A Conceptual Design Framework for Big Data Based Learning Analysis

Yanlin Zheng, Shigui Ruan
Faculty of Computer Science and Information Technology
Northeast Normal University
Changchun, China
e-mail: yanlinzheng@nenu.edu.cn; ruansg639@nenu.edu.cn

Luyi Li
Faculty of Education
Northeast Normal University
Changchun, China
e-mail: lily857@nenu.edu.cn

Abstract— **Beyond clipped or isolated data, big data highlights the integrated and all-around mining of the data related to some subject, which provides an innovative perspective for learning assessment, and big data based learning analytics explores a new approach to learning evaluation and improvement. Key points in terms of big data based process typically involve where to capture the big data, what data to use and how to handle the collected data. This paper proposes a conceptual design framework for big data supported learning assessment. A three-dimensional data resources model, a learning data cube, and a four-stage process model are separately suggested to respond to the mentioned *Where*, *What*, and *How* questions.**

Keywords-big data; learning assessment; learning analytics; design framework; data cube

I. INTRODUCTION

Qualified learning assessment is indispensible to instruction adjustment and learning improvement. There are many kinds of learning assessment according to different classification methods, like non-standardized, context-sensitive and non-comparable learning assessments based on different assessment tools [1], like formative or summative learning assessment based on distinct assessment directions [2], like teacher assessment, peer assessment and self-assessment based on different assessment subjects [3]. Nevertheless, the final quality of learning assessment is not depended on the diversities of evaluations, but is decided by the effectiveness and accuracy of the assessment method itself.

For example, formative assessment is often highlighted as an important approach to modern learning and teaching enhancement. However, to achieve effective formative assessment results, it is necessary to give attention to many issues, like domain considerations in its conceptualization, measurement principles, teacher-support demands and the impact of the larger educational system etc. [4]. In this context, the promotion of effective learning assessment deserves more exploration in its assessment range (such as what aspects of learning to be evaluated, who can participate in the assessment), scale (such as how much assessment information can be taken into account) and accuracy (such as whether the assessment results is valid and reliable).

In recent years, big data, not only referred to the volume of data quantity, but also related to the variety, velocity, variability, veracity and complexity of data quality, has become a popular keyword either in domain researches or the business circles [5]. Big data is not limited to a technical issue, but also provides a bran-new perspective for researchers and practitioners to analyze and deal with complex problems. Big data works based on information collection, storage, and analysis on a massive scale beyond the process of clipped and limited data, and can correspondingly mediate a deep insight of the relationships among multiple things or subjects [6]. There have been some discussions on the significance [7], or concrete application in a specific context of big data supported learning analytics [8][9]. However, big data or learning analytics is a new topic for educators, especially in the practice level. The key questions to be answered in terms of big data based strategy are typically divided into three aspects: where the big data is from, what data is captured and how the big data is dealt with and used. This study proposes a conceptual design framework for big data supported learning assessment as follows.

II. WHERE: THREE DIMENSIONS OF DATA RESOURCES

Various data resources are indispensable to the collection of big data. In an ideal design framework, any big data based application may expect to capture as much data as it can. However, in real practice, it is advisable to identify the resource types of data that are necessary for data process. As shown in Fig. 1, this paper suggests a three-dimensional model to describe the data resources for big data based learning analysis, including context data, process data and assessment data.
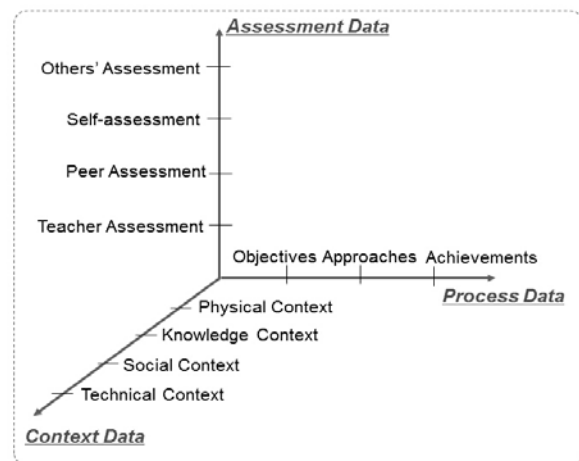

Figure 1 Three Dimensions of Data Resources

A. Context Data

Generally, in computing applications, context is often referred to "any information that can be used to characterize the situation of entities (i.e. whether a person, place or object) that are considered relevant to the interaction between a user and an application, including the user and the application themselves"[10]. This research discusses context data on a narrower sense, and defines context as background information for students' learning, not including the specific learning process information and related assessment information. In a technology-enhanced learning environment, learning context is divided into three categories: social context, knowledge context and technical context [11]. Learning situations may be embedded in a physical, or a technical, or a blended context. Therefore, this research proposes another type of context data, physical context. Physical context refers to the environmental context from real world, like places, temperatures, real tools etc.

B. Process Data

Process data is produced in a specific learning process, which is directly related to a specific learning objective, chosen learning approaches, and corresponding learning achievements. However, generally, it is hard to track process data in learning systems. First, it is necessary but difficult to identify which actions of the users must be monitored. Second, it is not easy to assure when the targeted action ends. Third, a single user may involve in several actions at the same time. Fourth, even though the actions can be tracked, it is still difficult to validate what characteristics should be extracted. In this context, it is necessary first to get some general principles for the representation and identification of learners' learning activities.

C. Assessment Data

Assessment data is specifically collected from assessment subjects. Big data based applications provide a new chance for educators to collect various assessment data, which may come from students themselves, teachers, peers, parents or others who care students' learning.

III. WHAT: CONSTRUCTION OF LEARNING DATA CUBE

Data cube provides an innovative perspective and a new approach to multiple-dimensional data representation and mining. Even though the data resource types have been marked clearly, it is of importance to have a concrete representation of a specific data.

Data ID is the identification of each data involved in big data based application. In above discussion, there are three types of data resource, and in each dimension, there are many actual and detailed data. Each detailed type of data is arranged with a unique ID. For example, in the context data dimension, taking social context as an example, there may be a data to be used to describe the interaction times between student A and Student B. This data should be given an ID code.

Data value is the actual value assigned to a specific data on a specific Time point. For example, the interaction times between Student A and Student B in May, 2015 is 10 times.

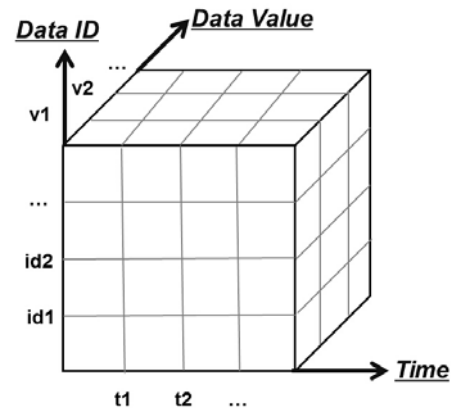May, 2015 is the specific time, and 10 is the actual data value.

Figure 2 Three Dimensions of Data Resources

IV. HOW: A FOUR-STAGE BIG DATA BASED LEARNING ANALYSIS PROCESS

As illustrated in Fig. 3, from data collection to data visualization, there are four basic stages for big data based learning analysis process.
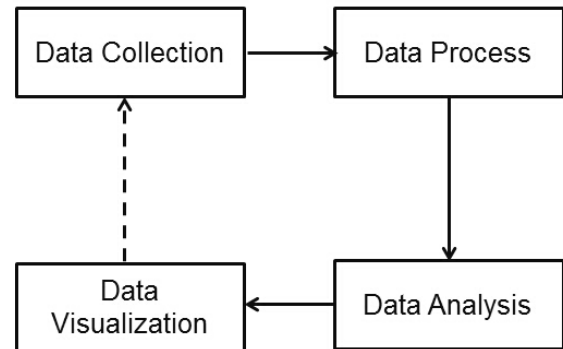

Figure 3 Four Stages of Big Data based Learning Analysis

A. The Collection of Big Data

As discussed above, the capture of big data from multiple contexts and dimensions is significant for the final application. Compared with small scale data, big data is not only involves the big volume characteristic, but also means that data is available on novel types of variables, and comes with less structure [12]. It is really a big challenge for educators to collect the data from different contexts, different devices and heteroideous systems.

B. The Process of Big Data

Big data process is often related to data cleansing, data filtering, and data clustering.

- Data clearance aims at removing unusable data from the whole data base or correcting inaccurate data for

use. This step is of use for improving data accuracy and reducing application system's data load.

- The meaning of Data filtering is twofold. Firstly, the source for filtering is the activity contexts of learners. Secondly, the goal of filtering is to get hold of the context data, process data and assessment data. There is no a uniform criterion or a set of fixed rules for data filtering because of the complications of activity contexts in different media spaces. On the one hand, there are different activities in different media spaces. On the other hand, even the same activity perhaps indicates different implications in different media spaces. Multiple algorithmic techniques may be applied for data filtering. For example, to mine learners' knowledge interests, content-based filtering strategies might be used. Collaborative filtering and Social Network Analysis(SNA) techniques might be adopted for exploring the social context in a learning space.

- Data clustering acts as a part of the process of big data, which is indispensible to distinguish learners' preferences and individualize their learning.

C. The Analysis of Big Data

Strictly speaking, data analysis covers all the process in terms of data collection, process and visualization. In this study, to emphasize the importance of educational goals, the stage of data analysis is especially marked as an independent stage, which is based on educational goals analysis. Around the educational goals, the analysis of big data may include the analysis of learners' behaviors, learners' academic achievements, instructional interactions and learning contents, etc.

D. The Visulization of Data Analysis Results

Data visualization regarded as a useful and powerful way to represent and communicate data promises to conquer the problem of "information overload", to lessen the burden on viewers trying to make sense of complex data, and to transfer cognitive load to perception task[13]. A social visualization is "a visual (or sonic or other perceptual) representation of information from which the presence, activities, and other characteristics of members of a social collectivity may be inferred, and, by extension, can provide the basis for making inferences about the activities and characteristics of the group as a whole" [14].Data may be visualized by two-dimensional static graphs or three-dimensional dynamical displays. With the development of big data applications, data visualization will not merely represent the recorded data, but also demonstrate the changing process of constantly generated data.

It is worthy to note that each stage discussed here does not exist in isolation, but rather they are interwoven and interact with each other. The main purpose of making such a distinction in terms of data process is to enhance the operability of the big data based applications design.

.

V. DISCUSSION

Firstly, it is of great significance to recognize the value of big data based learning assessment, which is a new approach to modern education evaluation [15]. Educational evaluation is one of the most important prerequisites for education policy-making and instructional decisions, while effective educational evaluation depends on overall and reliable proofs [16]. Based on deep mining and scientific analysis of multi-dimensional, large amounts of data in search of hidden relationships and value, big data is promising to help educators transform educational evaluation from small samples or incomplete information based suppositions into full data based evidences.

Secondly, although big data based applications are always built on technical tools and software, it does not mean that big data based applications does not need human design and participation. Any practice on big data application for educational evaluation should be mediated by deliberate design, choosing appropriate big data processing technologies, improving educators' data literacy, and providing multiple supports.

Thirdly, in recent years, with the development of the integration of information technology and education, many new learning and teaching models have been created, like flipped classroom, MOOC (Massive Open Online Courses), micro-lecture based individualized learning and teaching. It requires educators to consider more contexts and to pay attention to the collection, process and analysis of more data from these contexts. For example, although there are a variety of controversial discussions on MOOC, the effectiveness of MOOC teaching is crucial to the long-term development and deep application of MOOC. It is obviously unfeasible for educators and educational institutions to copy the effective teaching experiences and standards from traditional classroom contexts for MOOC teaching because of the essential characteristics of MOOC itself. To promote bid data based MOOC learning analysis, it is necessary to explore the principles of effective MOOC learning and teaching.

Fourthly, one particular focus should be given to teachers' needs analysis of using big data for teaching quality improvement. For example, according to a questionnaire based investigation, a research explored some elementary and middle school teachers' demands for the application of big data based learning analytics and assessment (BDLAA) from the following three aspects: (1) in the cognition dimension, how do the teachers judge the value of BDLAA; (2) in the usefulness dimension, what functionalities do the teachers expect for BDLAA; and (3) in the ease of use dimension, what influencing factors, preparations, and supports are concerned with teachers' adoption of BDLAA. The survey results showed that, a majority of teachers expect BDLAA can help get and analyze all the students' learning information from all-around aspects and whole-processes to realize a systematic assessment of learners' learning processes and achievements. Meanwhile, most teachers believe BDLAA is promising to promote all the learners' whole-person and personalized

1478

development, and to contribute to teachers' educational decisions and teaching improvement, and expect to be provided with more training and guide for the implementation of BDLAA.[17]

Finally, teacher data literacy, namely teacher should be able to use data to inform teaching, which has been becoming an urgent need of teachers' professional development with big data applications. Teacher data literacy, referred to teacher's awareness and ability to "continuously, effectively, and ethically access, interpret, act on, and communicate multiple types of data to improve outcomes for students in a manner appropriate to teacher's professional roles and responsibilities"[19], which plays a great role in using data to enhance learning and teaching for teachers. On the one hand, the development and implementation of big data in education enables teachers to access easily and quickly much more data than before, which will turn out to be an obvious expansion in the use of data in education domain. That is, the use of data is not only just for accountability and educational policy making, but also and what is more important, for the enhancement of learning and teaching in teaching practice level, which requires teachers' capacity to use data to inform teaching [20]. On the other hand, big data applications in education will result in an increasingly closer relationship and even one-to-one correspondence between data and actual teaching. One teacher who wants to get a more comprehensive understanding of learning and teaching and optimize his teaching continuously should be data literate to interpret and process data closely related teaching. In addition, the big impact on education from big data implementation is not just in the big "volume, variety, velocity, variability" value of big data itself, but also a big issue of student privacy, which demands teachers to be data literate to use data ethically and responsibly [21].

VI. CONCLUSIONS

Big data is potential for the deep analysis of learning processes and achievements, which is helpful for the improvement for learning performance and teaching quality. this paper proposed a conceptual design framework for big data based learning assessment and analysis based on a theoretical discussion, including three types of data resources (context data, process data and assessment data), a three-dimensional data cube (data ID, data value and time property) to represent data, and a four-stage big data based learning analysis process (data collection, data process, data analysis and data visualization). More discussions and exploration is worthy of attention to the practice and promotion of big data based learning and teaching assessment and analysis practices.

REFERENCES

[1] D.H., Kamens & A. Benavot, "National, Regional and International Learning Assessments: Trends among Developing Countries, 1960-2009," Globalisation, Societies and Education, vol. 9. 2011, pp.285-300.

[2] Crisp, G.T. "Integrative Assessment: Reframing Assessment Practice for Current and Future Learning,"Assessment & Evaluation in Higher Education, vol.37,2012, pp.33-43.

[3] Taylor, S.N., "Student Self-Assessment and Multisource Feedback Assessment: Exploring Benefits, Limitations, and Remedies,"Journal of Management Education, vol.38, Jun 2014,pp.359-383.

[4] Bennett, R.E., "Formative Assessment: A Critical Review,"Assessment in Education: Principles, Policy & Practice, vol.18, Feb 2011,pp.5-25.

[5] Wikipedia.Big data[ED/OL], http://en.wikipedia.org/wiki/Big_data,acessessed on 1st,May 2015.

[6] Viktor, M.S. & Kenneth, C. "Big Data: A Revolution That Will Transform How We Live, Work, and Think," Boston, MA:Houghton Mifflin Harcourt Publishing, 2013.

[7] Data Quality Campaign(2015).ESEA Reauthorization: Why Data Matter[EB/OL].http://www2.dataqualitycampaign.org/files/ESEA%20Reauthorization_DQC%20Recommendations.pdf/,2015-02-06.

[8] Ronald S.T.(2011). My Nine 'Truths' of Data Analysis[EB/OL].http://www.edweek.org/ew/articles/2011/06/15/35thomas.h30.html?tkn=UNZFK1u+dOgIrdSXpbrf/tuim4YZchGy/0Iw/, 2015-01-05.

[9] Ochoa, Xavier. 1st International Conference on Learning Analytics and Knowledge 2011〔EB/OL〕https://tekri.athabascau.ca/analytics/about,2015-02-06.

[10] Dey, A.K., Abowd, G.D., & Salber, D.(2001). A Conceptual framework and a Toolkit for Supporting the Rapid Prototyping of Context-aware Applications. Human-Computer Interaction, 16, 229-241.

[11] Zheng, Y.L.,& Yoneo Yano (2007). A framework of context-awareness support for peer recommendation in the e-learning context, British Journal of Educational Technology, Vol. 38, No. 2, pp.197-210. 2007.

[12] Einav,L. & Levin, J.D. (2013). The Data Reolution and Economic Analysis[ED/OL]. http://www.nber.org/papers/w19035

[13] Larkin, J.H. & Simon, H.A. (1987). Why a diagram is (sometimes) worth ten thousand words. Cognitive Science, 11(1): 65-99.

[14] Erickson, T (2003). Designing Visualizations of Social Activity: Six Claims.

[15] Data Quality Campaign(2014).Paving the Path to Success: Data Action 2014[EB/OL]. www.dataqualitycampaign.org/files/DataForAction2014.pdf/,2015-02-05.

[16] National Center for Education Statistics. Common Education Data Standards[EB/OL]. http://nces.ed.gov/programs/ceds/,2015-02-06.

[17] Zheng Yanlin，Li Luyi. "Needs Investigation for Big Data based Learning Analytics and Assessment--via the Perspective of Teachers," Modern Distance Education, Feb 2015,pp.36-42+47.

[18] Data Quality Campaign(2014).Teacher Data Literacy: It's About Time[EB/OL].http://www.dataqualitycampaign.org/files/DQC-Data%20Literacy%20Brief.pdf,2015-01-21.

[19] Schildkamp K, Karbautzki L, Vanhoof J. "Exploring data use practices around Europe: Identifying enablers and barriers," Studies in educational evaluation， vol 42, Sep 2014， pp.15-24.

[20] Mandinach E B, Parton B M, Gummer E S,& Anderson R.Ethical and Appropriate Data Use Requires Data Literacy[EB/OL]. http://datafordecisions.wested.org/wp-content/uploads/2015/03/Kappan-Ethical-and-Appropriate-Data-Use-Requires-Data-Literacy.pdf, 2015-04-13..