Int. J. Advance Soft Compu. Appl, Vol. 7, No. 3, November 2015
ISSN 2074-8523

An Approach for E-Learning Data Analytics using SOM Clustering

Nor Bahiah Ahmad[1], Mohd Khairulanwar Ishak[2], Umi Farhana Alias[3], and Nadirah Mohamad[4]

[1,2,3,4]Faculty of Computing
Universiti Teknologi Malaysia, 81310, UTM Skudai, Johor
e-mail: bahiah@utm.my, mohd.khairulanwar.ishak@gmail.com,
umifarhana.alias@gmail.com, deera36@gmail.com

Abstract

Within the field of e-learning, the accessibility of large amount of learning materials in web-based education systems often burden students to get the learning materials that they preferred. Recent study in e-learning has intensified the need for adapting students' behaviour and knowledge level in presenting learning materials to the students. However, the heterogeneous of students' behaviour, the diversity of learning materials and the complexity of hyperlinks for navigations become the limitation attributes of an adaptive learning environment system. Hence, there is a need to develop techniques to analyse the students' data in order to understand the student's behaviour in order to organize and maintain the learning materials in the domain model repository in the system. In this research, we propose a framework for adaptive learning environment by using self-organizing map (SOM) clustering approach for learner model data analytics and structuring the domain model content in order to present the learning content that suits with the student's need and to achieve higher performances in learning. This research contributes to big data analytics in e-learning which currently is the new focus among educators as the learning data grows tremendously every day.

Keywords: *E-learning Big Data, Data Analytics, Adaptive Learning, Self-Organizing Map (SOM), Learner Model, Domain Model.*

1 Introduction

The recent advancements in technology have produced big data and become the necessity for researcher to analyse the data in order to make it meaningful [1] and to reveal more patterns of the data [2]. In this regard, big data analytic is needed in order to provide techniques to analyse the data. However, big data analytic need faster, efficient [3] and more scalable solutions to store and process all this data, instead of using traditional data warehousing approaches that basically didn't work in this situation. Many research projects have been endorsed to find solutions in delivering the analysis of the data effectively and efficiently.

E-learning refers to using electronic applications and processes to learn. E-learning applications and processes include web-based learning, computer-based learning, virtual classrooms and digital collaboration [4,5]. Variety of features in E-learning systems gives flexibility for instructors to create, manage and administer online courses. Instructors are able to develop and deploy variety of activities such as quizzes, forums, examples, assignments and many other features. E-learning system also facilitate administrative issues such as monitoring the learners' progress and performance, students' enrolment, and grading [6,7].

Most education institutions had incorporated e-learning into educational process in order to increase teaching quality and to distribute the knowledge efficiently. The situation caused the amount of data coming out from these institutes multiplying considerably. These data which comes from the student's personal information, test scores, training materials, video lectures and others create the term 'Big data' in education and will be possible to have effects on the future of education.

In e-learning, the realization of individualized learning has increased the popularity of adaptive technology in e-learning. Currently, the lack of visual cues on the student's status and performance in the online environment cause students to feel isolated in hyperspace. Adaptive learning system is an interactive system which adapts the learning content, pedagogical model and interactions in order to cater the learner's unique needs with aims to improve the learner's performance [8]. E-learning with adaptive learning technologies can help students with different skill level and struggle with learning online feel comfortable in the learning environment that suit their preferences and knowledge level.

In this research, we propose a framework for adaptive learning environment by implementing learning analytics using SOM for clustering learners based on their learning behaviour in e-learning. The clustering result will be used to identify and deliver the right learning materials stored in the domain model that personalize with the students' needs and hence enable students to achieve higher performance.

2 E-Learning and Big Data

Big data in E-learning environment is not new. Since the rise of Internet, E-learning has become an essential medium in delivering education. In e-learning environment, every learner's actions and interactions with the system during the learning process, such as engaging in online assessments, involve in forum discussion and involve in other e-learning activities will produce data that are recorded in the log file. As more activities are captured in the log files, the number of data in the log files will increase and generate excessive load of information. These data are very useful in education and therefore one of the major issues in educational data is how to analyse the data in order to get meaningful information that able to improve educational approach.

E-learning which has astounding large user's data are limited in data exploitation which can cause delay in user action and thus, provide opportunities to assist the solution for the problem [9]. Ignoring this issue can give disadvantages to educational institutions or instructors. For example, without analytical of educational data, the educational institution or instructor will not know that there are students with decreasing performance, courses that are irrelevant to students and students who need different type of resource supports. The problem can be overcome or avoided by analysing the data.

There are advantages of advance research and analysis of big data for educational institution and instructor. Therefore, involving them in big data discussion is important. Research has concluded that making decision based on data-driven can improve organization and its productivity [10]. The benefits allow them to perform better in teaching and learning process.

In [4], the author highlights some significant advantages associated with e-learning big data whereby by analysing learner data, e-learning professionals will have information about the learners. Table 1 shows studies in e-learning big data in various areas that includes skills development to analytic and assessments.

Education on big data especially skills and course development are foundation to big data implementation. Zu and Zuo [11] and Piliouras [12] explore the importance of big data skills among students to setting up universities with big data speciality, while several researchers, such as [5,13,14] proposed big data course to be conducted as an effort to cater issue on lack of professional in data analysis. In other to predict student's career and readiness, AbuKousa and Atif [15] proposed a portal application which analysing big data of students higher education's lifetime.

Table 1: Studies on big data e-learning

Area	**Author**	**Studies**
Skills and course development	[11, 12]	Big data analytic skills and course enhancement
	[5,13,14]	Propose big data course
	[15]	Proposed portal application in other to predict student's career and readiness
Governance	[16]	Governance of the big data technology in education
Medium	[17]	Study on available MOOC provider and explained the differences between them
Analytic and assessment	[18,19]	Provides review on tools, techniques and application of big data in education
	[20]	Proposed a technique to continuously capture data of student's interaction
	[21,22]	Present assessment using peer assessment
	[23]	Social network analysis parameter with students outcome
	[24]	Rubric in multimodal assessment

Self, [16] covers on governance of the big data technology in order for the data to be safely integrated into infrastructure. Meanwhile, Massive Open Online Course (MOOC) which is now the new trend of learning is synonym with big data. Pang, [17] studied on several existing MOOC provider and explained the differences between them. The examples of MOOC providers include Coursera, Udacity and Edx. Compare to other providers, most Edx courses are free while courses in Udatacity are from industry and provided according to the learner’s level.

West and Siemens [18,19] provide reviews on tools, techniques and application of big data in education which prominent analytic techniques include modelling, relationship mining and knowledge domain modelling. While application involves trend analysis and prediction, personalisation or adaptive learning and structural analysis. In a study on MOOC, Govindarajan [20] proposed a technique to continuously capture data of student's interaction. Admiraal [21] and O'Toole [22] presented experiment using peer assessment in order to conduct assessment for learning, Hernández-García [23] using social network analysis by assessing the students’ interaction to predict academic performance, Burnett [24] using rubric for multimodal assessment. Those are methods that have been proposed and presented in analysing big data. From the study done, it can be concluded that the studies are covered in several areas which are skills, governance, e-learning medium and e-learning data analytic. These areas are important in implementing any e-learning big data analytic research or project. Skills and courses on big data are needed in providing resources for analytic, governance is important in ensuring the flow of data processing, medium of data is needed as a medium for data resources and lastly, analytic is important in transforming data into meaningful value.

3 Learning Analytic

Learning analytics (LA) is the process of discovering learning patterns among learners. The analytic usually involves variety of techniques including machine learning techniques to mine the data. It involves the process of select and capture data, aggregate and report, predict, use, refine and share [26]. In the early phase of learning analytic, select and capture relevant data from e-learning database need to be done [27]. Next, data pre-processing and processing is conducted that include data aggregate and producing report. The tools or methods need to be selected wisely to fit the data type. From the process, come out the application of several techniques on the result which is useful for prediction [27]. The analytic result is shared with learning centre for strategic decision making and to inform students of their progress. Those processes are enhanced with big data engine to process large data efficiently. From other perspective, the process involved are to make the raw data to reach wisdom level as depicted in knowledge continuum [26].

Most of the data that used in LA applications comes from LMS either from the login information, rates of participation in specific activities, assessment results or grades and time students' interaction with online resources. The knowledge gained from e-learning data can help in improving the adaptive learning environment [10]. LA can assist by providing a more personalized learning experience through the use of historical and current data to respond to students' needs.

Several benefits of learning analytics that able to motivate researchers to consider the importance of data analysis in e-learning are as follows:

- Help instructor understand how learners assimilate. The understanding levels vary among students. Learning analytics can assist instructors to know learners' learning style. The instructor then can use the information to develop better course materials based on specific student's learning processes.
- Facilitate instructor using instant data [28]. In e-learning, analysing instant data such as quiz and test results, total of students taking a course and participation records, can facilitate the educator develop model to measure each student's participation levels compare to their grades. This information helps them to adjust their schedule, approach and teaching process to ensure students' participation in activities.

- Predictive analysis give benefit to instructor who concern more with personal development [29]. They can analyse the data results such as to know which course they need and seeking professionals for educational fields. The development programs can improve both the instructor's performance and income.
- Personalize learning through learning analytic. The analytic can be used in identifying isolate students (who appear to be left behind than other students) or student who appears to have poor result which less likely to succeed academically. Also, from the analysis result, instructor able to see opportunity to intervene in other to assist and encourage them for better outcomes. Other than that, instructor can modify curricular or learning activities that fit to improve learning.

Based on the benefits listed above, it is thus very important for instructors to analyse e-learning data in order to improve learning performance among students and also to improve the teaching approaches.

4 SOM Clustering

Self-Organizing Map (SOM) is a type of Artificial Neural Network (ANN) architecture that were introduced by [30] to convert high-dimensional input to low-dimensional space. It has been widely used in industry, finance, natural sciences, linguistics, massive textual database and bioinformatics data [31] and [32] but least in clustering for educational data. Clustering by using SOM can produce certain output pattern [33] that can be used in improving future teaching and learning process. The advantages of this method are based on implementation simplicity, execution speed and a shorter training process using unsupervised learning better than other neural network [34] which are advantages for big data. Processing large data set efficiently becomes one of the main challenges of big data. The steps to process SOM clustering are:

i. Create the initial value of input vector by using random-based, input-based or linear initialization.

ii. Select the input vector depend on the type of parameter

iii. Identify the winning node by using finding the Best Match Unit (BMU)

iv. Identify the radius of neighborhood node based on winning node

v. Weight updated for winning node

vi. Step repetition (iii. to v.) until maximum iteration

Data mining using SOM clustering is an analytical tool to analyse the student's clustered group based on the behaviour pattern while using the E-learning system. Students are freely use the system anytime and anywhere to access the system and browsing the learning material prepared by the educator. In this study, Waikato Environment for Knowledge Analysis which is popularly referred as Weka is used as a data mining tool. Weka is a free software available under the GNU General Public License and was developed at University of Waikato [34].

5 Proposed Adaptive Learning Framework

Fig. 1 illustrates the system architecture of the proposed adaptive learning framework which is adapted from [35]. The main purpose of this framework is to enable students to navigate the learning materials intelligently based on the adaptation features. The architectural design of the proposed framework is composed of four main modules: the adaptation model, the learner model, the domain model and the analytic component. In e-learning environment and big data, this framework facilitates two main issues, which are taking the advantages of valuable log data to understand student's behaviour toward online learning. Moreover, this framework proposed a new technique for organizing variety types of learning materials in the domain model and presenting to the learner with preferred learning materials by considering two adaptation features, which are the student's learning style and knowledge level.

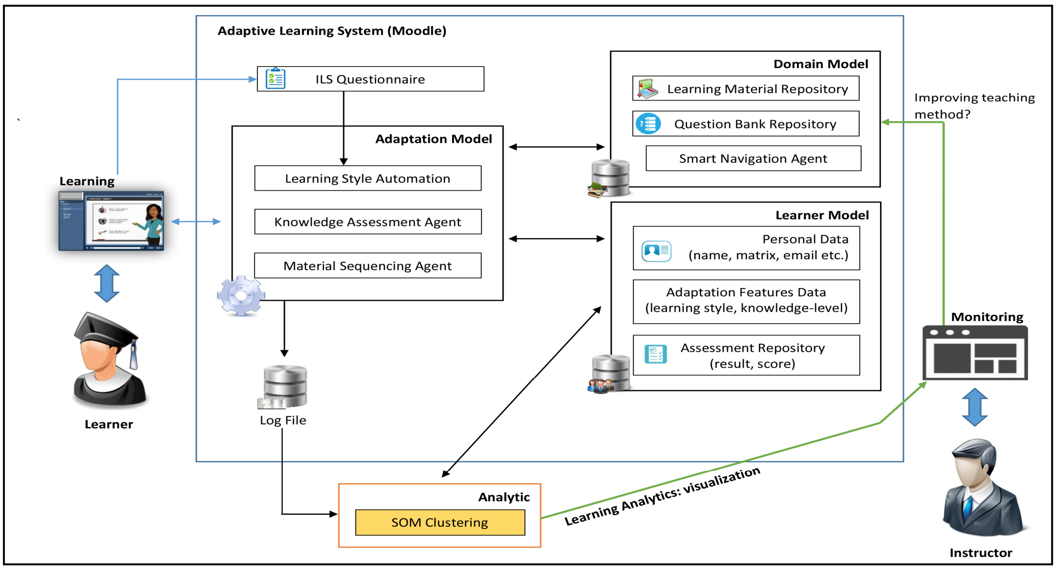

Fig. 1: Framework of the adaptive learning system using Moodle

1.1 Adaptation model

The adaptation models consist of three components, which are the learning style automation, the knowledge assessment agent and the material sequencing agent. Initially, when the learners login in the system for the first time, they need to answer the Index Learning Style (ILS) questionnaire adopting from [37]. Based on the answers given by the learner in the questionnaire, the learning style automation will identify the learning style of the learner and stored in the learner model repository. The knowledge assessment agent will operate once the learner finished doing assessment such as exercises, quiz or test. The score obtained from the assessment will be used induct the student's knowledge level either as beginner, intermediate or advanced learner. This component will identify the learners' knowledge level iteratively during the learning process. Both components interact continually with a learner model to adapt learner behaviour. The material sequencing agent deals with generating learning path and relevant material which influencing the adaptation features of the learner. This component also will be used to provide assessment and suggested materials for the learner as additional information for them. Currently, Moodle provide a feature to record every action perform by the learners during learning and assessment activities and stored the information in the log file. Those valuable data will be used by the analytic component to analyse learner's browsing behaviour.

1.2 Domain model

The domain model performs as repository for storing curriculum and knowledge about the domain. The model is developed on a concept of graph and hierarchical of nodes and arcs. The node represents the knowledge concepts and arc represents the link between nodes. The instructor or course coordinator can direct access to the domain model to insert and modify the learning materials. The domain model framework consists of dependencies configuration, repository of domain model and content adjusting principle.

For each learning material, the instructor needs to specify the learning level and the difficulty level of the learning materials as shown in Fig. 2. This information will be used for adaptation model to propose suitable learning materials to the learner. This additional configuration is implemented to the current Moodle platform in order to enhance the environment to support adaptability features.

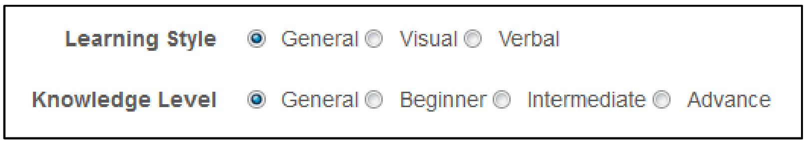

Fig. 2: Dependencies configuration

The organization of the content in the repository is structured in hierarchical order, starting with the courses and each course associates with the chapters. Each chapter composes of topics and each topic will link to the learning objectives and assessments. Every learning material will have a set of different resources associated with it. Depending on the learning objectives, the cross-reference resources may exist such as glossary, definition, or terms by using a URL. Fig. 3 depicts repository structure in organizing knowledge domain in the domain model.

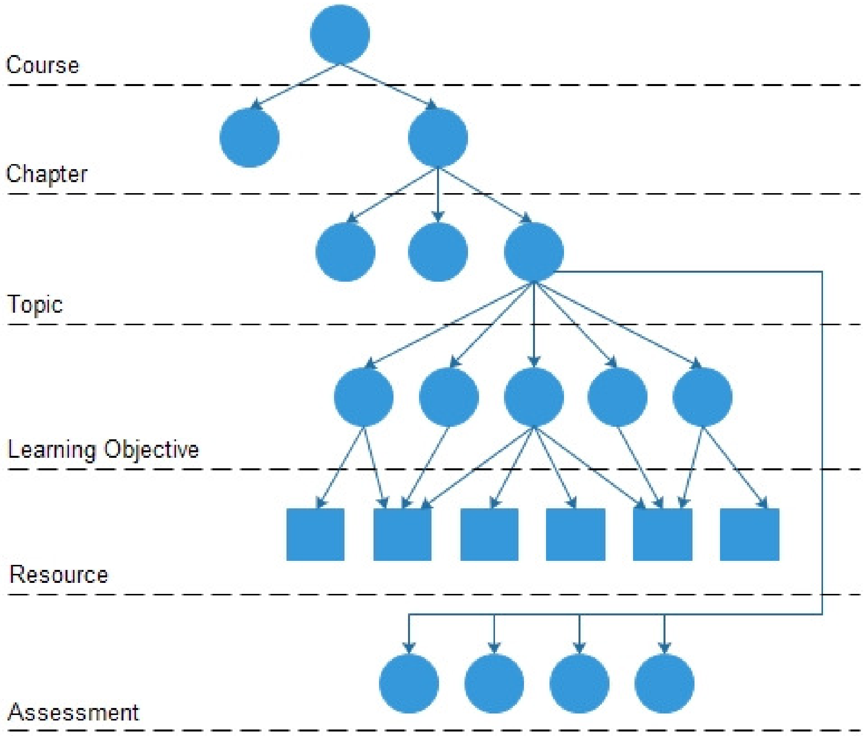

Fig. 3: Repository of domain model

The content adjusting principle is designed based on the characteristics of the adaptation features; learning style (visual and verbal) and knowledge level. The knowledge strategies of the proposed system are given in Table 2. The components of subject unit are classified into the following six categories:

- Text component: text contain in the subject unit
- Graphic component: the illustrative including video, image and figures
- Example component: the sample or additional information
- Basic component: component that contain basic contents of the subject; title, concept and corresponding text, graphic, example and assessment
- Additional component: contain supplementary resources that are helpful to student in extending their understanding and learning scope.

The smart navigation agent is used to guide the learner to traverse the various links and page of the system. The agent also is responsible in presenting all the materials in a structured way in the presentation layout framework.

Table 2: Content adjusting principle

		Learning Style	
		Visual	Verbal
	Beginner	Text: Basic	Text: Basic +

		Graphic: Basic + Additional Example: Basic Assessment: Quiz	Additional Graphic: Basic Example: Basic + Additional Assessment: Quiz
	Intermediate	Text: Basic Graphic: Basic + Additional Example: Basic Assessment: Quiz + PS	Text: Basic + Additional Graphic: Basic Example: Basic + Additional Assessment: Quiz + PS
	Advance	Text: Basic Graphic: Basic + Additional Example: Basic Assessment: PS	Text: Basic + Additional Graphic: Basic Example: Basic + Additional Assessment: PS

1.3 Learner model

The learner model represents information about the learner. In this model, three strands of information are maintained, which are student's personal details, adaptation features data and assessment score information. Among the personal details stored are the student's name, matrix number, email address, current CGPA, etc. The adaptation features data include type of learning style (visual or verbal) and knowledge-level (advanced, intermediate or beginner). The assessment repository stores the results and scores of each quiz and test conducted by the students. The information of learning style and knowledge-level will be updated automatically by the system and highly dependent on the results of ILS questionnaire answered by the learner and the assessment score obtained from the test or quizzes.

1.4 Analytic component

The last module in this framework is the analytic component. The purpose of this module is to offer more quantifiable evidence for instructor to provide and prepare personalization and to help decision making on the student's status and behavior. In this proposed framework, we applied SOM data mining technique to identify student browsing behaviour from the data captured in the log file. SOM will use the historical data from log file and the personal information from learner model as an input to analyse. The navigation agent will use the results generated by the analytic component to improve learning method and to provide additional materials for specific learner if necessary.

6 Experimental Setup

In this research, a preliminary experiment has been done by using data mining technique to analyse the browsing behaviour while using E-learning system. The process of experimental setup for this study are described as follow:

1.5 Data collection and pre-processing

The data set used in this study was obtained from UTM Moodle E-Learning environment log file. The history of student's action is collected from Data Structure and Algorithm subject (SCSJ2013). There are 19 students involved from the section 3, semester 1 for 2014/2015 session. The recorded data are from first week until week 14^{th} where the teaching and learning process occur. Data is downloaded from e-learning system in the excel format. Fig. 4 shows a sample view to download the data of log file from the e-learning system. There are choices to choose the information to download and for this study require the data for all students, all days and all the action recorded in the log file history.

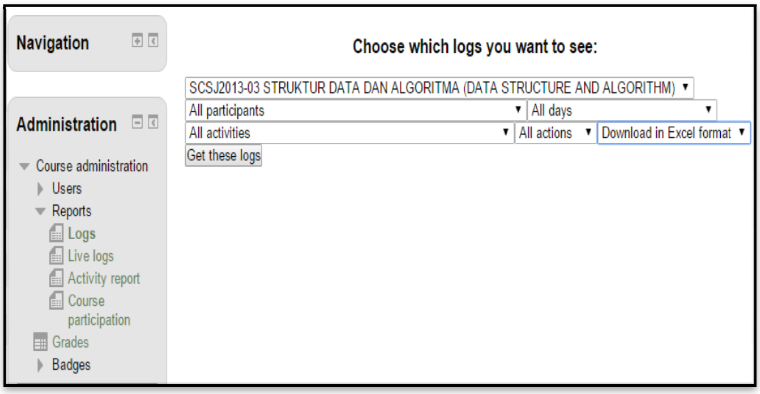

Fig. 4: A Sample view to download E-learning log file

Hence, the data collected contain several information such as the course name, the time accessed, the IP address, the user id, the user actions and the information for each student as shown in Fig. 5. Data need to be cleaned by removing the noise data and data that are not related to this study such as the action of the admin, educator and other users that used the system for this subject. Therefore, the data is filtered so that only the data related to the students is used. In the action information column contain the information of the module prepared by educator for learning material such as assignment module, forum module, resource module and others. Furthermore, this study focuses on data from course module, assignment module and resource module only.

3	Course	Time	IP addres	User ID	Actioı	Information
4	SCSJ201:	####	10.60.84.1	stud1	assig	Submission statement accepted by user stud1
5	SCSJ201:	####	10.60.87.3	stud1	assig	Submission status: Submitted for grading. The n
6	SCSJ201:	####	10.60.84.1	stud1	assig	Submission status: Submitted for grading. The n
7	SCSJ201:	####	10.60.84.1	stud1	assig	Submission status: Submitted for grading. The n
8	SCSJ201:	####	10.60.100	stud1	assig	Submission status: Submitted for grading. The n
9	SCSJ201:	####	10.60.84.1	stud1	assig	Submission status: Submitted for grading. The n
10	SCSJ201:	####	10.60.84.1	stud1	assig	Submission status: Draft (not submitted). The nu
11	SCSJ201:	####	10.60.86.1	stud1	assig	Submission status: Submitted for grading. The n
12	SCSJ201:	####	10.60.98.3	stud1	assig	Submission status: Submitted for grading. The n
13	SCSJ201:	####	10.60.98.3	stud1	assig	Submission status: Submitted for grading. The n
14	SCSJ201:	####	10.60.85.6	stud1	assig	Submission status: Submitted for grading. The n
15	SCSJ201:	####	10.60.84.2	stud1	assig	Submission status: Submitted for grading. The n
16	SCSJ201:	####	10.60.84.1	stud1	assig	Submission status: Submitted for grading. The n

Fig. 5: A sample view data of log file

Therefore, the data is prepared by creating categories of action for 14 weeks teaching and learning process as shown in Table 3. There are nine actions (variable) for each week to be used to analyse the student's action while browsing the learning materials.

After the cleaning and filtering process, from 8503 raw data recorded, the remaining data used in this study are 2356 hits of the student's action while browsing the E-learning system. Fig. 6 shows the sample data sets that have undergo data pre-processing. The hits recorded in the datasets are based on the actions done every week for 14 weeks. Next, SOM clustering will be applied to the datasets and describe in the next subtopic.

The e-learning data extracted from the log file contains user's interaction history is prepared in ARFF format. The normalization process need to be done before SOM clustering can be conducted. Fig. 7 shows the visualization from the clustering process based on the clustered group for action viewing and downloading notes in week 1. The result from the experiments will be discussed in the next section.

Table 3: The attributes filtered from E-learning

Module	Variable	Action	Explanation
Course	courseV	Course View	Contains the information of the subject, resource display on learning environment
Assign	assignV	Assignment View	Allow to view the information about the assignment
	assignS	Assignment Submit	Allow student to submit the works as request
Resource	asgI	Assignment Individual	Allow student to view and download the assignment materials and work it individually
	asgG	Assignment By Group	Allow student to view and download the assignment materials and work it by group
	Exe	Exercises	Allow student to view and download the

			exercise for certain topics
	P_exam	Test Materials	Allow student to view and download the test materials for examination preparation
	Notes	Notes	Allow student to view and download the notes for each topic
	Example	Example	Allow student to view and download the example for certain topic

ID	courseV1	assignV1	assignS1	asgl1	asgG1	Exe1	P_exam1	Notes1	example1
stud1	3	3	1	0	0	1	0	3	0
stud2	4	4	2	0	0	0	0	3	0
stud3	5	13	3	0	0	3	0	10	0
stud4	19	6	1	0	0	1	0	12	0
stud5	3	3	1	0	0	0	0	3	0
stud6	3	1	0	0	0	0	0	1	0
stud7	9	1	0	0	0	0	0	0	6
stud8	14	7	2	0	0	0	0	5	0
stud9	9	4	1	0	0	1	0	11	2
stud10	15	4	1	0	0	3	0	8	0

Fig. 6: A sample view of log file data for SOM clustering

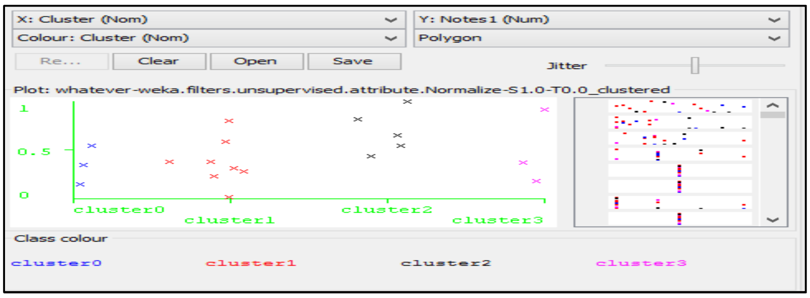

Fig. 7: The visual cluster of assignments using SOM

1.6 Result and discussion

As mentioned in section 6.1, the data used in this study are from UTM Moodle E-Learning environment data captured in the log file for 14 weeks of teaching and learning. The data sets contain the sum of hits for 19 students based on the attributes and week. There are 126 attributes to be analysed. The clustered groups are based on the attributes for 14 weeks. Based on the cluster analysis, there are four clusters formed for 19 students which are 3 students in cluster0, 8 students in cluster1, 5 students in cluster2 and 3 students in cluster3 as shown in Fig. 8. The clustering result were obtained from SOM unsupervised learning. Meanwhile, Fig. 8 shows the percentage of hits of action for each cluster. It shows that for the whole semester, students in cluster 2 actively browsing the materials by 72% hit, followed by cluster1 (70%), cluster3 (64%) and lastly cluster0 (57%). Although cluster1 have the highest number of students compared to other clusters, but as shown in Fig. 8 cluster2 have high value of browsing behaviour in most of the attributes for 14 weeks. Students in cluster2 likes browsing and downloading the example from resources, do more actions in course view for the subject and active in the assignment module.

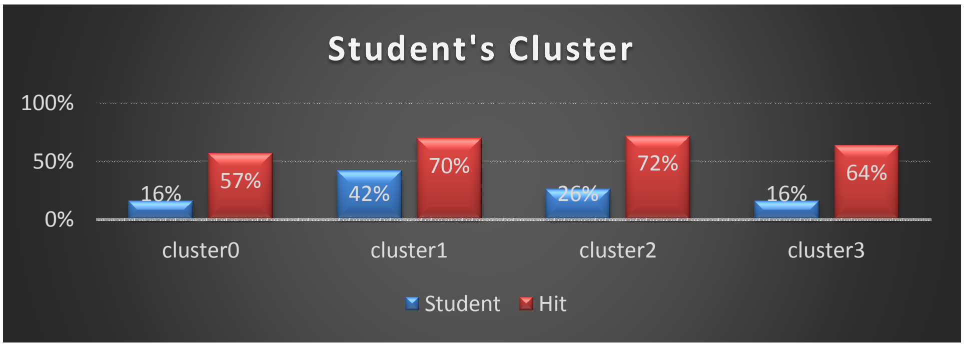

Fig. 8: Student's cluster based on SOM clustering

Fig. 9 shows the sample of cluster analysis for week 1 of the semester. At the beginning of the semester, each cluster shows that there are browsing and downloading activities for certain attributes and cluster2 shows the students in the cluster actively do some course view and downloading the notes. For the first week of teaching and learning process, students in each cluster mostly accessed the course view activity to get an overview of the planning of this subject. Usually, during the early week of the semester, the lecturer gave not much activity, so the students simply download the notes for the class preparation. The value 0 shows that there are no hit for the attributes in the cluster among the students. Moreover, there are no assignments or exercises to submit in the first week.

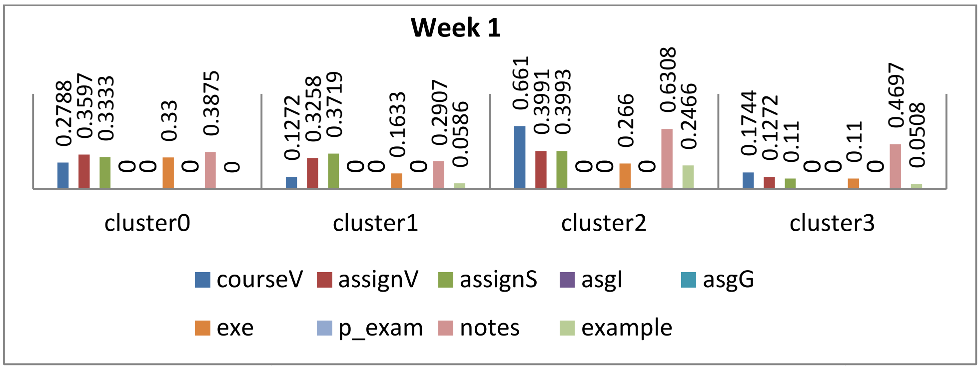

Fig. 9: Student's cluster based on browsing behaviour for week 1

Fig. 10 shows the cluster analysis for week 8. In week 8 the students need to sit for an exam to add up the carry marks for their grade at the end of semester. The result shows that there a slightly changes in student's behaviour, whereby the students are more focus on preparing for examination by doing more reviews in p_exam's action. Students in cluster3 do not have any hits for assignment modules as they are more focus on the exam preparation.

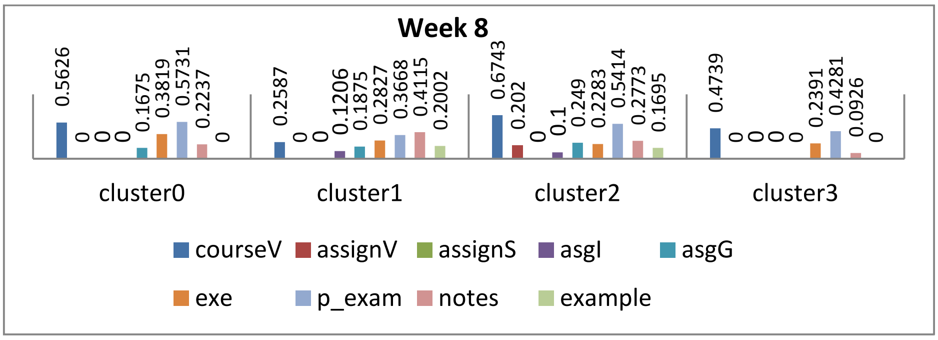

Fig. 10: Student's cluster based on browsing behaviour for week 8

Fig. 11 depicts the cluster distribution for week 14. In week 14, which is the last week of the semester, the hits on the example are the highest hits compared to others activity. However, the hits for other action are low. The result shows that toward the end of the semester, the students are eager to view the examples for their examination preparation.

Overall the SOM clustering results show that each group of cluster has different browsing pattern towards the resource materials in the E-learning environment. In week1, students like to have an overview of the subject and only download the notes compared to other actions. Throughout the semester such as in week8, students perform different actions and focus more on exam preparation during test 1. For the last week of teaching and learning process (week 14) students activities were slow down by having more reviewing in the example modules in order to increase their understanding of the learning materials during the final examination. From the experimental result, it shows the variety of student's behaviour from the beginning until the end of semester. Thus, it can help the educators to predict the future student's performance by preparing the learning material based on the browsing behaviour. For example, educators can provide more examples and exercises at the end of the semester to help student in their final examination.

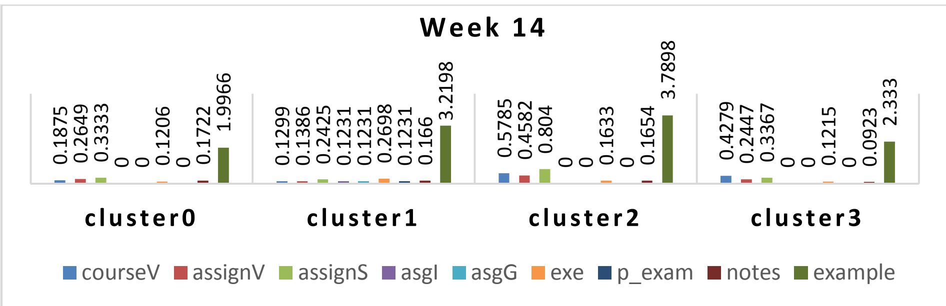

Fig. 11: Student's cluster based on browsing behaviour for week 14

7 Conclusion and Future Works

This paper has introduced a framework for presenting learner with suitable learning materials according to adaptation features, which are learning style and knowledge level. This framework enhances the current architecture in Moodle to support adaptive features. Analytic component is added to analyse the learner model data. SOM clustering able to show that there exists various clusters based on the student's hits in Moodle. The students behave differently in various weeks of the semester. By analysing the current behaviour and knowledge level of the student, it is hoped that this research will provide valuable solution in the education process, whereby the student may achieve higher understanding of their learning outcome by getting the suitable and relevant learning material according to their learning style and knowledge level.

This research has found several area that has potential for further research which to implement the analytic with larger dataset using big data engine. The implementation then can process live data directly and transfer the result for visualization, which will be simplified through dashboard technology. In the future, the student behaviour patterns identified by the analytic component will be integrated with real-time data, such as the student's performance and interactions with LMS. In this way, we can identify at-risk students who are not on-track and designing data visualization for student to notify their current status in the learning.

ACKNOWLEDGEMENTS

This paper are supported by Fundamental Research Grant Scheme (FRGS), vote number 4F496 and Flagship Grant, vote number Q.J130000.2428.02G69. The authors would like to express our deepest gratitude to Research Management Centre (RMC), Universiti Teknologi Malaysia (UTM) for the support in R & D, and Soft Computing Research Group (SCRG) for the inspiration in making this study success.

References

[1] Gupta, B., Goul, M., & Dinter, B. (2015). Business Intelligence and Big Data in Higher Education: Status of a Multi-Year Model Curriculum Development Effort for Business School Undergraduates, MS Graduates, and MBAs.*Communications of the Association for Information Systems*, *36*(1), 23.

[2] Erwin, K., Bond, M., & Jain, A. (2015). Discovering the Language of Data: Personal Pattern Languages and the Social Construction of Meaning from Big Data. *Interdisciplinary Science Reviews*, *40*(1), 44-60.

[3] Dittrich, J., & Quiané-Ruiz, J. A. (2012). Efficient big data processing in Hadoop MapReduce. *Proceedings of the VLDB Endowment*, *5*(12), 2014-2015.

[4] A. Ashraf, H. El-Bakry, S. M. A. El-razek, and Y. El-Mashad. (2015). Handling big data in e-learning. *International Journal Advanced Research in Computer Science & Technology (IJARCST 2015), 3(1)* on (pp. 47–51).

[5 Okur, M. C., & Buyukkececi, M. (2014, May). Big data challenges in information engineering curriculum. In *EAEEIE (EAEEIE), 2014 25th Annual Conference* (pp. 1-4). IEEE.

[6] Romanowski, C. J., & Raj, R. K. (2013, October). Catching the wave: Big data in the classroom. In *Frontiers in Education Conference, 2013 IEEE* (pp. 405-406). IEEE.

[7] Graf, S., & Liu, T. C. (2010). Analysis of learners' navigational behaviour and their learning styles in an online course. Journal of Computer Assisted Learning, 26(2), 116-131.

[8] Beldagli, B., & Adiguzel, T. (2010). Illustrating an ideal adaptive e-learning: A conceptual framework. *Procedia-Social and Behavioral Sciences*, *2*(2), 5755-5761.

[9] Bloom, C. P., & Loftin, R. B. (Eds.). (1998). *Facilitating the development and use of interactive learning environments*. CRC Press.

[10]Hollands, F. M., & Tirthali, D. (2014). Why do Institutions Offer MOOCs?.*Online Learning*, *18*(3).

[11]Zhu, Y., & Zuo, J. (2014, August). Probe into setting up big data processing specialty in Chinese universities. In *Computer Science & Education (ICCSE), 2014 9th International Conference on* (pp. 566-569). IEEE.

[12]Piliouras, T., Yu, P. L., Fei, Y., Zeng, Y., Chen, J., Wu, R., Yan.R and Needham, N. (2014, October). Finding gold (star students) in the data mine. In Emerging Technologies for a Smarter World (CEWIT), 2014 11th International Conference & Expo on (pp. 1-6). IEEE.

[13]Wlodarczyk, T. W., & Hacker, T. J. (2014, December). Problem-Based Learning Approach to a Course in Data Intensive Systems. In *Cloud Computing Technology and Science (CloudCom), 2014 IEEE 6th International Conference on* (pp. 942-948). IEEE.

[14]Demchenko, Y., Gruengard, E., & Klous, S. (2014, December). Instructional Model for Building Effective Big Data Curricula for Online and Campus Education. In *Cloud Computing Technology and Science (CloudCom), 2014 IEEE 6th International Conference on* (pp. 935-941). IEEE.

[15]AbuKhousa, E., & Atif, Y. (2014, December). Big learning data analytics support for engineering career readiness. In *Interactive Collaborative Learning (ICL), 2014 International Conference on* (pp. 663-668). IEEE.

[16 Self, R. J. (2014, December). Governance Strategies for the Cloud, Big Data, and Other Technologies in Education. In *Utility and Cloud Computing (UCC), 2014 IEEE/ACM 7th International Conference on* (pp. 630-635). IEEE.

[17]Pang, Y., Wang, T., & Wang, N. (2014, August). MOOC Data from Providers. In *Enterprise Systems Conference (ES), 2014* (pp. 87-90). IEEE.

[18]Siemens, G. (2013). Learning analytics: The emergence of a discipline.*American Behavioral Scientist*, 0002764213498851.

[19]West, D. M. (2012). Big data for education: Data mining, data analytics, and web dashboards. *Governance Studies at Brookings*, 1-10.

[20]Govindarajan, K., Somasundaram, T. S., & Kumar, V. S. (2013, December). Continuous Clustering in Big Data Learning Analytics. In *Technology for Education (T4E), 2013 IEEE Fifth International Conference on* (pp. 61-64). IEEE.

[21]Admiraal, W., Huisman, B., & Van de Ven, M. (2014). Self-and peer assessment in massive open online courses. *International Journal of Higher Education*, *3*(3), p119.

[22]O'Toole, R. (2013). Pedagogical strategies and technologies for peer assessment in Massively Open Online Courses (MOOCs).

[23]Hernández-García, Á., González-González, I., Jiménez-Zarco, A. I., & Chaparro-Peláez, J. (2015). Applying social learning analytics to message boards in online distance learning: A case study. *Computers in Human Behavior*, *47*, 68-80.

[24]Burnett, R. E., Frazee, A., Hanggi, K., & Madden, A. (2014). A Programmatic Ecology of Assessment: Using a Common Rubric to Evaluate Multimodal Processes and Artifacts. *Computers and Composition*, *31*, 53-66.

[25]Mandala, S., Abdullah, A. H., & Ismail, A. S. (2013, June). A survey of e-learning security. In *ICT for Smart Society (ICISS), 2013 International Conference on* (pp. 1-6). IEEE.

[26]Lias, T. E., & Elias, T. (2011). Learning Analytics: The Definitions, the Processes, and the Potential.

[27]Larusson, J. A. and White, B. (2014). *Learning Analytics: from research to practice*. Springer.

[28]Siemens, G., & Long, P. (2011). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE review*, *46*(5), 30.

[29]van Barneveld, A., Arnold, K. E., & Campbell, J. P. (2012). Analytics in higher education: Establishing a common language. *EDUCAUSE learning initiative*, *1*, 1-11.

[30]Kohonen, T. (1998). The self-organizing map. *Neurocomputing*, *21*(1), 1-6.

[31]Grajciarová, L., Mareš, J., Dvorák, P., & Procházka, A. Biomedical Image Analysis Using Self-Organizing Maps.

[32]Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks*,*37*, 52-65.

[33]Hasan, S. (2010). *Enhanced self organizing map with particle swarm optimization for classification problems* (Doctoral dissertation, Universiti Teknologi Malaysia, Faculty of Computer Science and Information System).

[34]Cabada, R. Z., Estrada, M. L. B., & García, C. A. R. (2011). EDUCA: A web 2.0 authoring tool for developing adaptive and intelligent tutoring systems using a Kohonen network. *Expert Systems with Applications*, *38*(8), 9522-9529.

[35]Ramesh, V., Parkavi, P., & Yasodha, P. (2011). Performance analysis of data mining techniques for placement chance prediction. *International Journal of Scientific & Engineering Research*, *2*(8), 1.

[36]Pitigala Liyanage, M. P., Gunawardena, K. L., & Hirakawa, M. (2013, December). A framework for adaptive learning management systems using learning styles. In *Advances in ICT for Emerging Regions (ICTer), 2013 International Conference on* (pp. 261-265). IEEE.

[37]Soloman, B. A., & Felder, R. M. (2005). Index of learning styles questionnaire.*NC State University.*