International Journal of Hybrid Information Technology
Vol. 9, No.11 (2016), pp. 229-238
http://dx.doi.org/10.14257/ijhit.2016.9.11.20

Academic Management and Analysis Method of MOOC

Fei Lang[1], Siwen Zhang[1], Shaobo Li[2] and Guanglu Sun[2]

[1]School of Foreign Languages, Harbin University of Science and Technology, Harbin, Heilongjiang, China
[2] School of Computer Science and Technology, Harbin University of Science and Technology, Harbin, Heilongjiang, China;
Corresponding author: Fei Lang, langfei@gmail.com;

Abstract

MOOC (Massive Open Online Courses), is a new mode of distance education courses based on the Internet technology. To solve problems in MOOC, such as lacks of management, supervision and comprehensive evaluation etc., this paper aims to study MOOC academic management and analysis method in contexts of colleges and universities. To supervise and trace students' learning process, an academic management system of MOOC was constructed. To assess and predict students' learning trends and achievements, an academic analysis model was established based on learning and interaction data analysis. Experimental results prove the above method insures expected learning effects when applying MOOC in colleges and universities.

***Keywords**: MOOC; distance education; academic management; academic analysis; academic prediction*

1. Introduction

Recently the new tend of the distance education is MOOC (massive open online course). Since 2012, Stanford University, Massachusetts Institute of Technology and Harvard University have launched hundreds of MOOC courses, which attracts millions online learners and initiates attentions and popularization of MOOC all over the world.

MOOC is usually built by the cooperation between the education industry and commercial companies. Education institutions represented by the top colleges and universities provide course videos and learning materials of MOOC, while commercial corporations are responsible for constructing online-learning platform, making and transmitting good course learning resources. Three suppliers of MOOC – Coursera, Udacity, edX, have many partners in the education industry colleges, which plays an important role in the development of MOOC in the whole world. Take Coursera for an example. By the end of 2015, Coursera has constructed cooperative relationship with hundreds of top colleges and universities worldwide. In China, MOOC courses are launched by Beijing University and Tsinghua University on the platform of edX. Fudan University and Shanghai Jiaotong University also signed a contract with Cousera for MOOC cooperation. Obviously, MOOC develops very fast under the participation of the elite colleges and universities in the world. Under the participation of the elite colleges in the world, does MOOC's fast development illustrate that the MOOC is more scientific and effective than the previous distance education?

Firstly, as it lacks teachers' guidance and communication, the validity and reliability of peer review is difficult to be sustained on a high level. In the process of learning, students conduct learning discussion mainly through course forum and online-learning group. This may cause intermingled academic discussions and cannot ensure to solve learning problems effectively. Additionally, it even leads to a host of impolite online behaviors.

ISSN: 1738-9968 IJHIT

Second, the lack of academic management and supervision during the learning process directly causes high dropout rate [1]. Compared with traditional classroom teaching, the success of distance education decided by teacher's management of learning activities [2]. As to the problems existing in the learning process like using time effectively, dealing with difficulties, individualized learning needs and so on, it is hard for MOOC to give resolutions.

Third, the ways of learning evaluation are simple, which cannot insure its reliability and validity. Now, learning evaluation of MOOC mainly relies on software based exercises and tests, which may help technique-related courses application have a sound effect, while humanity courses cannot gain results as good as technique ones. As the only platform providing a small size of humanity courses, Coursrea adopts the peer review mechanism to grade students' task. But there are still many disadvantages of this mechanism in that most students' abilities are not good enough to point out the real problems in the task, nor to give advices and objective evaluation. In addition, even if there is an argument, it is difficult to discuss and modify in depth because this evaluation is a kind of one-way feedback mechanism.

In conclusion, it is difficult to integrate MOOC's online class learning pattern and learning platform into teaching process of higher education and distance education. However, because of MOOC's advantages of large-scale course resources, excellent guidance material, and fluent video effect and so on, it is vital to explore an effective system of academic management and analysis adapted to higher education. This paper puts forward MOLMA (MOOC-oriented Learning Management and Analysis) under the environment of the colleges and universities. Supervisory mechanism by the teacher is introduced by building academic management system for MOOC courses, which can guide and monitor students' learning. Based on the methods of statistical analysis, machine leaning and graph model with the learning and interactive behavior data, academic analysis model is built to evaluate and predict students' learning trend and results. With the guarantee of the teachers' supervision and academic analysis model, students may obtain the best results by MOOC learning.

2. Related Works

2.1. Research on MOOC

According to different educational ideas and learning models, MOOC can be categorized as two types: cMOOC and xMOOC. cMOOC is the initial form of MOOC based on Connectism, which emphasizes on learners' contribution and feedback to the knowledge network [3]. Learners construct course content together and participate in the study through some topic discussions based on blogs, synchronous network meeting and other approaches. While xMOOC is the main stream of current MOOC which is similar to the traditional classroom learning decided by the course designer. Learners can enroll in courses from various platforms of xMOOC and accomplish learning process through watching lesson video, doing exercises and homework, taking test and other learning methods. The openness and advantages of xMOOC lie in its unlimited class size, no costs, and high-level course resources. This paper conducts the research of MOOC in terms of xMOOC.

So far, research on MOOC are mainly from five perspectives: 1) Introductory studies. giving definitions, explanations and illustrations of MOOC [4]; 2) Conceptual studies. Analyzing the opportunities and challenges for higher education brought by MOOC [5]; 3) Case studies. For example, Rodriguez conducted comparative analysis between MOOC and AI-Stanford courses [6]; 4) Technical studies. Ardis and Henderson did research on software engineering of MOOC [7];

Fini analyzed problems existing in MOOC from the viewpoint of techniques [8]; 5) Education theory research. Cabiria analyzed theories of Connectionism in terms of MOOC [9]; deWaard explored the teaching methodology fit for MOOC [10]. However, there are few researches on MOOC's academic management and analysis.

2.2. Research on Academic Management and Analysis

During teaching process, effective academic management and scientific evaluation are always gained the most concerns. As the development of network techniques, Learning Management System (LMS) had been applied extensively in higher education. More than 90% American colleges and universities and 95% British colleges and universities use LMS [11]. LMS in the early time is on the basis of Web based application system and it is with functions of distributing, managing and searching course materials as well as assisting teacher's job in class at the same time [12]. Currently LMS based on the Internet can provide students learning services of using all kinds of interactive tools, such as communication software, chat room, wiki and blog, and help them to participate in learning activity [13]. All of these enable LMS to promote students' knowledge construction, learning autonomy and creativity, rather than simple knowledge transmission that is more suitable for being applied into distance education [14].

In the field of academic predication, a researcher has carried out many studies. Wang and Liao applied neural network approach to predict students' academic behaviors [15], and Sun explored determining factors in testing by using information entropy approach [16]. From the perspective of cooperation influence and group effects, Schoor conducted the research about students' social behavior in cooperative courses, including interaction between individual student and his or her collaborators and its analysis [17]. In terms of academic indicators, Kerr and Chung studied recognizing students' academic performance indicators and considered those indicators as important basis for teaching strategy design [18]. On modeling academic performance, Palmer showed applying Binary logistic regression models for modeling and evaluating the academic performance of students majored in Science and Engineering [19].

From the above previous research review it can be seen that related works are mainly about traditional teaching in class and blended teaching combining online and offline teaching. And those studies' mainly lay focus on the application of tools provided by LMS to help teachers and students reach the teaching goals, and how to evaluate indicators produced by LMS and LMS' functions. There are few research directly aiming at massive open online class like MOOC.

In total, researchers at home and abroad has conducted a large number of studies on MOOC and LMS respectively, but academic management and analysis for MOOC, especially based on the platform of microblogging, are still waiting to be filled the gap.

3. Academic Management, Analysis and Prediction for MOOC

This paper chooses to integrate the LMS into MOOC academic management and analysis for distance education of colleges and university. Based on MOLMA system's function of academic analysis, it can use the data from academic management system to analyze and evaluate students' learning status in order to identify the students whose learning and cognition are in a high-risk situation and give them educational interference in time. At the same time, the reasons of success and failures in teaching can be analyzed, which helps improve teaching patterns and methods.

The main objective of MOLMA system is to manage MOOC learners in the distance education of colleges and through tracing learners' academic situation to analyze their academic progress and predict their academic performance. Thus it can be seen there are two main functions of MOLMA: a. academic management function; b. academic analysis and prediction function.

3.1. Research on Academic Management and Analysis

There are a great many of tools and materials in the MOLMA system to support learning behaviors, mainly including detailed files and multimedia resources relating to courses learning, evaluation tools with functions of online testing and assignment submission, communication tools including e-mail, forum for chatting and asynchronous discussion, course management tools of recording and storing students grades, ranks, releasing announcement and showing course time, learning management tools of helping students review their course performance and trace learning progress. Its main functions are displayed in Table 1.

3.2. Academic Analysis and Prediction Function of MOLMA System

Through the academic management, MOLMA system can get and store abundant and complicated users' behaviors and interactive data. By analyzing these data, it can trace these variables including online frequency and time, applications of system tools, the number of read and post messages, webpages being visited, *etc*. These data can be caught in real time and can be theoretically used in any stages of the course to dig more information. Collecting these data through MOLMA is non-invasive, and need no manual intervention and obtaining. What is important is that the data may represent many aspects of learners' learning behaviors such as learning pattern, participation extent and pattern they devoted to the learning system, *etc*. Those behaviors are difficult or even impossible to get through other means [11]. The content that can be provided by the learning management system has been presented in Table 1. Through preliminary analysis, abundant academic information can be searched and got from these content and system log of MOLMA. Table 2 lists basic academic indicators this paper focuses on, which can be used to test learning behavior of students and correlation with their academic performance.

4. Methods

In this paper, it employed three research methods including the statistical method, the graph model method and the machine learning method.

4.1. The Statistical Method

The main objective of MOLMA system is to manage MOOC learners in the distance education of colleges and through tracing learners' academic situation to analyze their academic progress and predict their academic performance. Thus it can be seen there are two main functions of MOLMA: a. academic management function; b. academic analysis and prediction function.

Table 1. Main Functions of the Academic Management In MOLMA System

Functions	Function description	Implementation methods and techniques
Syllabus Distribution of course materials	Teacher posts course learning materials	Combination of HTML and P2P shared files
Course resources sharing	Course resources External links resources	File construction and storage, notify users timely to update file by sending messages and RSS
Homework	Students upload their homework and	Transmission of online

	report Teacher grade, check and evaluate students' homework and report, and send them back to students	emails, attachments and files
Group synchronous discussions	Course discussion between teachers and students can be conducted in fixed time and the discussion content can be auto-saved into the system	Online chat room that can realize the text and audio function
Group discussion in asynchronous	Course discussion between teacher and students can be conducted at any time and the discussion content can be auto-saved into the system	Online forum that can realize text and audio function
Testing	Online testing	HTML Files uploading and downloading String comparison, duplicate checking and auto-evaluation
Grading and ranking	Recording and ranking students' grade of online testing to present their academic trend	HTML Database records and operation
WIKI	All online users can edit and maintain knowledge information of concepts and theories. Online users can add learning content. The history memory can be kept.	File writing tool with synergy function
Announcement	Release the announcement and notice to all the online users	Mass mailing BBS Sending system messages
Calendar	Teaching calendar shared with all the users posts teaching progress (*e.g.* assignments submission deadline, the due date of discussion, *etc*.)	Calendar- sharing function
Journal	Record users' online activities in the system	Log records Management and query

Table 2. Basic Academic Indicators

No.	Indicator	No.	Indicator
F1	Online time	F2	Number of read emails & messages
F3	Number of sent emails & messages	F4	Number of read posts in BBS
F5	Number of forwarded posts in BBS	F6	Number of posts replied in BBS
F7	Number of posts initiated in BBS	F8	Number of search applied
F9	Number of study trend applied	F10	Number of "who is online" applied
F11	Time in course chat room	F12	Number of visited web link
F13	Number of finished assignment	F14	Number of read files
F15	Number & time of finished tests		

In order to analyze the relationship between academic indicators and students' grades, the statistical method was applied to carry out the correlation analysis with data collected from the experiment. On the basis of extracting indicators' data of online activities, by using the statistical method, further data analysis can be conducted to evaluate the correlation between these indicators and final grade and/or academic performance trends. Linear combination and deeply integrated indicators can go a step further to reflect this correlation.

Regression modeling approach and correlation coefficient of Pearson are used to calculate the correlation between the academic indicators and academic performance as formula (1) shows. The larger correlation indicator represents the higher correlation. If the significance level is less than 0.05 (α < 0.05 or 0.01), the null hypothesis of the correlation coefficient being zero is rejected.

$$\rho = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}} \tag{1}$$

Cov represents covariance, E is expectation, and σ is standard deviation.

4.2. Naïve Bayes Model

On the basis of analyzing all the academic indicators and its correlation with students' grade, machine learning method, such as Naïve Bayes model can be used to predict and evaluate students' academic performance (normal or trouble) in every stage, so teacher can provide guidance and help in time and accurately for those students who have some troubles in their learning process.

Naive Bayes (NB) is a popular machine learning model for its simplicity [20]. It is easy to be implemented with the linear computational complexity and high accuracy comparing other elaborate learning algorithms.

Here, we give notations for the NB model. In an example dataset $\{(X^{(1)}, y^{(1)})...(X^{(m)}, y^{(m)})...\}$, $X^{(m)}$ denotes a vector containing features of the mth example. The corresponding label is $y^{(m)}$. The spam likelihood $P(y = trouble \mid X)$ is calculated using the Bayesian Formula:

$$P(y = trouble \mid X) = \frac{P(trouble)P(X \mid y = trouble)}{P(X)} \tag{2}$$

Likewise, the ham likelihood is calculated using

$$P(y = normal \mid X) = \frac{P(normal)P(X \mid y = normal)}{P(X)} \tag{3}$$

To model $P(y \mid X)$, xi is conditionally independent for a given y. This assumption is called the NB assumption, and the resulting algorithm is called the NB classifier. We now have

$$P(X \mid y) \quad = P(x_1 \mid y)P(x_2 \mid y)P(x_3 \mid y)...P(x_n \mid y) = \prod_{i=1}^{n} P(x_i \mid y) \tag{4}$$

In academic prediction, there is no need to estimate P(X). The quotient of (1) and (2) is given by:

$$\frac{P(y = trouble \mid X)}{P(y = normal \mid X)} = \frac{P(trouble)\prod_{i=1}^{n} P(x_i \mid y = trouble)}{P(normal)\prod_{i=1}^{n} P(x_i \mid y = normal)} \tag{5}$$

We can use (5) to classify a student being in normal or trouble based on its academic indicators. In (6), $P(trouble)$ is the a priori probability of students in trouble, and is expressed as a frequency in the trouble students category. The calculation formulas are:

$$P(trouble) = \frac{N(trouble)}{N(trouble) + N(normal)} \tag{6}$$

$$P(x_i \mid y = trouble) = \frac{N(trouble \text{ students } includes\ x_i)}{N(trouble)} \tag{7}$$

N *(trouble)* is the number of students in trouble, and N (normal) is the number of students in normal. The normal situation is similar to that of trouble.

4.2. The Graph Model

The graph model is applied to analyze social relationships among students [21]. And it analyzes students' social networks' influence on their academic performance by selecting particular students' social networks. Both the statistical method and the graph method are convenient for teachers to monitor and intervene students' participation and learning process, and to assess the effect of teaching activities.

The interactive networks can also be easily illustrated with a form of graph by selecting and analyzing data from MOLMA, which will clearly show the advantages of interaction in teaching process. Aiming at the interactions among students, and even the networks analysis of the whole system, it first conducted graph based modeling: G=(V, E), V = {e1.1, e1.2, e1.2, …, , e1.150} is the set of nodes in the networks. E = {bi.j} is the edge set of the graph. Each node in this social network graph represent a student or a teacher who participates in the course discussion. A node's relative size stands for "degree"--- the relative number of direct connection made by each user. The degree of a node is an important indicator to describe the interactions among the students. The importance of a node can be evaluated by its degree in a graph model which both describes the distribution of nodes' degrees of all students, and interactions between a student and his/her classmates.

Moreover, the improved PageRank graph model is applied to evaluate various node measurements in the system (formula 8) [22]. Students' social networks and their sociability's influence on academic performances will also be found.

$$P(i) = (1-d) + d \sum_{(i,j)\in E} \frac{P(j)}{Oj} \tag{8}$$

5. Experimental Results and Analysis

To explore the academic management and analysis method under MOLMA, it carried out an empirical study in a university. 200 students took the course of College English were selected randomly. They are from several schools of the university including Applied Science, Electrical Engineering, Materials Science, Law and Arts. The experiment starts from their first year study of College English and finishes at the end of the second year study. The duration of the experiment is four academic terms. The experimental results are as follows.

5.1. Results of the Correlation Analysis

The experimental results show that not all the indicators mentioned in 3.2 have great influence on students' academic performance. The effective indicators in MOLMA can be seen from Table 3.

The results of correlation analysis show that only six indicators have strong correlations with students' grades in MOLMA context, and the rest of indicators bring weak or no effect on students' academic performance. The number of posts initiated in BBS (F7) is responsible for the strongest correlation with students' grades. The coefficient correlation (CRR) is 0.845, and the significance level is approximately zero, which means it has a strong correlation with student' graded. The number of posts replied in BBS (F6) is the indicator with the second strong correlation, which is followed by the number of sent emails & messages (F3). The number of visited web link (F12) takes the fourth rank. The last two indicators that have strong correlation with students' grades are time in course chat room (F11) and the number of forwarded posts in BBS (F5).

Table 3. Results of Correlation Analysis

Indicato rs	Pearson Correlation	Sig. (2-tailed)	Indicators
F7	0.845	0.000	F7
F6	0.843	0.000	F6
F3	0.759	0.000	F3
F12	0.742	0.000	F12
F11	0.732	0.000	F11
F5	0.726	0.000	F5

5.2. Results of Analysis Based on Machine Learning

Aiming at all the academic indicator we drew out and past academic performance of students, prediction model of academic performance using NB model has been built. Academic indicator can be translated into numeric type, academic performance into two classes (normal, trouble). 10-fold cross validation is utilized to validate the prediction model. Precision, Recall and F-measure are used to evaluate the results. Table 4 shows the prediction results of the students based on 6 indicators and all 15 indicators which produced in the MOLMA system.

The results of academic prediction show that the prediction with six indicators is better than that with 15 indicators. The best performance is 97.6% in Precision, 95.2% in Recall and 96.4% in F-measure.

Table 4. Results of Academic Prediction

The number of Indicators	Prediction class	Precisi on	Recal l	F-measure
6	normal	97.6%	95.2 %	96.4%
	trouble	77.8%	87.5 %	82.4%
15	normal	95.4%	92.3 %	93.8%
	trouble	70.2	82.6 %	75.9%

5.2. Results of the Graph Model Analysis

Degree of importance of node can be measured according to the degree of the node in a graph model. Figure 1 describes the distribution of the node degree in a graph model of students in a class. Figure 2 describes the interactive condition between students A and other students specifically.

Because the interactive diagram can presents clearly the participation of the group and individual in the class, teacher can recognize easily from the network which students are participating in the class discussion and which students are drifting away from the learning. With this information, educator can realize intervening measures of learning. Through managing social structure and promoting relation and diversity of the network density, a learning community with active and functionality can be constructed to help students study.

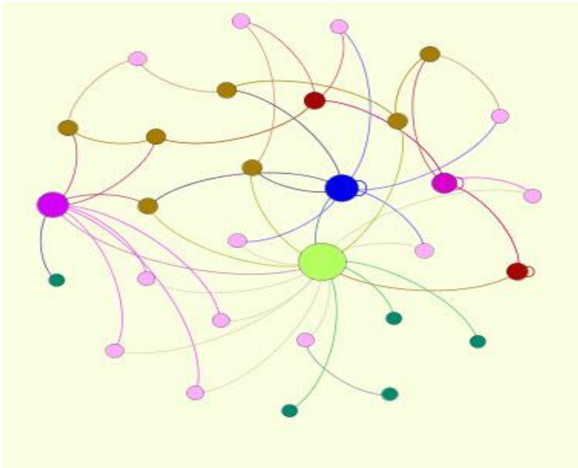

Figure 1. Interaction Diagram of Class Students

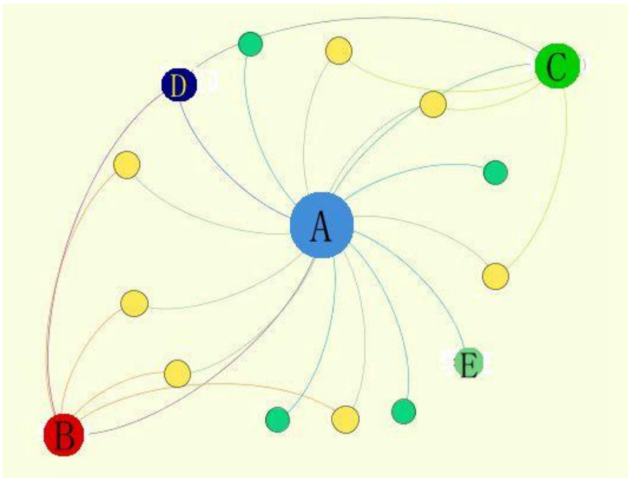

Figure 2. Interactive Diagram Between Students *A* and Other Students

6. Conclusion

Increasing colleges and universities in China are introducing or plan to introduce MOOC course. To guarantee good teaching results, the keys to MOOC education are effectively manage and trace students' learning process, and reasonably evaluate students' progress and academic performance in the teaching process. Therefore it is necessary to conduct research on MOOC's academic management tools and strategies under the environment of higher education. It has a practical meaning to effectively utilize MOOC's excellent course resources, cultivate students' autonomous learning ability, and promote the teaching level of the higher education. In addition, due to MOOC's lack of face to face learning patterns and its over-sized class, it is difficult to well know all students' academic performance in time just depending on examination and tasks in written form and tasks. In this paper, it proposes MOLMA system with functions of academic management, analysis and prediction for MOOC. This system not only supports teacher to manage and participate in students learning activities in MOOC courses, but also can model, analyze and evaluate students' academic performance by applying the data from MOLMA, which helps students benefit from the MOOC.

References

[1] TR. Liyanagunawardena, "MOOCs: A systematic study of the published literature 2008-2012", International Review of Research in Open & Distance Learning, vol.14, no. 3, **(2013)**.

[2] A. Seoane, "Lifelong learning onlinetutoring methodology approach", International Journal of Continuing Engineering Education and Life Long Learning (IJCEELL), vol. 17, no. 6, **(2007)**, pp. 479-492.

[3] J. Daniel, "Making sense of MOOCs: Musings in a maze of myth, paradox and possibility", Journal of Interactive Media in Education, **(2012)**, pp. 1-26.

[4] K. Masters, "A brief guide to understanding MOOCs", The Internet Journal of Medical Education, vol. 1, no. 2, **(2011)**.

[5] B. Mehlenbacher, "Massive open online courses (MOOCs): Educational innovation or threat to higher education?", Proceedings of ACM Workshop on OSDC, New York, vol. 99, **(2012)**.

[6] C. Rodriguez, "MOOCs and the AI-Stanford like courses: Two successful and distinct course formats for massive open online courses", European Journal of Open, Distance and E-Learning, **(2012)**.

[7] M. A. Ardis and P. B Henderson, "Software engineering education (SEEd): Is software engineering ready for MOOCs?", ACM SIGSOFT Software Engineering Notes, vol. 37, no. 5, **(2012)**, pp. 14-16

[8] A. Fini, "The technological dimension of a massive open online course: The Case of the CCK08 course tools", The International Review of Research in Open and Distance Learning, vol. 10, no. 5, **(2009)**.

[9] J. Cabiria, "Connectivist learning environments: Massive open online courses", The 2012 World Congress in Computer Science Computer Engineering and Applied Computing, **(2012)**, pp.16-19.

[10]I. de Waard, "Exploring the MOOC format as a pedagogical approach for mLearning", In Proceedings of 10th World Conference on Mobile and Contextual Learning, **(2011)**, pp.138-145

[11] S. Lonna, D. Stephanie and B. Teasley, "Saving time or innovating practice: Investigating perceptions and uses of Learning Management Systems", Computers and Education, vol. 53, **(2009)**, pp. 686-694

[12] R. West, "Understanding the experiences of instructors as they adopt a course management system", Educational Technology Research and Development, vol. 55, no. 1, **(2007)**, pp. 1-26

[13] T. McGill and J. Klobas, "A task–technology fit view of learning management system impact", Computers & Education, vol. 52, **(2009)**, pp. 496-508

[14] P. Coaldrake and L. Stedman, "Academic work in the twenty-first century. Occasional paper series, higher education division", Australian Government Publishing Service, **(1999)**.

[15] Y. H. Wang and H. C. Liao, "Data mining for adaptive learning in a TESL-based e-learning system", Expert Systems with Applications, vol. 38, no. 6, **(2011)**, pp. 6480-6485

[16] X. Sun, "Finding dependent test items: an information theory based approach", In Proceedings of the 5th international conference on educational datamining, **(2012)**, pp. 222–223

[17] C. Schoor and M. Bannert. "Exploring regulatory processes during a computer supported collaborative learning task using process mining", Computers in Human Behavior, vol. 28, no. 4, **(2012)**, pp. 1321-1331.

[18] D. Kerr and G. Chung, "Identifying key features of student performance in educational video games and simulations through cluster analysis", Journal of Educational Data Mining, vol. 4, no. 1, **(2012)**, pp. 144-182.

[19] S. Palmer, "Modelling Engineering Student Academic Performance Using Academic Analytics", International Journal of Engineering Education, vol. 29, no. 1, **(2013)**, pp.132—138.

[20] A. Ng and M. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes", presented at the Advances in Neural Information Processing Systems, Canada, **(2002)**.

[21] W. Aiello, F. Chung and L. Lu, "A random graph model for massive graphs", Proceedings of the thirty-second annual ACM symposium on Theory of computing. Acm, pp. 171-180, **(2000)**.

[22] M. Bender, T. Crecelius and M. Kacimi, "Exploiting social relations for query expansion and result ranking. Data engineering workshop", IEEE 24th International Conference on, **(2008)**, pp. 501-506.

