Design of a Personalized
Massive Open Online Course Platform

Junfu Xi, Yehua Chen[✉], Gang Wang
Xingtai Polytechnic College, Xingtai, China
`beij-08@163.com`

Abstract—Focusing on the massive open online course (MOOC) platform, the purpose of this study is to realize personalized adaptive learning according to the needs and abilities of each learner. To this end, the author created a personalized adaptive learning behaviour analysis model, and designed a personalized MOOC platform based on the model. Through the analysis of learning behaviours on the MOOC platform, the model digs deep into the pattern of learning behaviours, and lays the basis for personalized intervention in the learning process. The comparison experiments show that our prediction method is more accurate than the other prediction algorithms. This research sheds new light on the design of learner-specific MOOC platform.

Keywords—massive open online course (MOOC); big data; personalized adaptive learning; learning analytics

1 Introduction

A massive open online course (MOOC) [1] is an online course aimed at unlimited participation and open access. First introduced in 2006, MOOC has developed into a novel and popular platform of distance learning in recent years [2-4].

The open access to content, structure and learning goals is an essential feature of early MOOCs [5], which promotes the reuse and remix of resources. With the proliferation of MOOCs, an increasingly number of new features continue to emerge. Despite the partial similarity in the learning data, the MOOC platform differs greatly from traditional classroom in behaviour collection and analysis. Unlike traditional classroom, MOOC platform can record various user operations and capture each submission from the user. Besides, the platform contains much greater details on user behaviours than traditional classroom. The gigantic amount of data makes it possible to implement the technology of big data analytics in the platform.

The era of data-driven education has come, big data technology has a profound impact on education based on MOOCs. In MOOC platforms, teachers and administrators can use big data technology to inform teaching decisions, give grades, award credits and so on. And even big data analytics can predict and adjust student progress to inform both instructional and institutional choices.

The big data-based learning analytics was defined as: "*the measurement, collection, analysis and reporting of big data about learners…for purposes of understanding and optimizing learning and the environments in which it occurs*" [6]. This technology predicts the future performance of learners based on their learning progress, enables them to self-define the learning contents and select preferred courses, and notifies instructors in advance when learners need academic guidance. In this way, the learners can unlock their full learning potential and have a good command of knowledge.

This paper proposes a personalized adaptive online learning analysis model for MOOCs. The model mainly analyses the big data generated by the learner operations on MOOC platform, and provides effective guidance on the teaching and learning on the platform.

The remainder of this paper is organized as follows. OVERVIEW OF MOOCS briefly introduces MOOCs; LITERATURE REVIEW reviews the related studies; PERSONALIZED MOOC PLATFORM BASED ON BIG DATA ANALYTICS puts forward the PAOLA model based on big data analytics; EXPERIMENTS AND ANALYSIS carries out several experiments and analyses the experimental results; CONCLUSIONS wraps up this research with some meaningful conclusions.

2 Overview of MOOCS

The concept of MOOC was invented by Stephen Downes, as an online course aimed at unlimited participation and open access via the web [7]. The first MOOCs emerged from the open educational resources movement in 2008. With the development of MOOCs, there appear to be two distinct types: cMOOC and Xmooc [8]. Following the connectivist philosophy, cMOOCs require that teaching resources must be remixable and re-purposable. The basic methods of cMOOCs attempts to connect learners to each other to complete the learning process. By contrast, xMOOCs have a much more traditional course structure. They employ elements of the original MOOC. The instructor is the knowledge provider and problem solver, but learner interactions are usually limited to asking for assistance and advising each other on difficult points.

The timeline of MOOCs is shown in Figure 1.

Instructor, learner, course, resource and context are five essential factors of a MOOC system [9]. The instructor needs to simplify the learning process and give guidance to the learner; the learner should study the course by logging in the MOOC platform; the course is the externalization of the resource and the context on the platform; the resource of the platform can be accessed via multiple education methods; the context represents such components of the platform as online social networks, IT solutions, communication systems, and so on. Through the integration of the above five factors, MOOCs can support the learning of a large community, allowing the students to acquire knowledge anytime, anywhere [10].

Compared to traditional distance education, MOOCs mainly have the following advantages [11]. First, MOOCs help the students replace passive reception of instruc-

tions with autonomous learning; second, MOOCs encourage cooperation and team-work, which are not highlighted in traditional distance education.

Over the years, MOOCs have led to numerous changes to higher education [12]. They can realize team-based course design. The construction of MOOCs requires the collaboration of all faculty members, and the support from designers, software developers, teaching researchers, librarians and videographers. They can focus on teaching process. MOOCs shine a light on teaching and learning in universities. They offer a proving ground to proving ground, a viable strategy to promote active learning in traditional courses. They can give space for innovation. The supporting structures of MOOCs provide an ideal space for innovation in teaching and learning, and give birth to thoughtful and bold ideas for higher education in a digital era.
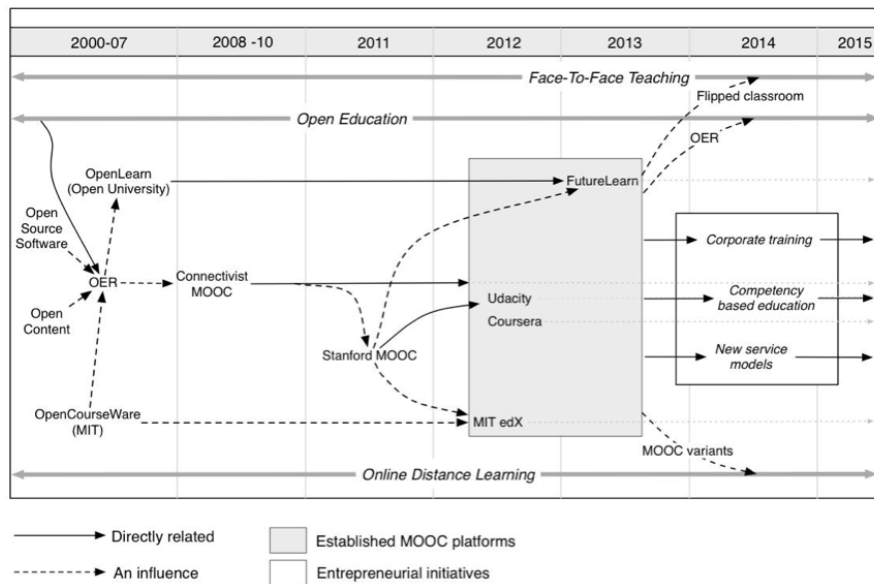

Fig. 1. The MOOCs timeline

3 Literature Review

The rapid development of MOOCs creates a perfect environment for the integration of big data and teaching research. The lifecycle of the big data generated from MOOCs is illustrated in Figure 2.
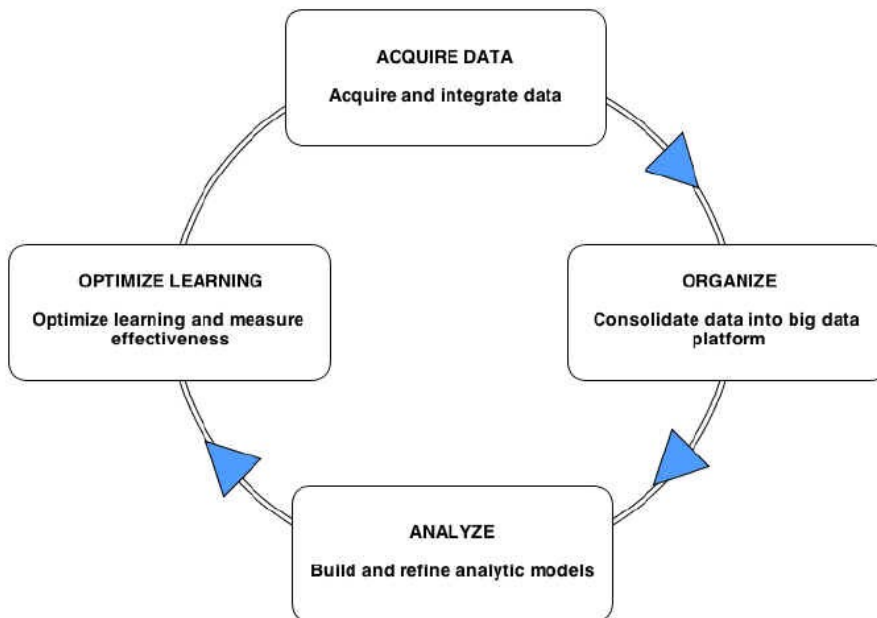
Fig. 2. The lifecycle of big data generated from MOOCs

Some MOOC data are the same with those generated in traditional classrooms, namely, teaching resources, enrolment information, and test scores. The major difference between MOOCs and traditional classroom teaching lies in the behavioural data of learners. The MOOC platform can capture and record mouse clicks, video controls and even all submissions to the platform. The application of big data analytics to these behavioural data opens up an effective way to enhance the learning/teaching efficiency and effectiveness.

In the past, learner behavioural data have been relied on to determine the influencing factors on the drop-out rate of MOOCs. For example, Thille et al. developed a personalized content delivery method based on the average time-on-page of learners, seeking to identify different cohorts of learners [13].

Besides, many strategies have been developed to tackle the prediction issue. Eriksson et al. proposed a method that can predict the performance and drop-out probability of learners in MOOC-based learning [14]. Similarly, Hughes et al. constructed a reliable model to forecast the early drop-outs [20]. Romero et al. predicted GPA with a regression analysis algorithm. Based on big data analytics [21], Polyzou et al. set up grade prediction models for specific students and courses [15]. These models are applicable to pre-registration grade prediction and in-class grade prediction. El-badrawy et al. projected the next-term grades of students after integrating new performance prediction techniques into recommender systems [16]. To sum up, the final results of MOOC-based learning rely heavily on theoretical framework, research design and analysis methods.

Much research has also been done on personalized learning behaviours and functions of the MOOC platform. Through scientific means, Ren et al. compared the personalized learning features of several well-known MOOC platforms [17]. Marin et al. developed a personalized feedback mechanism for learners to reflect on their understanding of MOOC courses [18]. Focusing on several specific issues, Zeide evaluated big data-driven learning environments, and advised educators and policymakers to consider the explicit implications of data-driven education [19].

In spite of the multi-angle research on MOOC-based learning, the previous studies fail to provide personalized adaptive learning according to the needs and abilities of each learner.

Therefore, this paper explores the learning process, learner behaviours and learning rules through big data learning analysis model, so that the MOOC platform can automatically recommend reasonable learning path and learning resources to each learner.

4 Personalized Mooc Platform Based on Big Data Analytics

This section aims to design a personalized MOOC platform that provides timely and accurate feedbacks to learners. The platform should automatically adjust the learning path, provide learners with adaptive resources, and enable instructors to give personalized guidance according to the learning behaviours and needs of each learner. To this end, the collaborative filtering was introduced to push the learning information of a user to those with the same or similar interests.

4.1 Personalized adaptive learning behaviour analysis model

The basic data flow of our personalized adaptive learning behaviour analysis model is shown in Figure 3.

The proposed model consists of six components. In Figure 3, (1) stands for the content presentation and delivery component, which interactively delivers personalized contents to learners and evaluates their learning performance; (2) refers to the big data repository of student learning, which stores the learner inputs and behaviours captured during their learning on MOOC platform; (3) represents the future behaviour/performance prediction model, which relies on the learning and behavioural data from the big data repository; (4) symbolizes a function that provides learners with visible feedbacks based on the predicted results; (5) indicates the adaptation engine that adjusts the content presentation based on the predicted results, and delivers resources depending on learner-specific performance and interests; (6) expresses the intervention engine, which improves the learning process by instructor intervention in the automated platform.
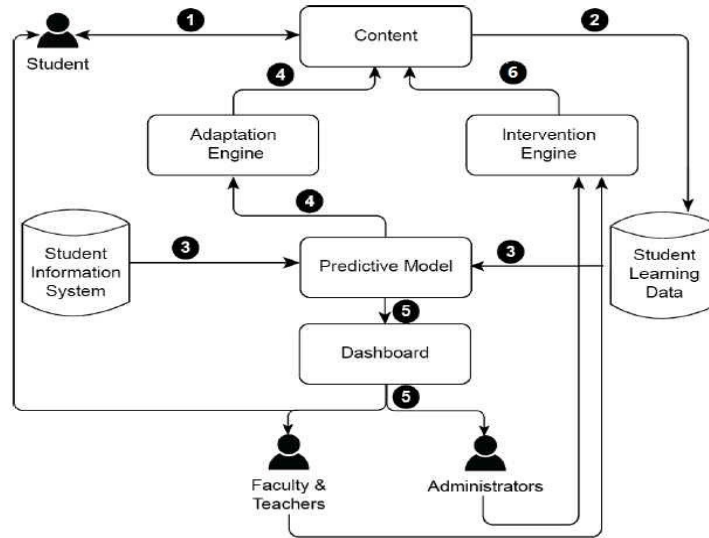
Fig. 3. The basic data flow of personalized adaptive learning behaviour analysis model

The prediction model is obviously the centrepiece of our model. Two methods were investigated for the establishment of the model. In the first method, the model is created by linear regression algorithm, with learner data and related features being the predictor variables. In the second method, the model is built up based on matrix factorization algorithm, aiming to find a low-dimensional space that represents both the learner and the content.

In this research, the personalized linear regression model employs a linear combination of m learner-specific regression models. The predicted value $\hat{p}_{ij}$ for learner i in course j can be expressed as:

$$\hat{p}_{ij} = v_o + t_i + s_j + S_i \times U \times X_{ij} \tag{1}$$

where v_o is the global bias value; t_i is a bias term for learner i; s_j is a bias term for course j; S_i is the $1 \times n$ vector for learner i; U is the $n \times m$ coefficient matrix; X_{ij} is the feature vector associated with learner i and course j. The two bias terms reflect the average score of a learner in the past and the average score of a course in the past. The information in X_{ij} includes the factors related to learners and those related to courses.

In our matrix factorization algorithm, each learner i and each course j are respectively described by k-dimensional feature vectors v_i and v_j. The formula below predicts the score of learner i in course j.

$$\hat{p}_{ij} = \sum_{t=1}^{k} v_{i,t} v_{j,t} \tag{2}$$

4.2 Design of personalized MOOC platform

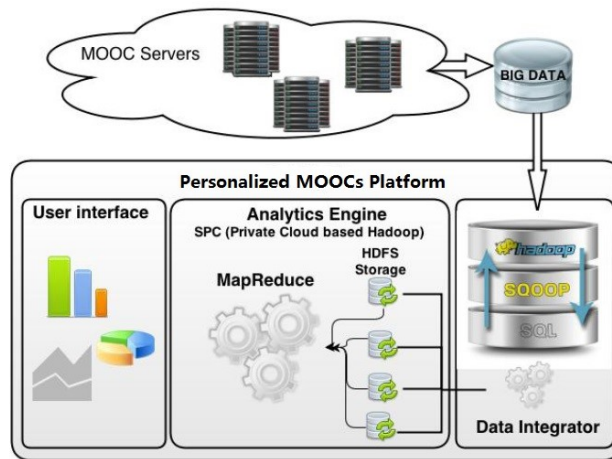Figure 4 shows a prototype of the personalized MOOC platform.

Fig. 4. The prototype of personalized MOOC platform

In our personalized MOOC platform, learners can receive feedbacks on their future choices and activities during the learning process. The next activities are recommended on the learner dashboard (Figure 5).
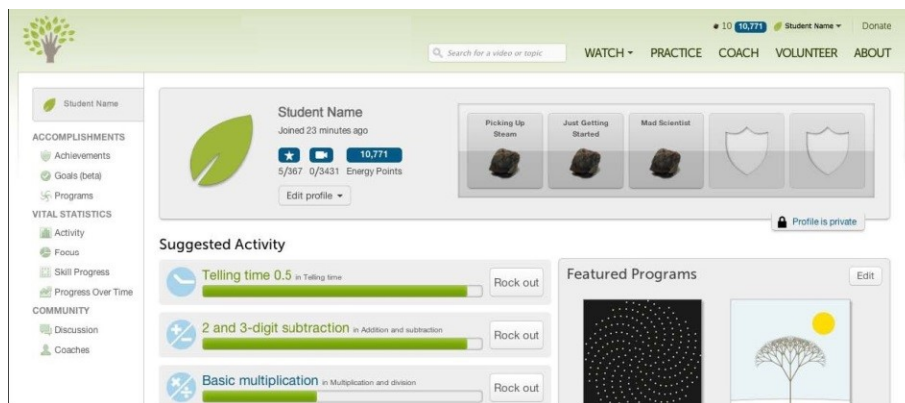

Fig. 5. The learner dashboard showing the detailed learning information

Apart from the recommended activities, the learner dashboard also compares the learning performance of a learner with that of other learners. In accordance with the comparison, the learner can adjust the learning path and content on the platform.

The instructor dashboard lists the performance of each learner (Figure 6).
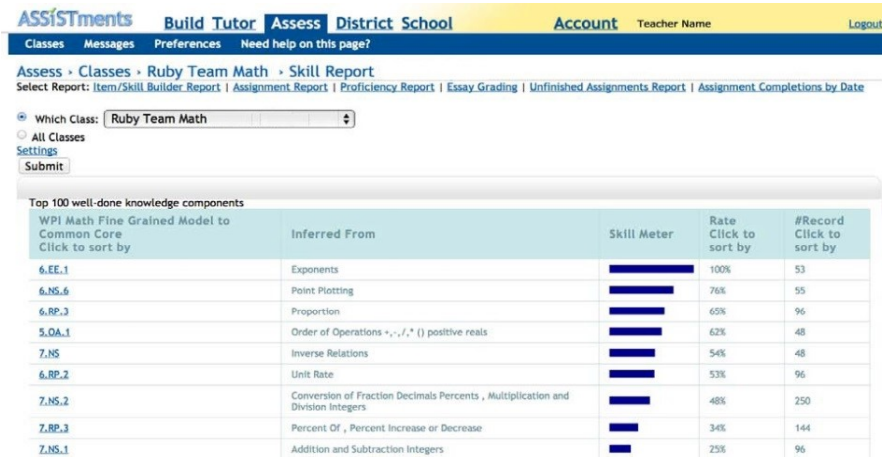
Fig. 6. The instructor dashboard showing performance of each learner

Based on the centralized display of learner performance of the whole class, instructors can adjust their teaching behaviours and the learning pace of each learner.

The administrator dashboard lists the detailed data on different classes (Figure 7).
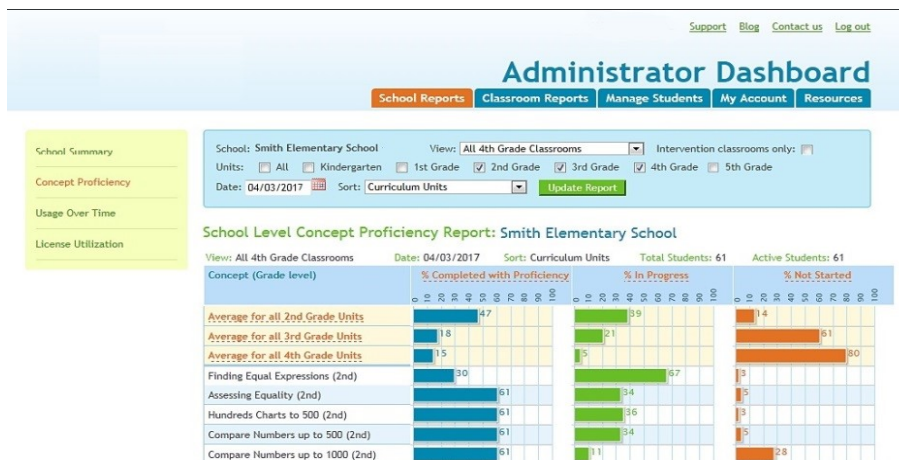

Fig. 7. The administrator dashboard showing the detailed data on different classes

The information on administrator dashboard explains the effect of a particular policy on learning performance. In view of the information, administrators can improve the teaching and learning policies.

5 Experiments and Analysis

To verify the effect of our model, the related data were gathered from a personalized MOOC platform deployed in Qingdao University.
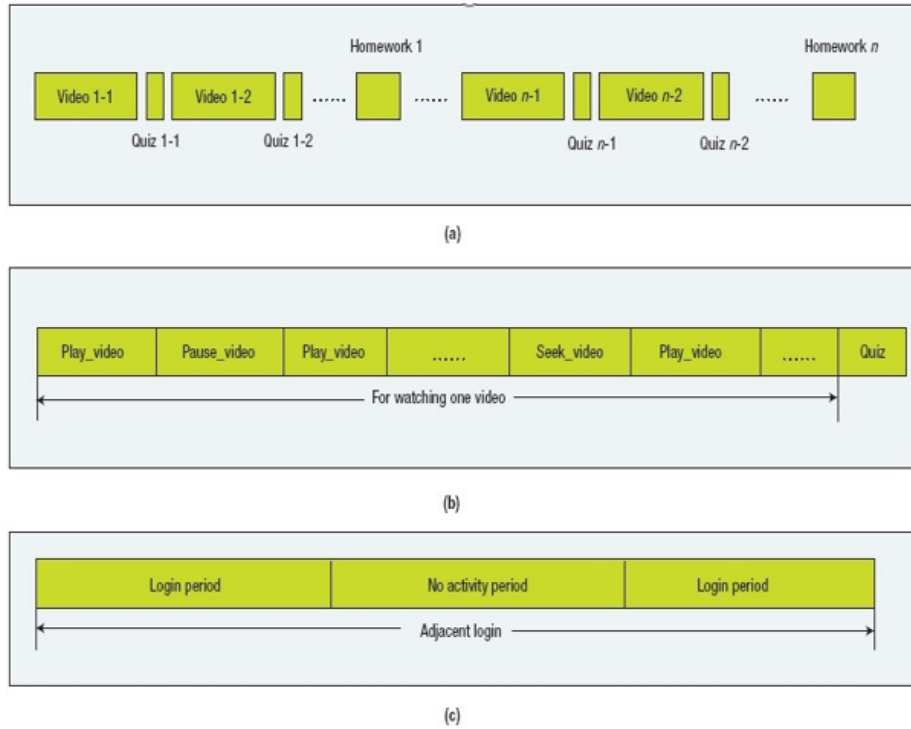
Fig. 8. Various learner behaviours in the personalized MOOC platform ((a)Video lecture learning process; (b) Video-viewing operations; (c) Multiple login sessions)

Firstly, our matrix factorization (MF) algorithm was contrasted with several commonly used prediction algorithms. Table 1 lists the predicted results on the personalized MOOC data by these algorithms.

Table 1. Predicted results of several algorithms

Method	Root-mean-square error (RMSE)	Mean absolute error (MAE)
Matrix factorization (MF)	0.7326	0.525
Personalized linear multi-regression (PLMR)	0.7877	0.564
Random forest (RF)	0.7983	0.572
Mean of means	0.8327	0.602
Uniform random guessing	1.9823	1.532

From Table 1, it can be seen that our MF algorithm had the lowest error in prediction. For the density of learner-course score matrix, the MF algorithm outperforms course-specific regression (CSpR) algorithm and course-specific matrix factorization (CSpMF) algorithm. Figure 9 presents the predicted results on the personalized MOOC data by these algorithms.
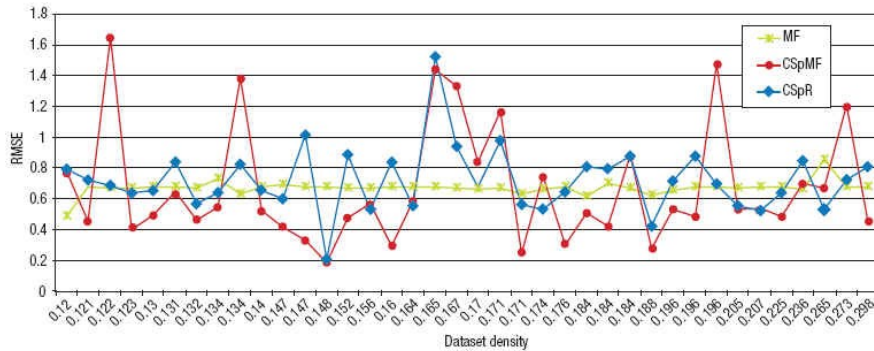

Fig. 9. Predicted results of three algorithms

Secondly, the author tested the influence of cold start on MF, random forest (RF) and polarimetric L-band multibeam radiometer (PLMR) algorithms. The predicted results of these algorithms for cold start (CS) and non-cold start (NCS) records are given in Table 2.

Table 2. Predicted results comparison of three algorithms for cold start and non-cold start records

Group	Percent	Method	RMSE	MAE
NCS	46.8	MF	0.7385	0.5183
		PLMR	0.7893	0.5429
		RF	0.7981	0.5678
CS student only	41.3	MF	0.8091	0.5862
		PLMR	0.9653	0.6821
		RF	0.7364	0.5256
CS course only	0.165	MF	0.7466	0.5238
		PLMR	1.2153	0.8643
		RF	0.7865	0.5376
CS both	0.419	MF	0.8352	0.6235
		PLMR	1.2105	0.8573
		RF	0.8296	0.5768

Table 2 shows that the MF algorithm is the one least affected by cold start problem.

Next, the author evaluated the performance of the personalized MOOC platform. During the construction of the platform, the MapReduce-based structure was intro-

duced with different number of parallel nodes. The introduction may affect the speedup of big data analytics. Through simulation, the relationship between the number of login learners and platform speedup is shown in Figure 10.
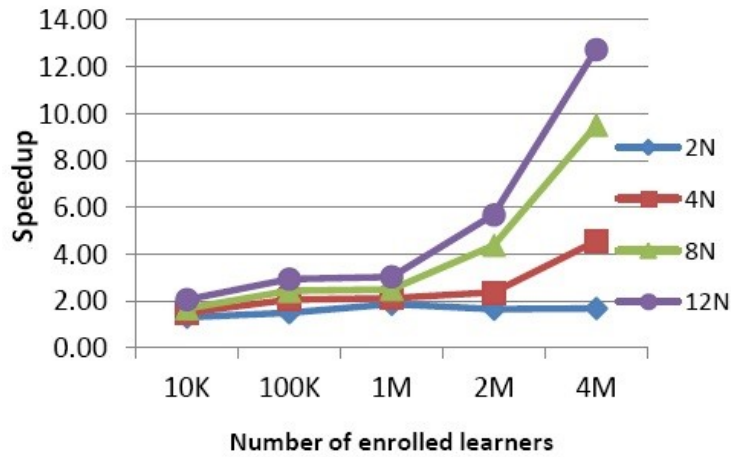

Fig. 10. The relationship between the number of login learners and platform speedup

As shown in Figure 10, the advantage of MapReduce was not fully displayed at a small number of login learners. However, the speedup rocketed up when that number increased continuously. Hence, a large number of nodes is conducive to the performance of big data analytics.

6 Conclusions

With the popularity of data collection and mining to feed learning analytics, it will influence more and more of MOOCs education. The data created by each learner is part of the "big data", and each learner is a producer and a consumer of big data. Under the support of big data analysis, there had been lots of studies for content push and quality analysis of learning resources on MOOC platforms. Most of the existing studies on MOOC data analytics had focused on predicting the drop-out rate or the learner performance, and overlooked the practical use of personalized MOOC platform. To make up for the gap, this paper develops a personalized adaptive learning behaviour analysis model, and designs a personalized MOOC platform based on the model. The model digs deep into learning behaviours, and reveals the relationship between learning behaviours and implicit data. Our work helps to grasp the essence of the learning process and implement personalized teaching. Through the big data analysis model, we can explore the learning process of learners, discover learning rules, and provide personalized adaptive learning according to the needs and capabilities of every student. In the future, the author will examine even more features of learning behaviours, and design suitable learning materials for each learner.

7 References

[1] Kaplan, A.M., Haenlein, M. (2016). Higher education and the digital revolution: About MOOCs, SPOCs, social media, and the Cookie Monster. Business Horizons, 59(4), 441-450. https://doi.org/10.1016/j.bushor.2016.03.008

[2] Siemens, G. (2013). Massive open online courses: Innovation in education. Open educational resources: Innovation, research and practice, 7(5), 5-15.

[3] Breslow, L., Pritchard, D.E., Deboer J. (2013). Studying Learning in the Worldwide Classroom Research into edX's First MOOC. Research & Practice in Assessment, 11(8), 13-25.

[4] Zemsky, R. (2014). With a MOOC MOOC here and a MOOC MOOC there, here a MOOC, there a MOOC, everywhere a MOOC MOOC. The Journal of General Education, 63(4), 237-243. https://doi.org/10.5325/jgeneeduc.63.4.0237

[5] Reich, J. (2014). MOOC Completion and Retention in the Context of Student Intent. Silnki Spalinowe, 99(2), 387-392.

[6] Kop, R., Sitlia, H. (2011). The value of learning analytics to networked learning on a personal learning environment, International Conference on Learning Analytics and Knowledge, ACM, 104-109.

[7] Jordan, K. (2014). Initial Trends in Enrolment and Completion of Massive Open Online Courses, International Review of Research in Open & Distance Learning, 15(1), 133-160. https://doi.org/10.19173/irrodl.v15i1.1651

[8] Prpic, J., Melton, J., Taeihagh, A. (2015). MOOCs and Crowdsourcing: Massive Courses and Massive Resources. Social Science Electronic Publishing, 20(12), 77-89. https://doi.org/10.5210/fm.v20i12.6143

[9] Kop, R. (2011). The Challenges to Connectivist Learning on Open Online Networks: Learning Experiences during a Massive Open Online Course. International Review of Research in Open & Distance Learning, 12(3), 19-38. https://doi.org/10.19173/irrodl.v12i3.882

[10] Milheim, W.D. (2013). Massive Open Online Courses (MOOCs): Current Applications and Future Potential. Educational Technology, 53(3), 38-42.

[11] Penuel, W.R., Roschelle, J., Shechtman, N. (2007). Designing formative assessment software with teachers: An analysis of the co-design process. Research and Practice in Technology Enhanced Learning, 2(1), 51-74. https://doi.org/10.1142/S1793206807000300

[12] Sandeen, C. (2013). Integrating MOOCs into traditional higher education: The emerging "MOOC 3.0" era. Change: The magazine of higher learning, 45(6), 34-39. https://doi.org/10.1080/00091383.2013.842103

[13] Thille, T., Candace, S., Emily, K. (2014). The Future of Data-Enriched Assessment. Research & Practice in Assessment, 9, 17-21.

[14] Eriksson, T., Adawi, T., Stöhr, C. (2017). "Time is the bottleneck": a qualitative study exploring why learners drop out of MOOCs. Journal of Computing in Higher Education, 29(1), 1-14. https://doi.org/10.1007/s12528-016-9127-8

[15] Polyzou, A., Karypis, G. (2016). Grade prediction with models specific to students and courses. International Journal of Data Science & Analytics, 2(3), 1-13. Doi: 10.1007/s41060-016-0024-z. https://doi.org/10.1007/s41060-016-0024-z

[16] Elbadrawy, A., Polyzou, A., Ren, Z. (2016). Predicting Student Performance Using Personalized Analytics. Computer, 49(4), 61-69. https://doi.org/10.1109/MC.2016.119

[17] Marin, V.J., Pereira, T., Sridharan, S. (2017). Automated Personalized Feedback in Introductory Java Programming MOOCs. International Conference on Data Engineering, IEEE, 1259-1270.

[18] Ren, Y., Zhao, L., Liu, M. (2015). How Far the MOOCs Being Away from Personalized Learning-Based on the Analysis of 10 MOOCs Designs at Home and Abroad. Modern Distance Education Research, 12(6), 3-10.

[19] Zeide, E. (2017). The Structural Consequences of Big Data-Driven Education. Big Data, 5(2), 164-172. https://doi.org/10.1089/big.2016.0061

[20] Hughes, G., Dobbins, C. (2015). The utilization of data analysis techniques in predicting student performance in massive open online courses (MOOCs). Research & Practice in Technology Enhanced Learning, 10(1), 10-16. https://doi.org/10.1186/s41039-015-0007-z

[21] Romero, C., López, M.I., Luna, J.M. (2013). Predicting students' final performance from participation in on-line discussion forums. Computers & Education, 68, 458-472. https://doi.org/10.1016/j.compedu.2013.06.009

8 Authors

Junfu Xi received the M.S. degree in computer application technology from Yanshan University，Qinhuangdao, China, in 2010. Since 2010, he has been on the faculty of Xingtai Polytechnic College. His research interests include intelligent algorithm optimize, computer education.

Yehua Chen received B.S. degree in computer application technology from Taiyuan University of Technology, Taiyuan, China, in 2002, the M.S. degree computer application technology from Hebei University of Science and Technology, Shijiazhuang, China, in 2013. Since 2002, she has been on the faculty of Xingtai Polytechnic College. Her research interests include electronic commerce, computer education.

Gang Wang received the B.S. degree in physics from North University of China, Taiyuan, China, in 2007, the M.S degree in computer application technology from China University of Mining and Technology, Beijing, China, in 2015. Since 2015, he has been on the faculty of Xingtai Polytechnic College. His research interests include internet of mobile, technique of information fusion, algorithms and software systems.