Marzak, and H. Behja, "Investigating an Advanced Approach to Data Preprocessing in Moodle Platform" International Review on Computers and Software. Vol. 7 N. 3: pp. 977-982

[3] M Zorrilla, S. Mill´an, E. Menasalvas. "Data webhouse to support web intelligence in e-learning Environments". In: Proc. of the IEEE International Conference on Granular Computing, Beijing, China, 2005, GrC pp. 722-727.

[4] R. Hijon, A. Velazquez, "E-learning platforms analysis and development of students tracking functionality". In Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications, June 2006, pp. 2823-2828.

[5] M.-E. Zorrilla, D. Mar´ın, E. Alvarez, "Towards virtual course evaluation using Web Intelligence". In: Moreno D´ıaz, R., Pichler, F., Quesada Arencibia, A. (eds.) EUROCAST, Springer, Heidelberg 2007. LNCS, Vol. 4739, pp. 392–399.

[6] C. Romero, and S. Ventura, "Educational Data Mining: a Survey from 1995 to 2005". Expert Systems with Applications, 2007, Vol 33(1), p.135-146.

[7] R. Cooley, "The Use of Web Structure and Content to Identify Subjectively Interesting Web Usage Patterns". ACM Transactions on Internet Technology, 2003, Vol 3(2), pp. 93-116.

[8] R. Baker, and K. Yacef, "The state of educational data mining in 2009: A review and future visions", J. Educ. Data Mining, 2009, Vol. 1, no. 1, pp. 3–17.

[9] C. Romero and S. Ventura, "Educational data mining: A review of the state of the art". IEEE Transactions on Systems Man and Cybernetics Part C. Applications and Reviews, 2010, Vol 40(6), pp.601-618.

[10] M. Cakir, F. Xhafa, N. Zhou and G. Stahl, "Thread-based analysis of patterns of collaborative interaction in chat" the international conference on AI in Education (AI-Ed 2005), Amsterdam, Netherlands, 2005.

[11] A. Anjewierden, B. Kolloffel and C. Hulshof, "Towards educational data mining: Using data mining methods for automated chat analysis to understand and support inquiry learning processes", International Workshop on Applying Data Mining in e-Learning (pp. 27-36), 2007.

[12] A. Anaya and J. Boticario(2009), "A data mining approach to reveal representative collaboration indicators in open collaboration frameworks", In International Conference on Educational Data Mining, Cordoba, Spain, 210-218, 2009.

[13] P. Reyes and P. Tchounikine, "Mining learning groups' activities in forum-type tools," in Proc. Conf. Comput. Support Collaborative Learning: Learning, Taipei, Taiwan, pp. 509–513, 2005.

[14] M. I. Lopez, J. M. Luna,C. Romero and S. Ventura, "Classification via clustering for predicting final marks based on student participation in forums", In Proceedings of the 5th International Conference on Educational Data Mining, EDM 2012, pages 148-151, 2012.

[15] C. Reffay and T. Chanier, "How social network analysis can help to measure cohesion in collaborative distance-learning," in Proc. Int. Conf. Comput. Supported Collaborative Learning, Bergen, Norvege, pp. 1–6, 2003.

[16] R. Mazza and C. Milani, "GISMO: A graphical interactive student monitoring tool for course management systems," in Proc. Int. Conf. Technol. Enhanced Learn., Milan, Italy, pp. 1–8, 2004.

[17] T. Huang, S. Cheng, and Y. Huang, "A blog article recommendation generating mechanism using an SBACPSO algorithm," Expert Syst. Appl. J., vol. 36, no. 7, pp. 10388–10396, 2009.

[18] J. L. Hung, K. Rice and A. Saba, "An educational data mining model for online teaching and learning", Journal of Educational Technology Development and Exchange, 5(2), 77-94, 2012.

[19] W. Schramm, "Big media, little media", Beverly Hills, CA: Sage, 1977.

[20] C. J. Bonk and T.H. Reynolds, "Learner-centered Web instruction for higher-order thinking, teamwork, and apprenticeship". In B. H. Khan (Ed.), Englewood Cliffs, NJ: Educational Technology Publications, Web-based instruction (pp. 167-178), 1997.

[21] E. Mor and J. Minguill ´ on, "E-learning personalization based on itineraries and long-term navigational behavior," in Proc. 13th World Wide Web Conf., New York, 2004, pp. 264–265.

[22] L. Shen and R. Shen, "Learning content recommendation service based-on simple sequencing specification," in Proc. Int. Conf. Web-based Learning, Beijing, China, 2004, p. 363-370.

[23] Y. Ouyang and M. Zhu, "eLORM: Learning object relationship mining based repository," in Proc. IEEE Int. Conf. Enterprise Comput., E-Commerce, E-Service, Tokyo, Japan, 2007, pp. 691–698.

[24] P. Karampiperis and D. Sampson, "Adaptive learning resources sequencing in educational hypermedia systems," Educ. Technol. Soc. J.,vol.8, no. 4, pp. 128–147, 2005.

[25] N. Sael, A. Marzak, and H. Behja, "Web Usage Mining: Data Preprocessing and Multi Level analysis on Moodle", IJCSI, Vol.10 Issue 2, March 2013, pp : 347 – 354.

[26] Cocea, M. and Weibelzahl, S.: 2007b, 'Cross-System Validation of Engagement Prediction from Log Files'. In: Duval.

[27] Vicente, A., Pain, H. (2002) Informing the detection of the students' motivational state: an empirical study. In S. A. Cerri, G. Gouarderes, F. Paraguacu, editors, Proceedings of the Sixth International Conference on Intelligent Tutoring Systems, volume 2363 of Lecture Notes in Computer Science, pages 933-943, Berlin. Heidelberg. Springer.

[28] Ding C. and He X., "Principal Components and K-means Clustering". LBNL Tech Report 52983. 2003.

Copyright (c) 2013 International Journal of Computer Science Issues. All Rights Reserved.