Application of Data Mining in MOOCs for Developing Vocational Education: A Review and Future Research Directions

Jianzhen Zhang, Jia Tina Du, and Fang Xu

Abstract—**Massive Open Online Courses (MOOCs) can be considered as one of the most prominent developments in education, which brings new opportunities for higher and vocational education. This paper presented an in-deep literature review on the application of data mining in MOOCs. We found there are 8 types of behavior data mainly researched by the existing publications, and then classified the main application of the data mining in MOOCs into 7 directions. However, there is as yet little evidence on the application of data mining on MOOCs for developing vocational education. Based upon the review findings, we presented 3 recommendations, including applying cluster to find the effective marketing area for vocational education organizations, applying association analysis to figure out vocational education course sets for the specific profession, and applying regression analysis to recommend the personalized career planning for candidates. This article can be useful for vocational institutes and MOOCs platforms to develop learner-centered strategies.**

Index Terms—**Data mining, MOOCs, personalized course list, vocational education.**

I. Introduction

Universal access to high quality education is the key to the building of peace, sustainable social and economic development, and intercultural dialogue [1]. MOOCs are developing a new way to those who are seeking the chance to enhance their skills and broaden their knowledge to move up the career ladder faster, further and easier than they dreamed possible. A wide range of courses are presented, from the Humanities, Social Sciences, Mathematics, Engineering, to Computer Science, and even other disciplines [2], and tens of thousands of participants are attracted to enroll.

The scale of user data generated by MOOCs platform daily during interactions between users and platform has reached TB(Terabyte) level with various data types, including learners' characteristic data, behavioral data, and performance [3], which provides education researchers with invaluable insights about how students respond to the "substance and form" of content at a very fine granularity: every online browsing event is recorded [4].

Data mining is the process of discovering "hidden messages," patterns and knowledge within large amounts of data and making predictions for outcomes or behaviors [5]. The main data mining techniques used to analyze educational data consists of cluster analysis, association analysis, regression analysis, etc. [6].

Generally, a course of MOOCs consists of lectures videos, quizzes, peer assessments, assignments, discussion forum, etc. So far, the existing literature has mainly studied participants' behavior accessing to the above MOOCs elements and then discovered the covering knowledge to support the e-learning improvement of universities and institutes and optimize the settings for course selection [7]. As for MOOCs in vocational education area, the existing researches have focused on conceptive applications, such as building MOOCs by universities and enterprises [8], training employee [9], and teaching reform. To our knowledge, there are few publications about MOOCs in vocational education field based on experimental evidence, yet mentioning the application of data mining technology on MOOCs for developing vocational education.

Vocational education plays a key role in individual professional and career improvement as well as socioeconomic development. Applying data mining technology on MOOCs to develop vocational education becomes necessary and urgent. This paper investigated the existing publications with the themes "data mining in MOOCs" and "MOOCs and vocational education", with the aims to find out the possibility and feasibility of applying data mining technology in MOOCs for developing vocational education and maximizing the value of vocational education.

The rest of the paper is organized as follows: Sections II studied the existing literature and examined user behavior data and explored the main applications. We presented three new recommendations to apply data mining on MOOCs for developing the vocational education with the concrete data mining methods in Section III. Finally, Section IV concluded the paper and suggested future work.

II. Related Work

We located English-language and mandarin-language refereed conferences and journal articles published between January 2012 and July 2017 by searching on Google Scholar and China Knowledge Resource Integrated Database, using key words 'Data Mining, MOOCs, vocational education', 'Data Mining, MOOCs' and 'MOOCs, vocational education'.

Manuscript received September 19, 2017; revised November 12, 2017. This work was supported by Shanxi Institute of Educational Science under Grant GH-17116 and Shanxi Institute of Mechanical & Electrical Engineering.

Jianzhen Zhang is with the Shanxi Institute of Mechanical and Electrical Engineering, Changzhi, 046011 China (e-mail: sxjdzjz@126.com).

Jia Tina Du is with the University of South Australia, Adelaide, SA 5001 Australia (e-mail: Tina.Du@unisa.edu.au).

Fang Xu is with the Soochow University, Suzhou, 215123 China (e-mail: xufangn@suda.edu.cn).

doi: 10.18178/ijiet.2018.8.6.1073

We did not get any papers directly related to ‘Data Mining, MOOCs, vocational education’, which indicates that there is limited research on applying data mining technology in MOOCs for developing vocational education. There are too many results matching with the second and third topics. Those papers being cited more than 30 times and more closely related to the key words were used. The publications are presented in Tables I and II. There are 37 papers related to ‘Data Mining, MOOCs’ and 13 to ‘MOOCs, vocational education’ were selected.

TABLE I: STATISTICS OF THE RESEARCHES ON BEHAVIOR DATA AND DATA MINING APPLICATIONS

Ref	D1	D2	D3	D4	D5	D6	D7	D8	A1	A2	A3	A4	A5	A6	A7
[1]		√	√	√			√	√		√	√				
[2]			√	√	√	√	√		√	√	√	√		√	
[3]													√	√	√
[4]					√			√					√		
[6]									√			√			√
[10]	√								√				√		
[11]	√	√	√		√	√		√			√	√		√	
[12]	√		√		√	√	√	√				√			
[13]	√	√	√		√	√	√			√		√			
[14]						√			√					√	
[15]	√	√			√	√						√			
[16]					√							√		√	
[17]	√				√	√	√		√		√	√			
[18]					√			√	√		√			√	
[19]					√	√	√	√			√	√			
[20]						√	√		√		√	√			
[21]								√	√	√				√	
[22]	√	√	√		√	√				√		√		√	
[23]		√			√	√						√			
[24]	√	√	√	√	√	√		√				√		√	
[25]		√	√			√	√					√	√		
[26]	√	√	√		√		√				√	√	√	√	
[27]			√	√	√		√				√	√			
[28]					√	√	√	√	√	√		√		√	√
[29]				√		√					√	√		√	
[30]			√		√	√	√	√	√		√	√			
[31]	√								√	√				√	
[32]					√		√								
[33]				√	√	√		√	√		√				
[34]		√	√					√	√						
[35]		√	√		√	√						√			
[36]	√	√	√		√		√	√					√		
[37]	√	√	√		√			√		√	√			√	
[38]	√	√	√	√	√	√						√			
[39]		√	√		√		√								
[40]	√	√			√	√					√	√			
[41]					√				√		√		√		
Total	15	16	17	7	26	20	15	15	15	8	15	22	7	14	3

D1= video, D2= quiz, D3= assignment, D4= peer assessment, D5= discussion forum, D6= completion, D7= drop out, D8= course resource.
A1= assistant instruction decision-making, A2= assistant teaching evaluation, A3= improving students' experience, A4= assistant students completion, A5= assistant commercial application, A6= improving MOOCs platform, A7= assistant students selecting courses.

TABLE II: STATISTIC OF MOOCS CONCEPTIVE APPLICATIONS ON VOCATIONAL EDUCATION

Ref	employee training	school-enterprise cooperation	Amplify enrollment	Teaching revolution
[8]	√			√
[9]	√	√		
[42]	√			
[43]	√			√
[44]		√		√
[45]	√		√	
[46]	√			
[47]	√			
[48]	√			
[49]	√			√
[50]	√			
[51]				√
[52]	√			
Total	11	2	1	5

A. Analysis of Enrollers’ Behavior on MOOCs

We divided learners’ behavior in terms of MOOCs characteristics into 8 aspects: watching video, downloading the course resource, taking part in the quiz, submitting the assignment, peer assessment, discussing on forum, completingthe task or getting the course certification, dropping out.

In MOOCs learning-model, participants learn knowledge and skills mainly from watching video offered by MOOCs platform. Videos are central to the student learning experience [11]. Guo studied how video production decisions affect student engagement in online educational videos. Using data from 6.9 million video watching sessions, they found that shorter videos are much more engaging thereby developed a set of recommendations to help instructors and video producers take advantage of the online video format. Kizilcec, Piech and Schneider [12] classified learners based on their patterns of interaction with video lectures and presented MOOCs design directions for potential improvement, Breslow *et al*. [13] examined the students’ use of videos by time spending on. Sometimes, learners download the course resource with the aims of in-deeply study or reinforced knowledge, the course resource refers to video, text, and other lecture notes. Jia, Miu and Wang [14] explored the students’ download behavior changes along with time going on by using 6 MOOCs data offered by Chinese University MOOCs. Just like traditional teaching model, MOOCs consist of quizzes, such as week quizzes, mid-exam test and final exam. Xu, Zhao and Xu [15] judged the learners’ level of mastering the knowledge using the score of every week, Liu *et al*. [16]studied the impact of the final testing grades on students’ experience. Peer assessment is a new learning way created by MOOCs platform. Adamopoulos [2] proposed that peer assessment has a positive effect (0.17, $p<0.01$) compared to automated feedback indicating that better technological solutions for automatically providing feedback and evaluating the assignments of the students are still needed.

Discussion forums are the primary interaction mode among instructors and students designed by MOOCs providers[18], so this kind of data source attracted many researches interesting, for instance, Huang et.al studied contribution behavior in discussion forums to find superposters, Tharindu, Kizilcec *et al*. [17], [19] examined the discussion on forums to identify learners' learning behavior characteristics. Drop out included that course task was not completed and giving up study was at the half-way. Drop out was taken as one of the most valuable research to improve MOOCs. Yang *et al*. [20], [21] explored the case of learners' drop out and presented some early warning system.

Fig. 1 outlines the number of publications about application of data mining on MOOCs from 2012-2017 covered in this paper. As can be seen from the data resource distribution, discussion forum and completion are obviously the main data resource for data mining on MOOCs, presumably because discussing on forum is the important way for learners to explain their opinion on MOOCs study which is also used by MOOCs providers to understand what the learners think and need. After learners completed all tasks they expect to get certification which means their completion or success on their study. As a result, completion rate becomes another hot topic. It is strange that little attention is paid to peer-assessment which is a novel assessment way contrast to tradition education.

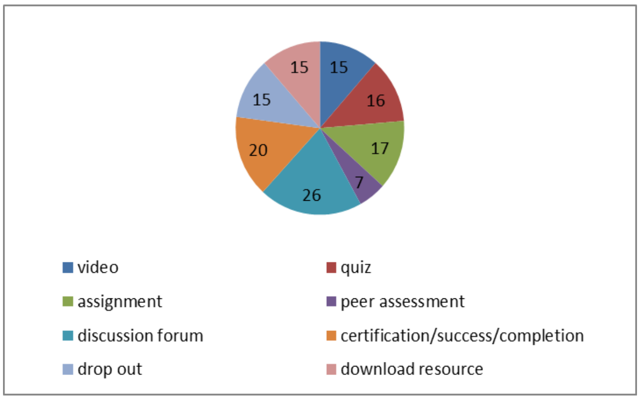


Fig. 1. Distribution of application of data mining on MOOCs from 2012 to 2017.

B. Analysis of MOOCs Existing Application Based on Data Mining Technology

Much hidden information will be obtained from the huge number of data created by MOOCs when applying data mining technology on MOOCs, such as to provide decision-making and maximize the education benefits. Therefore, how to gain insight into the value of the data from the large data in MOOCs platform is becoming an important topic. We found there are 7 categories of applications of data mining on MOOCs, including assistant instruction decision-making, assistant teaching evaluation, improving students learning experience, assistant students completion, improving MOOCs platform, helping students selecting courses, assistant commercial application.

Assistant instruction decision-making is to provide decision-making support, modify teaching model in time, by mining learners' educational background and learning behavior feature. Liu [22] purposed to provide the more reasonable teaching strategy by mining learners' learning behavior. Assistant teaching evaluation focuses on evaluate the teaching activity, which will help on getting the real opinions from students. Jiang, Zhang and Li [23] explored predicting the possibility of learners getting the certificate based on the learning behavior analysis, help to MOOCs teaching evaluation in future. Assistant students completion means helping students completing all tasks required by MOOCs and then getting the certificate by providing the warning and intervention based on mining students' patterns of login, watching video, submitting assignments and so on. Xu *et al*. [24] studied the relationship between students' behavior and success. Improving MOOCs platform needed mining the problems which in a way hindered students gained comfortable experience from MOOCs learning, providing recommendations to MOOCs platform designers. Kay *et al*. compared the main features of the six famous platforms so as to find the most effective MOOCs platform design [25]. Helping students selecting courses is from mining students' educational background and enrollment aim. Goyal and Vohra [7] worked on measuring students retention rate to help students selection of courses. Most of MOOCs platforms offered free courses in previous years, while it is not a sustainable way. Assistant MOOCs platform commercial is being gradually aware. Anderson [26] presented business models for assistant MOOCs platform commercial.

Fig. 2 charts the general trend on application of data mining on MOOCs. It is obviously that helping students' completion attracted most work from 2012 to 2017, which is followed by improving students' experience and assistant instruction decision-making. It is well known that MOOCs has high enrollment as well as high drop-out rate [27], [28], which can be reasonably explained that students' experience is the key factor for students to keep on learning and the key premise for MOOCs getting success. Improving MOOCs platform is another important application area, ángel Fidalgo-Blanco *et al*. [29], [53] compared different platforms to try to find the proper platform for different types of courses. Interestingly, a small number of publications worked on assistant teaching evaluation and assistant students selecting courses. The papers involving assistant commercial application is rare.

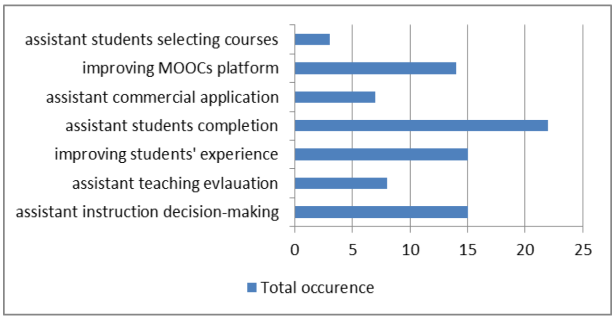


Fig. 2. The general trend on application of data mining on MOOCs.

C. Analysis of MOOCs' Conceptive Application on Vocational Education

Vocational institutes and organizations are gradually aware of MOOCs popularity with higher education and are finding the proper way to apply MOOCs model on vocational education. Interestingly, there are yet little publications on experimental evidence for applying MOOCs on vocational education. Apart from general higher education, vocational education aims to build students work skills, therefore, most of the reviewed publications focused on employee training, teaching reform, school-enterprise cooperation, and enlarging enrollment.

Vocational institutes and organizations are gradually aware of MOOCs popularity with higher education and are finding the proper way to apply MOOCs model on vocational education. Interestingly, there are yet little publications on experimental evidence for applying MOOCs on vocational education. Apart from general higher education, vocational education aims to build students work skills, therefore, most of the reviewed publications focused on employee training, teaching, school-enterprise cooperation, and enlarging enrollment.

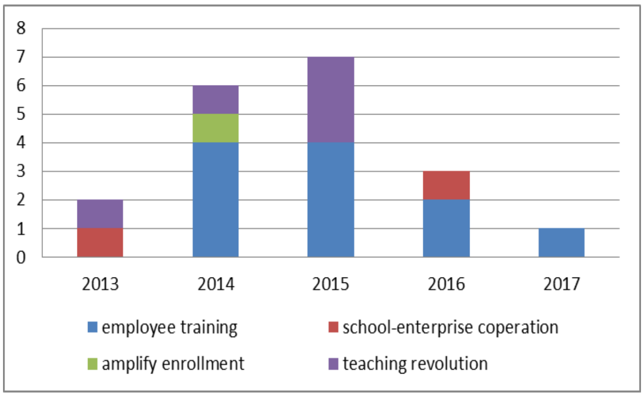


Fig. 3. Distribution of MOOCs in vocational education field.

Based upon the review of existing work, we concluded that there are many publications about the application of data mining on education data but there is not enough research on vocational education. According to Breslow's findings, 1100 out of 6381 students reported they were in their 20s and 30s, and the majority of learner are between 20 to 40 years old, in addition, over half the survey respondents reported the primary reason they enrolled in that MOOCs was for the knowledge and skills they would gain [13], which fully demonstrates that many people registered to MOOCs for improving their workplace competitiveness. Application of data mining on MOOCs for improving vocational education should not be ignored any more.

III. Exploring to Mine MOOCs Information for Serving Vocational Education

Although little experimental researches focused on mining vocational education value from MOOCs data, differences between theoretical and professional MOOCs had been recognized. Li *et al.* [30] divided MOOCs into two types: academic-based and vocation-based, there are obvious professional education features in vocation-based MOOCs, which insisted on market-oriented and student-centered, aimed to improve learners' professional ability. Based on the big data from MOOCs platform, mining valued information for serving vocational education is becoming a hot topic.

We explored to dig MOOCs' professional education value from 3 aspects.

A. Identifying the Target Groups for Vocational Institutes

Clustering is the task of assigning a set of objects into groups called clusters so that the objects in the same cluster are more similar to each other than those in other clusters [7]. Clustering's goal is to find data points that naturally group together [53]. Jiang categorized learners according to reported information by clustering way [31]. In this paper, we aim to categorize learners according to reported address information, which can be informed by analysis IP access and registered location, then courses enrolled by this area were clustered, taking the economic structure of this area into account, the list of popular courses finally will be concluded. Based on the data about enrollment and economic development in this area, vocational institutes can carry out online and offline vocational education publicity targeted this area, which will improve support for local groups of learners [18] as well as help vocational institute success.

Fig. 4 illustrated the process of mining the target area with the cluster analysis. A, B and C represented respectively three courses learner location clustering groups. We can easily find all of three courses shared D round area, which means there is the potential demand for A, B and C courses. The vocational institute offering these courses will be more likely to get success by carrying on marketing in this area. Using MOOCs as a marketing tools transfers the enrolled participants into real registered students was mentioned by the University of London in 2013 [54]. Kumar *et al.* demonstrated the methodology on how to help educational organizations compete with others by mining the hidden enrollment patterns in some area and predict registration [55], [56].

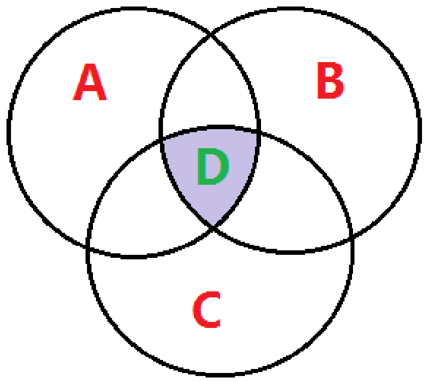


Fig. 4. The process of mining the target area.

B. Figure Out Vocational Course Sets for the Specific Profession

Association analysis is to find if-then rules of the form that if some set of variable values is found, another variable will generally have a specific value [53]. The same type of profession possibly shared more similar requirements on knowledge and skills than differences. Aher and Lobo compared association rule algorithms for course recommender system in E-learning, and found Apriori association algorithms perform better than others in predicting the course selection based on student choice [57]. Vocational institutes and MOOCs platforms can use association rule mining to analyze learner job information registered, associate the courses information registered by those learners sharing the same job, in addition, sequential pattern of learning MOOCs can also be gained. Finally, vocational institute can work out knowledge and skills outline on the learner job and create a special course list for targeted job. Fig. 5 shows the process of generating a special course list.

The well-known MOOCs platforms already practiced developing some special course lists on some research area or professional skills, for example, Coursera platform deliveries some special course list, including Machine learning having 5 courses on the special course list, Big Data with 6 courses on the special course list and so on.

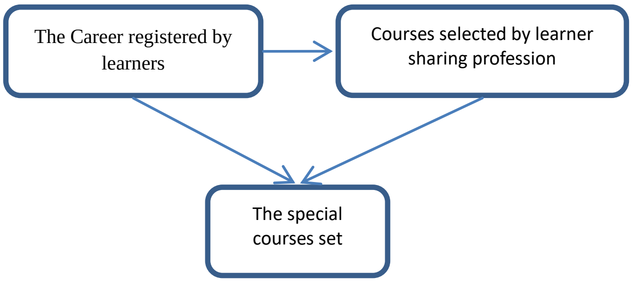


Fig. 5. The process of figure out vocational courses sets.

C. *Recommending the Personalized Career Planning for Learners*

Regression analysis refers to the characteristics of the attribute values in the transaction database in time aspect, producing a function which maps data items to a predictive variable value, and finds the interdependent relationships between variables or attributes. Regression analysis is always used to predict the trend characteristics of data sequences and co-relationship between data.

Distinguished from recommending the courses to learners, recommending the personalized career planning for learners is consistent with the idea of a person-oriented education. In order to make the proper career planning, personal profession, learning habits, and learning performance should all be taken into account. It is feasible and practicable for vocational institutes and MOOCs platforms to work out the personalized career planning for learners with data mining technology in MOOCs information. For personal profession, data can be directly gained from registered information; For learning habits, Wang conducted research on how to build up personalized learning model [3]; For learning performance on MOOCs platform, every online browsing events, such as watching videos, posting on forum, submitting assignments, and completing tasks, are recorded by MOOCs platform. Hou combined the algorithms of data mining with course information to construct the knowledge map and push the personalized course list that learner need quickly and accurately [32]. The proper career planning for learners will be worked out and delivered after mining on frequency, duration and means of learners.

IV. Conclusion and Future Work

He [58] said that MOOCs consisted of educators' learning wisdom, IT experts' technology wisdom, and entrepreneurs' business wisdom. Application of data mining on vocational MOOCs will be an important way to improve vocational education. Based on the review of key literature published from 2012 to 2017 around the topic 'data mining and MOOCs' and 'MOOCs, vocational education', this paper investigated the main user data researched and categorized the applications of data mining on MOOCs. It found limited literature on applying data mining on MOOCs for improving vocational education. Three recommendations were presented, including using data mining to help vocation institute find the target group, figure out a vocational course set for the special profession, and work out a personalized career planning for learners.

However, vocational education aims to enhance workplace skills, so practical skill training is an important part of professional education, which is the key point of vocational education differing from general higher education. It is still a challenge for vocational MOOCs to deliver effective skills training online. In the future, we will classify professional ability, explore the most effective ways to present vocational MOOCs by mining information from video, discussions on forum, and resource use.

Acknowledgment

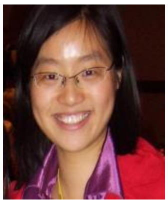
The authors would like to thank the anonymous reviewers for their detailed reviews and constructive comments. Despite their invaluable assistance, any errors remaining in this paper are solely attributed to the authors.

References

[1] T. Daradoumis, R. Bassi, F. Xhafa, and S. Caballé, "A review on massive e-learning (MOOCs) design, delivery and assessment," in *Proc. 2013 Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*, pp. 208-213, 2013.
[2] P. Adamopoulos, "What makes a great MOOCs? An interdisciplinary analysis of student retention in online courses," in *Proc. the International Conference on Information Systems*, pp. 1-21, Italy: ICIS2013, 2013.
[3] Q. P. Wang, "Research on data mining of MOOCs platform based on large data environment," in *Proc. 4th International Conference on Mechatronics, Materials, Chemistry and Computer Engineering*, pp. 1991-1994, China, 2015.
[4] F. Dernoncourt, C. Do, and S. Halawa, "MOOCsViz: A large scale, open access, collaborative, data analytics platform for MOOCs," *NIPS Workshop on Data-Driven Education*, pp. 1-8, 2013.
[5] J. Luan, "Data mining and its applications in higher education," *New Directions for Institutional Research*, vol. 113, pp. 17-36, 2002.
[6] R. Vohra and M. Goyal, "Applications of data mining in higher education," *International Journal of Computer Science Issues*, vol. 9, no. 2, pp. 113-120, March 2012.
[7] Mycos. The number of Chinese MOOCs users is over 10 million by this Year, Does it work well? [Online]. Available: http://www.sohu.com/a/115912532_121294
[8] S. Xie and J. Xia, "Vocational education development and reform in MOOCs area," *Chinese Vocational and Technical Education*, vol. 33, pp. 17-20, 2014.
[9] M. Ursache, E. C. Loghin, M. AvĂdanei *et al.*, "The use of massive open online courses (MOOCs) in textile education," in *Proc. the 12th International Scientific Conference eLearning and Software for Education*, pp. 532-535, Bucharest, April 2016.
[10] P. J. Guo, J. Kim, and R. Rubin, "How video production affects student engagement: An empirical study of MOOCs videos," in *Proc. the First ACM Conference on Learning@ Scale Conference*, pp. 41-50, USA: ACM (2014), 2014.
[11] R. F. Kizilcec, C. Piech, and E. Schneider, "Deconstructing disengagement: Analyzing learner subpopulations in massive open

online courses," in *Proc. the Third International Conference on Learning Analytics and Knowledge*, pp. 1-10, 2013.
[12] L. Breslow, D. E. Pritchard, and J. Deboer, "Studying learning in the worldwide classroom research into edX's first MOOCs," *Research & Practice in Assessment*, vol. 8, pp. 13-25, 2013.
[13] Y. Jia, J. Miu, and Q. Wang, "Big data analysis of learning behavior and performance on MOOCs," *Industry and Information Technology Education*, vol. 9, pp. 23-29, 2014.
[14] S. Xu, Q. Zhao, and J. Xu, "Analysis of monitoring indexes of MOOCs learning process based on data mining," *Modern Education Technology*, vol. 27, pp. 119-126, 2017.
[15] Z. Liu, W. Zhao, X. Chen. Data analysis and mining of MOOCs data. From Circuit Analysis Course. [Online]. Available: http://kns.cnki.net/kcms/detail/51.1653.N.20170503.1114.004.html
[16] J. Huang, A. Dasgupta, and A. Ghosh, "Superposter behavior in MOOCs forums," in *Proc. the First ACM Conference on Learning@ Scale Conference*, pp. 1-10, 2014.
[17] T. R. Liyanagunawardena, "MOOCs experience: A participant's reflection," *ACM SIGCAS Computers and Society*, vol. 44, pp. 9-14, 2014.
[18] R. F. Kizilcec and E. Schneider, "Motivation as a lens to understand online learners: Toward data-driven design with the OLEI scale," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 2, pp. 1-24, 2015.
[19] D. Yang, T. Sinha, and D. Adamson. "Turn on, tune in, drop out anticipating student," in *Proc. the 2013 NIPS Data-Driven Education Workshop*, pp. 1-14, USA, 2013.
[20] Y. Chen and M. Zhang, "Analysis and warning of students drop-out on MOOCs," *Industry and Information Technology Education*, vol. 9, pp. 30-36, 2014.
[21] Z. Liu, "Application of data mining technology in MOOCs," *J. Jianghan Univ. (Nat. Sci. Ed.)*, vol. 43, pp. 367-370, 2015.
[22] Z. Jiang, Y. Zhang, and X. M. Li, "Learning behavior analysis and prediction based on MOOCs data," *Journal of Computer Research and Development*, vol. 3, pp. 614-628, 2015.
[23] Z. Z. Xu, D. Goldwasser, and B. B. Bederson, "Visual analytics of MOOCs at Maryland," in *Proc. the First ACM Conference on Learning@ Scale Conference*, pp. 195-196, 2014.
[24] J. Kay, P. Reimann, and E. Diebold, "MOOCs: So many learners, so much potential," *IEEE InTeLLIGenT SySTemS*, vol. 6, pp. 2-8, 2013.
[25] T. Anderson, "Promise and/or Peril: MOOCs and open and distance education," *Commonwealth of Learning*, vol. 3, pp. 1-9, 2013.
[26] D. Coetzee, A. Fox, and M. A. Hearst, "Should your MOOCs forum use a reputation system?" in *Proc. the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 1176-1187, USA, 2014.
[27] M. M. Wen, D. Yang, and C. P. Rosé, "Sentiment analysis in MOOCs discussion forums," in *Proc. the 7th International Conference on Educational Data Mining*, pp.130-137, 2014.
[28] F.-B. Ángel *et al*., "Methodological approach and technological framework to break the current limitations of MOOCs model," *Journal of Universal Computer Science*, vol. 21, pp. 712-734, 2015.
[29] D. Gašević, V. Kovanović, and S. Joksimović, "Where is research on massive open online courses headed? A data analysis of the MOOCs research initiative," 2017.
[30] M. Li, Y. Zhang, and F. Ye, *Decoding MOOCs: Educational Research on Massive, Open, Online Courses*, Beijing: Tsinghua University Press, 2013.
[31] X. F. Jiang, "Study of MOOCs design based on data mining," *China Computer & Communication*, vol. 8, pp. 188-191, 2016.
[32] J. M. Hou, R. F. Liu, and Y. Zhang, "Constructing knowledge map for MOOCs using data mining methods," in *Proc. International Conference on Management Science, Education Technology, Arts, Social Science and Economics (MSETASSE 2015)*, USA: Atlantis Press, pp. 938-943, 2015.
[33] M. A. Chatti, A. L. Dyckhoff, and U. Schroeder, "A reference model for learning analytics," *International Journal of Technology Enhanced Learning (IJTEL) – Special Issue on "State-of-the-Art in TEL*," pp. 1-9, 2012.
[34] F. Han, K. Veeramachaneni, and U.-M. O'Reilly, "Analyzing millions of submissions to help MOOC instructors understand problem solving," *NIPS Workshop on Data Driven Education*, pp. 1-5, 2013.
[35] A. Ramesh, D. Goldwasser, and B. Huang. (July 2017). Modeling learner engagement in MOOC using probabilistic soft logic. [Online]. Available: https://www.umiacs.umd.edu/~hal/docs/daume13engagementmooc.pdf
[36] G. K. Balakrishnan, D. Coetzee. (July 2017). Predicting student retention in massive open online courses using HiddenMarkov models. [Online]. Available: http://www.eecs.berkeley.edu/Pubs/TechRpts/2013/EECS-2013-109.html
[37] D. O. Bruff, D. H. Fisher, and K. E. McEwen, "Wrapping a MOOC: Student perceptions of an experiment in blended learning," *Journal of Online Learning and Teaching*, vol. 9, pp. 1-22, 2013.
[38] S. H. Jiang, A. E.Williams, and K. Schenke, "Predicting MOOC performance with week 1 behavior," in *Proc. the 7th International Conference on Educational Data Mining*, pp. 273-275.
[39] J. Arguello and K. Shaffer, "Predicting speech acts in MOOC forum posts," in *Proc. the Ninth International AAAI Conference on Web and Social Media*, pp. 1-11, 2015.
[40] K. Chao and M. Wu, "Analysis and predicting of learning behavior based on edX platform," *Distance Education in China*, vol. 6, pp. 54-59, 2016.
[41] R. Eynon, I. Hjoth, T. Yasseri, and N. Gillani, *Data Mining and Learning Analytics: Applications in Educational Research*, Wiley, 2016.
[42] E. F. Zeer and A. V. Streltsov, "Technological platform for realization of students' individual educational trajectories in a vocational school," *Mathematics education*, vol. 11, pp. 2639-2650, 2016.
[43] L. R. Goldberg, E. Bell, and C. King, "Relationship between participants' level of education and engagement in their completion of the understanding dementia massive open online course," *BMC Mdeical Education*, vol. 60, pp. 1-7, 2015.
[44] M. Gao, X. Chen, and Y. X. Fang, "Reforming higher vocational education with MOOCs," *Chinese Light Industry Education*, vol. 6, pp. 70-72, 2013.
[45] L. Czerniewicz, A. Deacon, and J. Small, "Developing world MOOCs: A curriculum view of the MOOC landscape," *Journal of Global Literacies, Technologies, and Emerging Pedagogies*, vol. 2, pp. 122-139, 2014.
[46] G.-R. Israel, A.-H. Carlos, and Pérez-Sanagustín, "Scaffolding self-learning in MOOCs," *European MOOCs Stakeholders Summit*, pp. 43-49, 2014.
[47] M. Schmidt, C. D. Valentin, and A. Emrich, "A social and Personalized Learning platform for vocational social media education," *Multikonferenz Wirtschaftsinformatik 2014*, pp. 1-22, 2014.
[48] R. L. Guimarães and A. B. Mattos, "Exploring the use of massive open online courses for teaching students with intellectual disability," pp. 343-344.
[49] B. András, "More visual content into vocational education," *Opus et Educatio*, vol. 2, pp. 39-47, 2015.
[50] V. Shenoy, "MOOC: Recent advancements in personal trainings," *International Journal of Management, IT and Engineering*, vol. 5, pp. 39-44, 2015.
[51] H. Ming, "On higher vocational education in the era of the MOOC," *Journal of Wuhan Institute of Shipbuilding Technology*, vol. 2, pp. 4-7, 2015.
[52] M. Egloffstein and D. Ifenthaler, "Employee perspectives on MOOCs for workplace learning," *TechTrends*, vol. 61, pp. 65-70, 2017.
[53] S. J. Baker, "Data mining for education," *International Encyclopedia of Education (3rd edition)*, vol. 7, pp. 112-118, 2010.
[54] University of London International Programmes, *Massive Open Online Course (MOOCs) Report2013*, 2014.
[55] V. Kumar and A. Chadha, "An empirical study of the applications of data mining techniques in higher education," *International Journal of Advanced Computer Science and Applications*, vol. 3, pp. 80-84, 2011.
[56] C. Sandeen, "Integrating MOOCs into traditional higher education: The emerging "MOOCs 3.0" era," *Change: The Magazine of Higher Learning*, vol. 45, pp. 34-39, 2013.
[57] S. B. Aher and L. Lobo, "A comparative study of association rule algorithms," *International Journal of Computer Applications*, vol. 39, pp. 48-52, 2012.
[58] Y. He, S. Y. Chen, and Z. Lin, "Presenting a new education model in MOOCs view," *Internet of Things Technologies*, vol. 9, pp. 98-99, 2016.

Jianzhen Zhang is a lecturer in Shanxi Institute of Mechanical and Electrical Engineering. She was born in China in 1980. She got her master degree in computer science in Taiyuan University of Technology, Taiyuan city, Shanxi province, China. She has published more than 10 academic and teaching papers, excluding a patent on computer security technology.



Tina Du is senior lecturer of information studies in the School of Information Technology and Mathematical Sciences, University of South Australia. Tina is also an Australian Research Council (ARC) DECRA Fellow (2017-19), a highly competitive national research fellowship, and visiting fellow at the University of Oxford's Internet Institute (2016-17). She has published 63 peer-reviewed work, including one research book, three (3) scholarly book chapters and 59 papers in refereed journals and conference proceedings, including 15 SCI/SSCI journals.

Fang Xu is an associate professor at Soochow University. He got his PhD at Nankai University majored in LIS. He also is a member of ASIS&T and a visiting scholar at University of South Australia. He is a peer review of the journals like “Information & Management” (SCI), “Knowledge-based System” (SCI), and so on. His research interests focus on internet products and user experience, adoption and use of information systems in the Chinese context, competitive intelligence and crisis early warning. His publications have appeared in such journals as Journal of Library Science in China, International Journal of Information and Computer Science, among others.