Chen, W. et al. (Eds.) (2017). Proceedings of the 25th International Conference on Computers in Education.
New Zealand: Asia-Pacific Society for Computers in Education

Development of Alternative Conception Diagnostic System based on Item Response Theory in MOOCs

Yu-Cheng CHENG [a*], Jian-Wei, TZENG[b], Nen-Fu HUANG[ac], Chia-An LEE[a] & Meng-Lun Kuo[d]

[a]*Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan*
[b]*Center for Teaching and Learning Development, National Tsing Hua University, Hsinchu, Taiwan*
[c] *Netxtream Technologies Inc, Taiwan*
[d]*Institute of Service Science, National Tsing Hua University, Hsinchu, Taiwan*
*s104062514@m104.nthu.edu.tw

Abstract: With the popularities of Massive Open Online Courses, a great number of enrollments in MOOCs generate much educational big data in terms of online activities and logs, which might be valuable for academia and practitioners. More personalized and intelligent online learning environment could be potentially created through educational data mining and learning analytics techniques. Based on Item Response Theory (IRT), the current study builds an item analysis system to identify alternative concepts/misconceptions from leaners' response in exams. By calculating difficulty parameter and discrimination parameter from massive learners, our systems are believed to benefit both teaching faculties and online learners. With the affordances of the system, teaching faculties could assess leaners' learning performance and quality of test items while alternative conception of leaners would be identified for strategic learning. Other practical and technical implications will be discussed in this paper.

Keywords: Computer-Assisted Assessment, Item Response Theory, MOOCs (Massive Open Online Course), learning analytics, misconception, alternative conception, Item Analysis

1. Introduction

1.1 MOOCs

Massive online open courses (MOOCs) have emerged since 2012, providing high-quality online courses partnered with prestigious universities around the world. The innovative form of video-based learning has recently attracted attention from a variety of educational communities (Hsu, Huang, Chen, Lee, & Tseng, 2017). MOOCs allow learners from all over the world to enroll in a variety of online courses. Diverse learning modes and interaction could be found in MOOCs, including lecture videos, quizs and exercises, discussion forums, learning visualization, etc. Rather than only providing traditional course materials like OpenCourseware, MOOCs are famous for lively interaction among stakeholders of the courses (students, professors, and teaching assistants). The overwhelming trend of MOOCs' features makes it a contemporary way of 21-century learning (Huang, Hsu, Chen, Lee, Huang, Ou, Tzeng, 2017).

As MOOCs have been evolving, there have been two distinct types of MOOCs: cMOOC and xMOOC. cMOOC emphasizes the Connectivist view of learning, with a focus on leaners' autonomy and social nature of learning. xMOOC, on the other hand, is more closed to instructionism, implying the important role of lecture videos in the learning process (Breslow, Pritchard et al., 2013; Seaton et al., 2014). To be more specific, xMOOCs focus on the visual and audial representations in the instructional videos. Famous MOOC platforms such as Coursera, edX, and ShareCourse share such xMOOCs form of learning.

Jointly developed by National Tsing Hua University and Netxtream in Taiwan, SC (SC) is one of the largest MOOCs platforms in Chinese speaking society. More than 1,000 courses and over 100,000 active users have registered on SC since 2012. Furthermore, SC cooperates with over 70 universities and colleges. SC provides not only a good teaching and learning environment for online lecturers and learners, but also services like online discussion room, learning visualization system.

Although the developments of xMOOCs are widely accepted, the recent researches point out that the current progress of learning analytics systems is in a fancy stage (Ferguson & Clow, 2017). We believe that the great amount of educational big data in xMOOC platform might be valuable for course and platform improvement. IMS caliper and xAPI are both eager to accelerate the process of learning analysis through defined schema of learning activities (Santos et al., 2015 and Del et al., 2013). Among the types of activities, quizzes and exams are two primary ways of assessing learning performance in xMOOCs. Previous findings failed to focus on assessment mechanism. Moreover, most of analysis systems ignore the feedback of learners so that they lose the last arrow to close Learning Analytics Cycle or close the cycle in inefficiency way (Clow, 2012). There is still a lot of valuable information hidden in logs produced by xMOOCs platform.

1.2 Testing Analysis

Item response theory (IRT) is a statistical model used to evaluate the quality of measures (Kean & Reilly, 2014). Recently, there are several studies on Test-Item Analysis and Learning Management System(LMS). Fotaris, Mastoras, Mavridis & Manitsaris (2010) integrated LMS with an open source IRT tool, ICL, to improve the assessment quality, and found to (1) increase the efficiency of the testing process and (2) provide deeper guideline of test. Fotaris & Mastoras (2013) took advantage of IRT to optimize item pool on LMS. However, the author think the methodology is not limited to LMS. The environment of MOOCs with thousands of students might be more ideal for this experiment. Test-Item Analysis was also proposed to improve teaching and classification of students so that teachers could focus on learners' weakness (Kiat,1981). It is a more stable and rigorous method, which will not be influenced by sampled students' ability. The ideas of Item Difficulty and Item Discrimination are also thought of as practical way to estimate the item quality in traditional assessment.

However, the accuracy of IRT's analysis is limited by scale of data. IRT has better performance on sufficient data. In order to solve this problem, Professor Deng first introduced Grey System Theory that could be applied on small amount data (Julong, 1989). Naggi also proposed Rasch Model Grey Student-Problem chart based on Grey Relation Analysis, one of the methods in Grey System Theory. Sheu, Tzeng, Tsai & Chen (2012) obtained misconceptions that cause learner's confusion through the Rasch Model GSP chart even on insufficient data.

There have been some researches that perform IRT on data produced by MOOCs (Balint, Teodorescu, Colvin, Choi, & Pritchard, 2014 and Balint, Teodorescu, Colvin, Choi & Pritchard, 2015). To our knowledge, the current MOOCs services do not equip with item analysis features. This might be due to two reasons. First, the diverse and considerable number of learners deter the development of dynamic assessment system. Computing timely and precise data requires expansive computational power. Second, platform owners and designers put more emphasis on the interface and interaction design of the MOOCs platforms, failing to take advantages of assessment mechanism to improve learning effectiveness. Assessment-based learning is believed to be effective and goal-driven way of online learning. Through effective design of assessment system, learners' motivation and self-efficacy might be increased to foster learning. Therefore, this study developed a service integrating Test-analysis System into an existing MOOCs platform to offer alternative conception diagnostic service and testing item analysis service for both teaching faculties and online learners.

2. System Architecture

The whole service is built on two running instances, one is MOOCs service instance and the other one is datacenter. MOOCs service instance provides common online course as regular MOOCs platform like edX and Coursera. Datacenter instance provides restful API for the communication between two instances. MOOCs instance could send user's activity logs to data center and request for advanced analytics. This architecture provides flexibility for datacenter providing service to other MOOCs platforms and independency of two instances.

2.1 MOOCs Platform

In this paper, the MOOCs platform is SC. To communicate with the datacenter, the task of SC is required to sends logs with de-identification ID to datacenter through HTTP request. When students perform some activities on course material, activities logs will be packaged in JSON format, and send to the corresponding API route with specified header fields. The raw data will be stored in database waiting for analysis. After processing, the results will be also available through corresponding API route. SC sends a http request and reveals advanced analytics to teachers and students. The visualization and results tables are also available on our service.

2.2 API Server

The API server is the only interface that datacenter could connect to outside by. This would avoid exposing database and internal structure to external. It would validate the activity log received, filter logs that are from unknown source, and store the logs in specified schema in the database. Note that all the records stored on server are processed by de-identification. MOOCs doesn't send records with user' privacy, such as school, age, country, etc. Only unique identification is used to distinguish whom the records belong to. While MOOCs instance requests for certain analytic results, the API server fetches the results from database and returns the results to MOOCs instance similarly.

2.3 Alternative Conception Diagnosis Module

This module retrieves the misconceptions based on ordinal of gamma value, a float number list ranging 0 to 1. The problems with gamma value near 0.5 are considered as misconceptions. The alternative conception points out the concept that students need to pay more attention to. Moreover, it is proved that results can be produced under of even a small amount of people such as classical face-to-face education (Sheu, Tzeng, Tsai & Chen, 2012). This overcomes small number of passed students problem caused by the low pass rate problem while performing other analysis.

2.4 Analyzer

The analyzer is triggered by crontab daily. It calculates Testing Items, such as Index of difficulty and Index of discrimination for each exercise. Besides, it also mixes the result of *Alternative Conception Identify Module* with problem-concept mapping table to extract alternative conception of each course. All the analytics results above will be stored back to database.

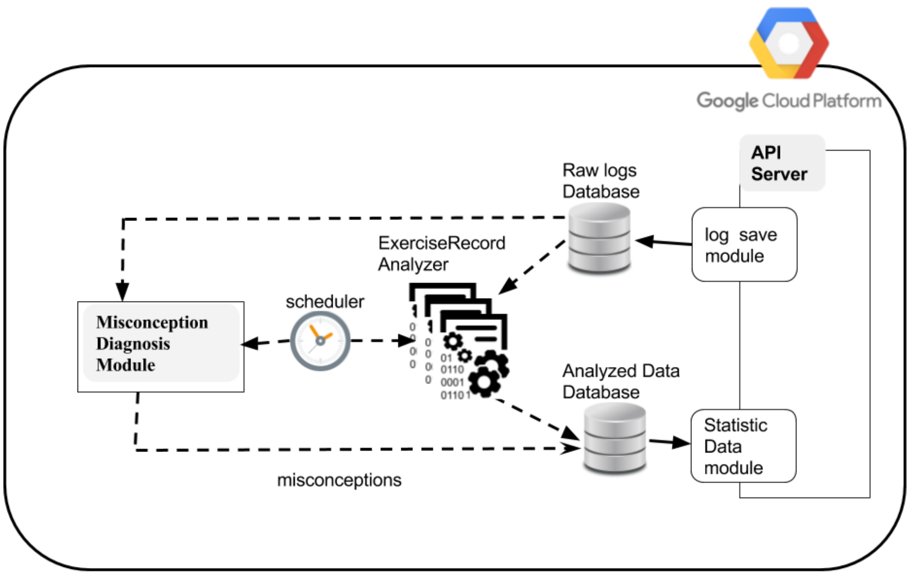

Figure 1. An insight view of Datacenter including four modules, Database, API Server, Misconception Diagnosis Module and Analyzer.

3. Algorithm

In this section, We illustrate the algorithm which analyzes answer records stored in database and comes out with Item difficulty index, Item discrimination index and the alternative conception questions.

3.1 Testing Item Index

In this paper, Kiat's definitions of high-low group Difficulty Index and Discrimination Index are applied. In the beginning, the program loads the learner's records of one course from database. Because of the high attrition rate and low completion of course content in MOOCs, the records of these learners who didn't complete the course would bias the analysis results. The following process includes only the records of learners who completed the course and took all the exams.

Before calculating, it is necessary to label student by their grade. If the student scores in top 27%, the student is labeled as 'high score group'. On the contrary, the students in last 27% are labeled as 'low grade group'.

The analyzer will aggregate the records of each problem and group by above-mentioned labels. Finally, it could combine the correct rate of the two groups to figure out the Item Difficulty Index and Item Discrimination Index. The difficulty index is within the range [0,1] and the discrimination index is within the range [-1,1]. The indexes of problems are all stored back to the database eventually.

Because it's not friendly to show raw index value to users and learners and tutors would not understand the meaning behind the value, the raw index value is transformed into signals like traffic light, great, normal and bad, so that users could easily catch the hint. These signals are shown in Chinese text on SC.

3.2 Alternative conception

In this paper, the methodology proposed by Sheu, Tzeng, Tsai & Chen (2012) is used to calculate the grey relation and to retrieve the alternative conception of the course. Unlike other algorithm, the gamma value is between [-1,1], which make it easier for comparison. Nagai's equation used in aforementioned paper is shown as follows.

$$\Gamma_{0i} = \Gamma(x_0, x_i) = \frac{\overline{\Delta}_{max} - \overline{\Delta}_{0i}}{\overline{\Delta}_{max} - \overline{\Delta}_{min}}$$

where x_0 is partial grey relation's reference vector and x_i is comparative vector and

$$\overline{\Delta}_{0i} = \|x_{0i}\|_\rho = \left(\sum_{k=1}^{m} [\Delta_{0i}(k)]^\rho\right)^{\frac{1}{\rho}}$$

Where $\overline{\Delta}_{max}$ and $\overline{\Delta}_{min}$ represent $\overline{\Delta}_{0i}$'s maximum and minimum. When Γ_{0i} is close to 1, it means that x_0 and x_i are highly related to each other. On the other hand, if Γ_{0i} is close to 0, the relationship between x_0 and x_i is lower.

4. Experiments and Results

In the experiments, two courses on SC are selected: Investment and Introduction of computer network. Both course have been taught so that there are students' responses of exam collected. We applied analysis on these records and provide results to processing course on the platform. In the course of Investment, our analysis focuses on 240 participants who completed both of the exams and our analysis in Introduction of computer focuses on 372 participants who completed both of the exams too. We also have requested for professor's agreement and asked teaching assistants for problem-concept mapping tables.

Figure 2 shows that students could see the discrimination index and difficulty index after the exam assessed. Note that we only show Chinese signals to represent level of discrimination index and difficulty index. Besides, we also provide these indexes with weekly exercises and item pool of tutors. Report of index reveals the quality of whole exam is provided to tutors, helping tutors to understand problems that need to be improved. At last, we show list of alternative conceptions in scope of midterm, as shown in table 1, to guide students to review the course material about these alternative conceptions.

Figure 2. Both Discrimination parameter and Difficulty parameter are attached to each exam problem.

Table 1: An example of misconceptions list of Investment midterm

Name of Course	Misconceptions of Midterm (In Chinese)
Investment	金融市場,資本市場,股票指數,融券放空

5. Conclusions & Future work

Due to the rapid development of MOOCs and the great number of online learners, massive online activity logs data would be generated and stored in web service. Online learning environment is different from traditional face-to-face classroom interaction, which teachers would be easier to understand learners' feedbacks for improving teaching quality. Moreover, the massive enrollment makes it even challenging for online instructors and teaching assistants to provide with adaptive intervention and guidance.

With the affordances of learning analytics, this paper develops an *Alternative Conception Diagnostic System,* automatically providing item analysis for teachers and alternative conceptions for online learners. Moreover, the system cooperates with a MOOCs platform that serves thousands of learners. This intervention makes learners identify alternative conceptions of course so that they could devote more efforts on these sections to get a higher performance. It also closes the learning analytics cycle and makes the online learning environments a better place.

Future work is to improve the presentation of alternative conceptions and figure out further usages. In this paper, only text table of misconceptions is produced. We are seeking for a better representation of alternative conceptions, such as knowledge map, which shows organized structure of course concepts. Highlighting the alternative conceptions on the Knowledge Map is a straightforward way to make learners understand, promoting the learning efficiency of MOOCs.

Acknowledgements

We would like to thank all the people who prepared and revised previous versions of this document.

References

Balint, T. A., Teodorescu, R., Colvin, K., Choi, Y. J., & Pritchard, D. E. (2015). Comparing measures of student performance in hybrid and MOOC physics courses. European Journal of Physics Education, 6(3).

Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research & Practice in Assessment, 8*.

Chauhan, A. (2014). Massive open online courses (MOOCS): Emerging trends in assessment and accreditation. *Digital Education Review*, (25), 7-17.

Chu, H. C., Hwang, G. J., & Huang, Y. M. (2010). An enhanced learning diagnosis model based on concept–effect relationships with multiple knowledge levels. *Innovations in Education and Teaching International, 47*(1), 53-67.

Clow, D. (2012, April). The learning analytics cycle: closing the loop effectively. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 134-138). ACM.

Colvin, K., Champaign, J., Liu, A., Fredericks, C., & Pritchard, D. (2014, July). Comparing Learning in a MOOC and a Blended, On-Campus Course. In Educational Data Mining 2014.

Del Blanco, Á., Serrano, Á., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2013, March). E-Learning standards and learning analytics. Can data collection be improved by using standard data models?. In Global Engineering Education Conference (EDUCON), 2013 IEEE (pp. 1255-1261). IEEE.

Ferguson, R., & Clow, D. (2017). Where is the evidence? A call to action for learning analytics.

Fotaris, P., & Mastoras, T. (2013, September). Integrating IRT analysis into LMS for item pool optimization. In *Proceedings of the 2nd Workshop on Technology Enhanced Formative Assessment*, co-located with the 8th European Conference on Technology Enhanced Learning (pp. 17-21).

Fotaris, P., Mastoras, T., Mavridis, I., & Manitsaris, A. (2010). Extending LMS to support IRT-based assessment test calibration. In *Technology Enhanced Learning. Quality of Teaching and Educational Reform* (pp. 534-543). Springer Berlin Heidelberg.

Hsu, H. H., Huang N. F., Chen, S. C., Lee, C. A., & Tseng, J. W. (2017, March). Misconceptions Mining and Visualizations for Chinese-based MOOCs Forum Based on NLP. In Big Data Analysis (ICBDA), 2016 IEEE International Conference. IEEE.

Hewson, M. G., & Hewson, P. W. (1983). Effect of instruction using students' prior knowledge and conceptual change strategies on science learning. *Journal of Research in Science Teaching, 20*(8), 731-743.

Hwang, G. J. (2007). Gray forecast approach for developing distance learning and diagnostic systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(1), 98-108.

Hwang, G. J., Tseng, J. C., & Hwang, G. H. (2008). Diagnosing student learning problems based on historical assessment records. *Innovations in Education and Teaching International, 45*(1), 77-89.

Julong, D. (1989). Introduction to grey system theory. The Journal of grey system, 1(1), 1-24.

Kean, J., & Reilly, J. (2014). Item response theory. Handbook for Clinical Research: Design, Statistics an*d Implementation*. (pp195-198). New York, NY: Demos Medical Publishing.

Kiat, T. W. (1981). Test-Item Analysis.

Liang, J. C., Sheu, T. W., Tzeng, J. W., Wang, B. T., & Nagai, M. (2011). 5W1H, GRA and GSM in the evaluation and identify on optimal design of bike lamps. *Journal of Convergence Information Technology (JCIT), 6*(12), 266-274.

Huang N. F., Hsu, H. H., Chen, S. C., Lee, C. A., Huang, Y. W., Ou, P. W., & Tzeng, J. W. (2017, March). VideoMark: A Video-Based Learning Analytic Technique for MOOCs In Big Data Analysis (ICBDA), 2016 IEEE International Conference. IEEE.

Pardos, Z. A., & Kao, K. (2015, March). moocRP: An open-source analytics platform. In Proceedings of the Second (2015) ACM conference on learning@ scale (pp. 103-110). ACM.

Ruiz, J. S., Díaz, H. J. P., Ruipérez-Valiente, J. A., Muñoz-Merino, P. J., & Kloos, C. D. (2014, October). Towards the development of a learning analytics extension in open edX. In *Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality* (pp. 299-306). ACM.

Santos, J. L., Verbert, K., Klerkx, J., Duval, E., Charleer, S., & Ternier, S. (2015). Tracking data in open learning environments. Journal of Universal Computer Science, 21(7), 976-996.

Seaton, D. T., Bergner, Y., Chuang, I., Mitros, P., & Pritchard, D. E. (2014). Who does what in a massive open online course? *Communications of the ACM, 57*(4), 58-65.

Sheu, T. W., Tzeng, J. W., Tsai, C. P., & Chen, T. L. (2012). Applying Problem-Concept Chart Combined with Structural Analysis to Investigate the Learning Misconcept-Simple Equation with One Variable for Example. *Journal of Grey System, 15*(1), 55-66.

Tzeng, J. W., Sheu, T. W., Liang, J. C., Wang, B. T., & Nagai, M. (2012). A new proposal based on Rasch model GSP chart and grey structural model with analysis and discussion. *International Journal of Advancements in Computing Technology, 4*(9), 111-121.