

Towards a Paradigm of Using Big Data in Education to Develop MOOCs

Qian Wang

Abstract—The Massive Open Online Course (MOOC) movement represents a new wave of enthusiasm for online resources development. Informed by relevant literature, the author designed a paradigm of using big data in education to develop MOOCs. It is argued that MOOC development can benefit from big data in education, the use of which enables MOOC developers to better reflect MOOC values in the overall design, better control the development process, and better meet users' needs. Big-data-driven MOOCs are characterized by better fit between developers' intentions and users' needs, enhanced usability and stronger adaptability. Issues concerning the use of big data in decision making, content selection, forms of expression and feedback mechanism are discussed, with the MOOC entitled "Psychology and Healthy Life" as a case in point.

Index Terms—MOOC, big data in education, development paradigm.

I. INTRODUCTION

With the coming of the data era, the "big data" concept which started from the astronomy and genetics has been applied in all fields that human beings committed to developing, marking that the human beings have taken a huge step forward in the period of seeking understanding of this world by way of quantization. In educational field, the opinion that big data change the educational research and educational practice has been proved with the increasing value of big data in education. While the educational experts are discussing the impact of big data on education, they are attracted by MOOC (Massive Open Online Course), the prominent feature of the higher education discourse in recent years as well. A MOOC is a model for education delivery typically defined as, "massive, with theoretically no limit to enrollment; open, allowing anyone to participate, usually at no cost; online, with learning activities typically taking place over the web; and a course, structured around a set of learning goals in a defined area of study" [1]. In 2012, a handful of key players, including Coursera, edX, and Udacity, have been instrumental in the development of the movement. As it were, MOOC had a great influence on the curriculum design, teaching and assessment mode of the traditional notions of classroom. However, when learners entered into MOOC learning environment, they are encountered with a lot of

problems. On the one hand, they are lack of motivation for continuous learning; On the other hand, they experience inadequate and ineffective interaction either with instructors or peers, which give rise to the low completion rate of MOOC. All these problems initiated the introspection and reflection upon MOOC development in educational circles. Although MOOC has been accepted by the mainstream within one year or even less than one year and had a space in educational circles, only the minority adopted MOOC[1]. Obviously, scholars' poise and reflection on the bubble of hype in the MOOC are grounded because MOOC as a matter of fact, did not collect the learning state and other relevant data of learners to improve its development paradigm. It is found that during the development of MOOC, many problems need to be resolved, such as ways of analyzing online learners' learning practice, relations between new technologies and learning mode, optimal teaching methods engaged in the teaching process and channels of adapting to the new learning environment, etc. In the discourse system supported by the big data researcher, the big data in education is the approach.

II. BIG DATA IN EDUCATION AND MOOC RESOURCE CONSTRUCTION DILEMMA

A. Big Data Thinking in Education

Since 2012, the latest data storage technology represented by massive data storage technology and real-time data warehouse technology have appeared, marking that the age of big data truly arrived. The ability of collecting, managing and analyzing data was greatly improved with the development of data mining technology. Data, contemporarily, is decisive to promote the social development and thus, the big data thinking becomes the core quality of modern society. According to Viktor Mayer, to realize big data thinking, analyzing things needs to focus on all relevant data rather than the part of data; we need to be willing to accept all kinds of data rather than the accurate data; we do not need to always seek the causal relationship but pay attention to the correlation between things [2]. This kind of thinking urges the education researchers to use big data to make decisions and predictions. Application of data is gradually extended in depth and breadth, which not only helps the education researchers to explore and predict the learning behavior of learners, better understand the learning attitude of learners, but also provides the learners with the appropriate learning materials, supporting service and instructors' intervention so as to improve the teaching and learning efficacy.

B. MOOC Resource Construction Dilemma

Manuscript received March 30, 2016; revised May 19, 2016. This study is supported by the Fundamental Research Funds for the Central Universities under grant No.3102014RW0023, Higher Education Research Fund under grant No. 2014GJY09 and College Curriculum Reform Funds under No. 14GH0166.

Qian Wang is with Northwestern Polytechnical University, Xi'an, Shaanxi, China (e-mail: wangqian1981126@126.com).

Since 2012, MOOC has been booming and chased after by universities around the world, and fueled by medium and researchers in educational circles. However, the problems hidden behind the zeal were exposed as well. The skeptics hold that MOOC do not resolve the problems brought about by distance education like lack of human warmth, lack of culture interaction, lack of in-depth and thought-provoking studying resources etc., which are deemed even more “inhumane” than the past online education. For example, interaction of MOOC mainly relies on the mutual evaluation coming from thousands of students, but it is a pity that even their work cannot be guided by instructors; although MOOC realizes the connection via technology, it cannot ensure the interaction [3]. At the symposium of Massachusetts Institute of Technology, the famous distance education expert, John Daniel, expressed his concerns that MOOCs, actually, ignore the developmental experience of open courses for 40 years [4]. What’s more, disagreement also appeared in resource construction of MOOC. The “knowledge copying” principle, which deviates from the exchanging ideas with students, was emphasized and developed wildly in MOOC environment. In terms of courses design, even no MOOC operator hired the people who have accepted the professional training about teaching design, educational technology and curriculum design. Especially, due to the dispersion and poorly-organized courses, there is no way to judge whether the learners complete learning the course or not. So to speak, MOOC experienced the enthusiasm in the early stage and entered into the calm reflection in the middle stage.

C. Resolving MOOC Resource Construction Dilemma with Big Data in Education

The best learning method is the individualized learning, which needs to be expressed reasonably by data embedded in resources. MOOC resources are no exception. Taking *Artificial Intelligence* taught by Thrun and Norvig as example, quality and practicability of this resource both have the public value, which was proved by 160,000 learners in the world. However, not everyone persists with learning with *Artificial Intelligence*. Among the learners, only 14% of them finished the course. The unsatisfying completion rate leads us to reflect on how MOOC responds to the learning needs of individuals? It needs to mine the related information during learning so as to form diversified resources catering to different levels of needs of learners. The main reason that could explain the tough situation of MOOC is that it does not provide education catered to the needs of the learners, and it lacks of cultural environment suitable for online learning. The current resource development paradigm places too much emphasis on the effectiveness of knowledge but ignores the needs of students, and attaches more importance to knowledge transmitting but ignores the acceptance of students. To our delight, big data in education provide us with the new paradigm since learners participated in MOOC course outnumber learners in traditional classroom, and open online education forms are diversified compared with the past closed network environment as well. Therefore, a lot of meaningful data, for example, length of the videos watched, records of participation in discussions and other activities after video-watching, the accuracy rate in exercises can be

produced. These data can contribute to analyzing students’ preferred learning resources, ways of constructing resources and directions of modifications after the resource has been established, etc.

III. MOOC RESOURCE DEVELOPMENT PARADIGM FROM BIG DATA IN EDUCATION

Paradigm is an interpretative framework, which is guided by a set of beliefs and feelings about the world and how it should be understood and studied. It is the shared belief, research tradition and value as well as accepted research method of an academic community. Aggregation, analysis and research of mass data are the pillar to promote the scientific development in big data era. Thus, data is called “the fourth paradigm of scientific research” by the computer expert J.Gray in America [5]. Development of educational science needs to face the new challenges that the social transformation brings to the education, needs to respond to the problems appeared in the education and teaching reform, and needs to evoke and intensify the self-awareness of the educational research. In fact, the role of big data in the educational science research can effectively helps to alleviate these problems; meanwhile, it also can solve all kinds of problems existing in MOOC resource construction. But, under the framework of big data in education, what should be considered to guide MOOC resource development?

A. MOOC Resource Development Principle Based on Big Data in Education

1) Big data awareness — the basic element to implement MOOC resource construction

Big data thinking should penetrate in every aspect of MOOC resource construction such as project approval, course design, resource construction, application and evaluation. With respect to the project approval, the data needs to be collected to demonstrate the online learning adaptability of courses, and for the courses with adaptability are supposed to consider the different needs of learners; with respect to course design, the learning state of learners, whether the teaching design implemented around knowledge points can disabuse learners and above all, time and place that the learners took part in learning are all useful data in affecting the learning outcome; with respect to the resource construction, it needs to fully consider how to collect the complex interactions between learners and resources, between learning peers as well as between learners and supporting personnel to trace the origin and characteristics of intervening data during learning; with respect to application, changes of learners’ state through human-computer interaction, including the changes of learning state and emotion in particular, etc. need to be collected. These data shall be gathered and sent to feedback system. Thorough analysis and modeling should be done so as to improve MOOC resources.

2) Overall design based on big data in education — the value standpoint of MOOC resource construction

The study habit and psychological state of mind of online learning cause the fragmentation in current education

resource construction, which is in line with the learning characteristics in digital age. Along with a large amount of micro-resources produced by the micro-information environment based on Web2.0 technology, the whole digital system is indeed characterized by fragmentizing resources. However, this kind of fragmentized learning has been criticized by educators for it, by no means, is beneficial to systematic command of knowledge. Researchers of online education started to research on how to outline the map for learners to acquire knowledge systematically, and help them to “fill the gap” during knowledge construction caused by micro-resources. However, this kind of help always occurs afterwards. The future MOOC resource construction needs the help of big data. There is no denying that the target-driven and goal-oriented learning is the effective mode to help learners to construct knowledge, especially in building up an environment which is fit for evocation, application and innovation of knowledge. For example, the “Advanced Mathematics” needs to be practiced intensively around a variety of knowledge points, so the mathematics is divided into different knowledge points and then the short and micro MOOC resources are constructed around these knowledge points, which has the obvious advantage for the learners to command. However, about the deep analysis of personality or characteristics of Wang Xifeng, the well-known figure in *Dream of Red Chamber*, we need referencing a lot of materials from her living environment and individuality formation process, and verify that information during her interacting with others. It will take a long time to help learners to analyze and form the general impression of Wang Xifeng’s personality in their mind. So, MOOC construction based on big data needs to integrate the text material, digital resources, running mode, marketing method, evaluation and data collection in the whole process of MOOC resource development to better accommodate learners’ different needs and types of motivation. What should be done is scientific prediction is supposed to be made through collecting the critical, sensitive and representative attributes of factors so as to ensure all kinds of resources in MOOC have the corresponding big data support, which can better reflect the value of MOOC in the aspect of improving personal comprehensive ability.

3) *Understanding big data in education — the internal logic of MOOC resource construction process*

Only the resource is understood, can it produce value. For a long time, the educational resource developers pay attention to learners’ experiential learning, but they are in the dark about what kind of experience in the spatial-temporal separation environment is better for learners; they always try their best to reduce the cognitive load of learners, but they are not clear what kind of load is best for the learners to comprehend and consolidate the knowledge; they understand loneliness and anxiety of online learning, but they have no idea what kind of psychological support and interaction can promote the effective cooperation among learners; they desire to obtain the needs of learners for resources through conversation, but they are unable to get the effective sample through questionnaire or survey. However, these troubles can be driven away in the logic of big data in education. A large

amount of data could be collected through all kinds of channels and expressed through graphs to serve as guidance. In addition, the team composed of resource constructors, teaching designers and course experts can effectively analyze and probe into the learning state of learners through gathered data, which can help to form the “zone of proximal development” suitable for learners. Effect of big data on the measurement is like the effect of microscope on the existence of material world [6]. Data about learners’ background, learning process and outcome and even data about subtle change of mind in learning process caused by clicking and touching keyboard each time could be obtained, which are very valuable for developing appropriate learning environment and cognitive load for learners, and certainly, can effectively promote communication and conversation between teaching and learning.

IV. MOOC RESOURCES CHARACTERISTICS BASED ON BIG DATA IN EDUCATION

A. *High Matching Degree*

In the traditional resource development paradigm, the developers analyze the feasibility and necessity of resource development, conduct the teaching design and embark on a series of resource development efforts relying on the questionnaire or survey in “small data” age. But, processing this kind of data needs the proper logic reasoning and theoretical analysis of conductors, which actually, to some degree, leads to the inaccurate interpretations of real learners’ needs. That can, partially, account for the fact that a lot of learners terminate the study in the middle way in MOOCs. Big data in education paradigm is a kind of learning model which can connect the knowledge, motivation, metacognition, attitude and behavior of learners without being intervened by human arbitrariness. Therefore, it can improve the matching degree between a multitude of learning resources and learners significantly.

B. *Good Practicability*

The big data in education can enhance the practical value of MOOC resources. First, project approval of MOOC guided by big data is based on the correlation between learning needs and resource construction, which guarantees the practicability of MOOC.; second, it can ensure the MOOC resource construction satisfy the needs of most learners through analyzing the needs shown by “fragmented” and “entire section” of teaching content; third, MOOC resource construction based on big data can increase mutual understanding between learners and MOOC resources. Data can “outline” the preferred learning time of learners and completion status of assignments, through which we can clearly know to what extent learners have understood learning materials, thus, practical value and circulation of knowledge can be ensured.

C. *Strong Adaptive Capacity*

In the logic of big data in education, MOOC resources can gradually evolve into the diversified resources featured by different levels, different types and different approaches. The

diversification and oneness, universality and specialty of resources, with the help of feedback guided by big data, can help learners to improve learning effectively and meanwhile, facilitate the flow of resources available online and their own needs for resources. For example, when recommending it to the learners, Coursera wrote: “Our course design, deeply rooted in reliable educational basics, can help you to master the new knowledge quickly and efficiently.” It is the utilization of big data in courses design that Coursera guaranteed the quality and practicability of their resources.

V. DEVELOPMENT METHOD AND PROCEDURE BASED ON BIG DATA IN EDUCATION

In accordance with the concept of paradigm, for MOOC resource development paradigm supported by big data, the big data as shared belief and research tradition need to be properly expressed in the practice of MOOC resource construction [3]. Specifically, from the aspect of big data, the following basic problems need to be concerned with: first, the necessity of project approval of course construction based on big data have to be identified; second, the big data generated in the platform when used by learners need to be collected and gathered; third, the dynamic resource system whose depth and breadth can be controlled through big data need to be developed; fourth, the feedback mechanism based on big data needs to be established to optimize resources. The method and procedure should be carried out in line with the needs and expectation of learners. Fig. 1 shows the development method and procedure: MOOC project is approved guided by MOOC construction target and the data about learners’ needs and expectations for the course; confirm the content and form of MOOC, and construct its resource; interaction data are generated through the learning experience of learners and the appropriate support service of teachers; the process data of study and feedback data are formed through course monitoring and measurement; data about expectation of learners and demand of the course are generated through data model. Procedure and method of whole frame are realized relying on big data, including the core steps as follows:

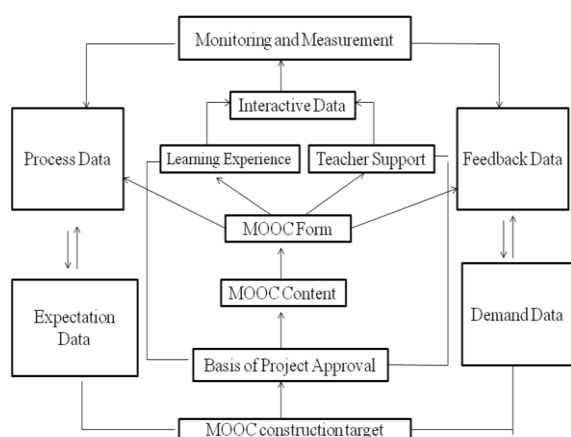


Fig. 1. MOOC resource development method and procedure based on big data in education.

A. Project Approval Based on Big Data in Education

According to causal logic, whether the resource needs to be constructed depends on defect of knowledge and learning

ability of the target population. MOOC resource is constructed according to the data from expectation of learners and course standards, combining with learners’ majors and social needs. The resource construction decision based on big data can fully and truthfully reflect the correlation between explicitness and implicitness during learning. From the data analysis for 179 largest listed companies, it can be found that 5% of growth rate cannot be explained. There are also a lot of correlations that cannot be explained with causal relationship in MOOC. This is also the reason why various educational decisions depend more and more on the big data. Establishment purpose of Program for International Student Assessment (PISA) organized by Organization for Economic Co-operation and Development (OECD) is to provide big data in education for satisfying the decision-making of government and organizations. They measured the data mainly to discover the student characteristics and learning experience in and outside of classroom related to the learning performance of students, and to make decisions according to these data. MOOC resource construction needs to deal with a lot of problems relying on big data as well. For example, although MOOC ecological construction needs a lot of course resources, the order of priority still needs to be followed. We should find the problems that need to be resolved first from resources based on learners’ needs, their major-related knowledge development needs and schedulable resources. The reason why A university has given the priority to the course of “Advanced Mathematics” is that the traditional teaching method of Advanced Mathematics met more and more problems, phenomenon of seeking for help in forum and retrieving relevant resources emerged, and downloads of such course at learning platform increased sharply; the passing rate of such course decreased, and the “incompetent teacher” of this course became the shared belief. Integrated these relevant data and combined with MOOC universal characteristics, it is reasonable to select the course-“Advanced Mathematics” to be first addressed in the MOOC construction project.

B. MOOC Content Selection Based on Big Data in Education

For MOOC resource construction, we have to consider the learning ability and pre-existing knowledge of learners. Applying the same resources for different levels of learners cannot realize the original intention of project approval of MOOC, which is also the main reason why we need carry out the learning analysis. Learning analysis is a kind of action which can discover the problem as soon as possible and implement the intervention strategies through detecting and predicting the learning process of students [7]. MOOC construction requires different levels of learning resources, and learners need to be recommended to appropriate levels of courses so that they can reach the required learning basis before entering into core content of the course. Through the learning data of the course “Psychology and Healthy Life” in 10 years shown in the open education platform of a university, the school-level excellent course “Psychology and Healthy Life” that imbibes MOOC concept extracted critical, highly sensitive and displayed words of such course: happiness (unhappiness) (166), grow up (144), learning (77), emotion (emotional control) (264), love (111), sleeping (206),

psychological problems (mental disease, depression) (86) and other seven types of words. Integrating students' concern with fragmentizing learning style that the learners are used to, relevant learning materials including 15 chapters and 64 sections are designed. In addition, the questions are set when the learners enter into learning experience, through which the compulsory and core contents that learners need to command are identified. Based on that, learners are given the options to select the learning content by themselves. Of course, the "Psychology and Healthy Life" is the course targeted for general education, so the previous experience of learners has limited influence on the learning outcome. But for most courses, relation between previous learning experience and the later result is of great important. If previous problems are not solved, the subsequent chapters will be much more difficult to understand. Therefore, we need to develop appropriate level of materials on the basis of big data. We shall ensure that the learners can enter into study easily in early period, but requirement for knowledge should not be lowered in later period to bring about the progressive learning on part of learners.

C. MOOC Form Based on Big Data in Education

MOOC mainly includes two forms: one is cMOOC based on connection, which focuses on using the connective knowledge to establish MOOC learning environment; another one is xMOOC based on behavior, which extensively utilizes micro-video, test and automatic feedback, etc. In the actual application process, two forms both have market space to run and practical value to fulfill, especially, faced with the different learning habits of so many different learners. Even though the difference of learners is not to be considered, the learning content as a matter of fact, sets up different requirements for MOOC resources. For example, for the study of procedural knowledge, xMOOC may be more effective. In the step-type chapters resolving the core concepts of "Psychology and Healthy Life", such as "Healthy life: whose field?" or "self-regulation of sleep problem", various concepts are explained, so xMOOC is more valuable. Learning for declarative knowledge, however, needs to build more complex learning environment, so cMOOC usually can obtain better effect. For example, the learning content of "Psychological health: can we do it by ourselves?" of "Psychology and Healthy Life" course has stronger subjectivity, thus we need to develop our own opinions on this kind of question based on our own living practice. According to the discussions about "Psychology and Health" in original open educational platform of this university, learners differ on this kind of question a lot. Therefore, it needs to leave more space for learners to think. What should be noted is that there is also dispute about length of MOOC. In 2013, micro course contest was held for 4 times in educational circles in China! The relative subjects and research papers showed a rapidly increasing tendency. Accordingly, many people keep skepticism and hold that such tendency will disappear soon just like previous upsurge of the information technology. It cannot be denied that fragmentized knowledge caused by the micro course may bring certain obstacles to establishment of systematic knowledge, but undeniably, such education resource also has strong adaptive capacity to special scenarios

of the digital age. "Small, micro, and focus" is the establishment mode of MOOC resource represented by the micro course theory and it can meet demands in specific period and a target population. In MOOC camp, a teaching video is generally controlled within 5-15 min and it can highlight the key points and difficult points so as to reduce cognitive load of students. The aforementioned systematic explanation also plays an irreplaceable role in many occasions, so we need to collect characteristics of all kinds of learning contents and classify them reasonably to form appropriate length. Knowing the length required by resource also needs to rely on the big data. We need to analyze the length that the course resources to be constructed shall show, by acquiring the relevant information of a large number of learners. For example, for "Advanced Mathematics", the micro-lectures centered on knowledge points can reduce the cognitive load to facilitate command of knowledge. In the course "appreciation and analysis of literature", analysis of figure's individual characters similar to Wang Xifeng needs the resources constructed with long time, however, about development history of literature, microlectures are more suitable for the needs of learners. Similarly, according to the experience of author participating in the "Psychology and Healthy Life" course resource development, for the contents that the learners need reflection, too short resources are not enough to make learners propose their own opinions; for the concepts that need to be consolidated by stage, too long content affects the accuracy in problem-solving like math problems. In short, the huge advantages of carrying out research based on big data is that it can help to overcome the negative effect caused by intuitional experience and theory prediction in education through integrating different learning features with systematic resources construction.

D. Feedback Mechanism Based on Big Data in Education

Feedback mechanism based on big data should be established to optimize MOOC. Taking advantage of feedback, we can make decisions on project launching and content selection. Feedback data stems mainly from learning experience, learning interaction, learning support, supervision and measurement. We shall prepare for effectively extracting the relevant data during learning by collecting data in the whole process including learners' participating in learning, immersion in learning (quitting learning), completing learning and resource push learning aimed at cultivating good learning habit. Of course, it is not an easy task. On the one hand, the course constructors need to be clear about data type and effect; on the other hand, the data recording technique is required. Realizing the goal needs the experienced data mining experts to develop the platform that can be transferred and copied, which could not be achieved if the collaborative innovation of education experts are absent. Learners need to learn through designed teaching content in advance. Learning performance data will be produced in this process, and then the system will construct the model according to performance so as to adjust content and form in time. This kind of feedback mechanism can not only function in the subsequent development of MOOC, but also is the basis of MOOC resource improvement and innovation. These data obtained can be applied to cognitive load controlling and

progressive resource construction in the future resource construction and development of MOOC, so it is a safe bet that MOOC could become the “organism” with vitality that can grow and flow freely.

VI. CONCLUSIONS

Specific to the future resource constructions including MOOC resource construction, big data in education is still “promising” and “creative” [8]. Big data in education provides infinite possibilities for the future development of MOOC, which can really realize freedom in learning. But at the same time, we would like to remind some open universities or institutions, which are eager to promote their own connotation and create their own brands through MOOC construction, to pay attention to the influence of big data education paradigm when developing an effective online learning environment so that better-informed MOOC design decisions can be made, and the programs on offer can be motivating and pedagogically effective.

ACKNOWLEDGMENT

This study is supported by the Fundamental Research Funds for the Central Universities under grant No. 3102014RW0023 and Higher Education Research Fund

under grant No. 2014GJY09

REFERENCES

- [1] Educause. (2013). Seven things you should know about MOOCs II. [Online]. Available: <http://net.educause.edu/ir/library/pdf/ELI7097.pdf>
- [2] J. Baggaley, “Reflect on MOOC,” *Open Education Research*, 2014, pp. 9-11.
- [3] D. Holton. What’s the “problem” with MOOCs? [Online]. Available: <http://edtechdev.wordpress.com/2012/05/04/>
- [4] J. Daniel, “MOOCs: What lies beyond the trough of disillusionment?” presented at LINC Conference, Boston: MIT, 2013.
- [5] S. Cooper and M. Sahami, “Reflections on Stanford’s MOOCs,” *Communications of the ACM*, vol. 56, no. 2, pp. 28-30, 2013.
- [6] V. Mayer-Schönberger, “Big data,” Hangzhou: Zhejiang People’s Press, 2013, pp. 28-29.
- [7] S. Lohr, “The age of big data,” *New York Times*, 2012.
- [8] B. He, “Moocs: Retrospections and reflections,” *Jiangsu Education Research*, vol. 1, 2014.



Qian Wang was born in Shaanxi, China, on Nov. 12, 1981. She got the MA in Applied Linguistics, Wuhan University of Technology, Wuhan, China, 2007. Her major field of study: linguistics and education, cognitive linguistics. She teaches English Northwestern Polytechnical University in Xi’an, Shaanxi, China. She is the assistant professor in School of Foreign Languages of Northwestern Polytechnical University. She has published several articles and books.