A Visualization Analysis of Research Status of Learning Analytics during Past Five Years in China

ZHANG Jiahua[1, a] , ZOU Qin[1,b]

[1] College of Teacher Education Zhejiang Normal University, Jinhua 321004, China;

[a]zjnuzjh@zjnu.cn, [b]etzoey@163.com

Keywords: learning analytics; word frequency analysis; co-word clustered analysis; correlative analysis; visualization analysis

Abstract. In order to understand the current situation of learning analytics, it tackled domestic research of learning analytics in China during past five years from the perspective of visual research. Multiple methods including literature analysis, word frequency analysis, and co-word clustered analysis are used in this research. It analyzed the research status from different aspects, such as process model, technological tool, typical applications, &etc. Finally it summed up the limitations of current research and put forward the opportunities and challenges facing the development of learning analytics in the future.

1. Introduction

Learning Analytics is a new technology in the field of educational technology, which is from the website analysis and business intelligence. Education researchers gradually pay close attention to it since the 2011 horizon report released. American Association of university educational information (EDUCAUSE) presents its definition of that "learning analytics uses data and models to predict student achievement and behavior and intervention"[1]. G.Siemens, an American scholar, thinks that learning analytics refers to the use of the data generated by the learners and the analytical model to explore the information and social links, in order to predict and recommend learning[2]. In 2011 the horizon report, learning analytics has been interpreted as "to assess academic achievement, predict future performance, find potential problems for the purpose, to collect large amounts of data to explain the process of learning"[3]. Learning analytics uses a variety of techniques and tools to capture the learning behavior of the data left behind. In order to optimize the teaching and learning, it studies the behavior of the learners in the learning process, predicts the academic risk, and provides the appropriate intervention and objective teaching feedback.

2. Research Methods

This paper mainly uses the word frequency analysis, Co word clustering analysis and correlation analysis. Word frequency analysis method is a method of literature metrology. The method of CO word clustering analysis is a research method of clustering algorithm and based on the frequency of the word. Correlation analysis is to study whether there is a certain dependence between the phenomenon.

3. Process Analysis

3.1 Sample Selection

According to the CNKI database of China and excellent dissertations as data source. Learning analytics as the keywords to search, author found that the domestic focuses on the analytics of the past five years. Therefore, the main articles of 148 are selected from January 2011 to December 2015. The meeting notices and other unrelated contents are removed and 144 valid documents are remaining.

3.2 Word Frequency Analytics

Author uses the Item Co-Occurrence Matrix Builder Bibliographic (BICOMB) to deal with keywords. Effective statistical literature keywords frequency, extraction of top 20 keywords, the results are shown in Table.1.

Table 1 High frequency keywords for "learning analytics"

Number	Keywords	Frequency
1	Learning Analytics	144
2	Big Data	31
3	Data Mining	26
4	Educational Data Mining	11
5	MOOCS	11
6	Online Learning	10
7	Social Network Analysis	9
8	Learning behavior	7
9	Personalized Learning	7
10	E-schoolbag	5
11	Learning Process	4
12	Online Education	4
13	Wisdom Education	4
14	Wisdom Learning	4
15	Model	4
16	Educational Informatization	4
17	Educational Big Data	4
18	Instructional Design	4
19	Big Data Era	4
20	Flipped Classroom	4

Table.1 High frequency keywords and Table.2 High frequency literature source (Right) for learning analytics. The results can be seen, the most closely associated with learning analytics of the key words are Big Data and Data Mining. This is also a hot area of current research. The sample is used again and selected 20 high frequency source journals. The results are shown in Table.2.

Table 2 High frequency literature source for "learning analytics"

Number	Keywords	Frequency
1	Journal of Distance Education	15
2	China Educational Technology	11
3	Open Education Research	8
4	The Chinese Journal of ICT in Education	8
5	Modern Educational Technology	7
6	Distance Education in China	6
7	Software Guide	5
8	Modern Distance Education	5
9	Northeast Normal University	4
10	Modern Distance Education Research	4
11	e-Education Research	4
12	Primary and Middle School Educational Technology	4
13	Journal of Guangzhou Open University	3
14	China Educational Technology & Equipment	2
15	Shanghai Normal University	2
16	Journal of World Education	2
17	Information Technology &Standardization	2
18	Journal of East China Normal University	2
19	Computer Education	2
20	Shanghai Research on Education	2

3.3 Construct the Co Word Matrix, Similarity Matrix and Tag Cloud

Co Word Matrix is a correlation matrix, and the data on the diagonal is represented by the frequency of a keyword. As the keyword of learning analysis appears a total of 144 times. After extracting high-frequency keywords, the threshold value is greater than or equal to 3 and less than or equal to 144. The formation is a 30*30 of the co word matrix (this study selects 10*10 matrix). Row and column respectively represent the keywords. When two words appear in the same time, it indicates that there is a relationship between them. Co word matrix values on the diagonal represent the keywords frequency[4]. As shown in Table.3. the hostest words include Big Data, Data Mining, Education Data Mining, MOOCS, Social Network Analysis, Online Learning.

Table 3 Co-Occurrence Matrix of High Frequency Words of Learning Analytics(part)

Co current entry	Learning Analytics	Big Data	Data Mining	Educational Data Mining	MOOCS	Online Learning	Social Network Analytics	Learning Behavior	Personalized Learning	e-Schoolbag
Learning Analytics	144	83	68	34	30	19	20	15	19	15
Big Data	83	31	14	10	6	1	5	2	5	4
Data Mining	68	14	26	4	6	3	5	4	3	1
Educational Data Mining	34	10	4	11	3	0	1	0	3	1
MOOCS	30	6	6	3	13	2	2	2	2	2
Online Learning	19	1	3	0	2	10	2	1	1	1
Social Network Analytics	20	5	5	1	2	2	9	1	0	2
Learning Behavior	15	2	4	0	2	1	1	7	1	2
Personalized Learning	19	5	3	3	2	1	0	1	7	2
e-Schoolbag	15	4	1	1	2	1	2	2	2	5

The data of Co Word Matrix is further analyzed by multivariate statistical analysis, which can be used to dig out the hidden knowledge information. With the aid of the Ochiia correlation coefficient algorithm, the Co Word Matrix is transformed to the similar matrix, and the data structure of the matrix is changed[5]. After statistical analysis, we can get the similar matrix as shown in Table.4.

Table 4 Similarity Matrix of High Frequency Keywords in Learning Analytics (part)

Co current entry	Learning Analytics	Big Data	Data Mining	Educational Data Mining	MOOCS	Online Learning	Social Network Analysis	Learning Behavior	Personalized Learning	e-Schoolbag
Learning Analytics	1.000	0.954	0.944	0.913	0.877	0.736	0.868	0.842	0.875	0.847
Big Data	0.954	1.000	0.956	0.967	0.913	0.798	0.902	0.895	0.932	0.933
Data Mining	0.944	0.956	1.000	0.918	0.919	0.836	0.912	0.909	0.901	0.881
Educational Data Mining	0.913	0.967	0.918	1.000	0.897	0.764	0.842	0.928	0.930	0.896
MOOCS	0.877	0.913	0.919	0.897	1.000	0.809	0.856	0.893	0.872	0.891
Online Learning	0.736	0.798	0.836	0.764	0.809	1.000	0.816	0.753	0.784	0.810
Social Network Analysis	0.868	0.902	0.912	0.842	0.856	0.816	1.000	0.799	0.805	0.886
Learning Behavior	0.842	0.895	0.909	0.928	0.893	0.753	0.799	1.000	0.835	0.822
Personalized Learning	0.875	0.932	0.901	0.930	0.872	0.784	0.805	0.835	1.000	0.912
e-Schoolbag	0.847	0.933	0.881	0.896	0.891	0.810	0.886	0.822	0.912	1.000

Note: the similarity in the range of 0-1. The similarity is closer to 1, on the contrary, close to 0.

The data in similar matrix is called similar data, and its size indicates the degree of the distance between the two keywords. From the Table.4, Big Data, Educational Data Mining, Social Network Analysis and other keywords have high similarity with learning analytics. Big Data(0.954),Data Mining(0.944), Educational Data Mining(0.913), MOOCS(0.877), Personalized Learning(0.736), Social Network Analysis(0.868), e-Schoolbag(0.847), Learning Behavior(0.842), Online Learning (0.736). In the published literature, the combination of the above areas are more closely to learning analytics.

The co-occurrence matrix generated by BICOMB2 is introduced into UCINET. It is saved to. ##h file format, and then opened in the NETDRAW. Then it generates high frequency keywords of the tag cloud, as shown in Fig.1. From Fig.1, we can see that the learning analytics as the center and its relationship with the closer the keywords is closer to the center, such as Big Data, Educational Data Mining, Social Network Analytics, etc. This shows the current research focuses on the learning analytics.

Fig.1 Tag Cloud

4.   The Conclusions

4.1 Theoretical Research and Application on Learning Analytics

Many scholars have conducted in-depth analysis and research on the process and model of learning analytics. Some scholars analyze and interpret the process and the steps of learning analytics[6][7]. The researchers put forward the behavior model of learning analytics, which complements the diversification of learning sources and the continuing problems of learning activities[8][9]. There are many methods used in learning analytics, such as statistical analysis and visualization, clustering, prediction, relational mining, content analysis and so on[10]. With the continuous deepening of the theory of learning analytics, the research on its practical application is also increasing. Based on some learning platforms, the researcher analyze the process of teaching and learning with the aid of the analysis methods and tools, and present the results[11][12][13].

4.2 The limitations of the study and analysis

In recent years, although there have been a number of papers and academic monographs that can be used for reference. But there are still a number of research issues to be further explored and discussed. There are some limitations in the study.

(1)Learning analytics develops slowly by the traditional teaching mode. The application of the basic education is weak.(2)The data collected in the learning analysis is not complete. Data analysis techniques need to be reduced.(3)Most of learning analysts are teachers and researchers. Self-analysis of learners is seldom.(4)The research depth and breadth of learning analytics is not enough. The corresponding prediction and intervention mechanism need to be further studied.

4.3 Opportunities and Challenges

Siemens believes that there are two important points in the prospects of learning analytics. One is the real world data collection that used to explain the learning situation and the environment. Another is the massive data provided by mobile learning[14]. Zhu Zhiting points out that the educational application of big data technology to promote the development of the learning analytics[15]. The research paradigm of big data is used for reference to discuss its application. Future challenges include three aspects as follow.

(1) Data processing technology challenges. Such as the diversity of data sources. The diversity of learning platforms is accompanied by a variety of data formats. The methods and techniques of data storage, acquisition, preprocessing and analysis are very high. At the same time, the data interoperability, compatibility requirements are relatively high. This will bring technical problems to the learning analytics.

(2) The challenge of policy and ethics, that is, the problem of the moderate openness of data. What do educational institutions do with learning analytics? Whether to carry out a certain strategic planning, and give the corresponding policy support? This is an important influencing factor for the sustainable development of the learning analytics. How to make the data collection and analysis in the premise of guaranteeing the right to know and the privacy of students and teachers, especially the students, has become another challenge in this research field.

(3) The universality of the learning analytics model and the limitation of the scope of application. At present, the study of learning analytics has not yet formed a good and general analysis model.

Some application studies have been done only on the shallow data mining, which has not been widely used in practice. In addition, most of the studies are collected from online learning activities of the data. There are a lot of real environment data does not enter the scope of the collection. Therefore, the scope of application of learning analytics is limited. All can't rely too much on it.

5. Acknowledgment

6. References

[1]. Brown,M. Learning analytics :the coming third wave[DB/OL].[2011-4-1]. http://net.educause.edu/ir/library/ pdf/ELIB1101.pdf.

[2]. Song Ling-qing. Learning analysis: a better understanding of the process of students' individualized learning——interview expert Professor Siemens George [J]. China Educational Technology,2013,(9):1-4.(In Chinese)

[3]. The Horizon Report 2011 Higher Education Edition[DB/OL]. [2011-06-25]. http://wp.nmc.org/horizon2011/.

[4]. Zhang Jie, Wang Hong. The analysis of learning research based on word frequency analysis and comparison of mobile visual word network map[J]. Modern Distance Education, 2014,(2):76-83. (In Chinese)

[5]. Hu Shui-xing. An empirical analysis on the educational application of big data and its key tech-nologies [J]. Journal of Distance Education，2015(5):46-53. (In Chinese)

[6]. Li Feng-qing, Qian Wan-zheng. Learning analytics: a new field of research and practice of uni-versity teaching information [J]. Modern Educational Technology, 2012,(7):5-10. (In Chinese)

[7]. Li Yan-yan, Ma Shao-qian, Huang Rong-huai. Learning analytics technology: service learning process design and optimization [J]. Open Education Research, 2012,(5):18-24. (In Chinese)

[8]. Yu Xiao-hua,Gu Xiao-qing. Learning activity flow: a behavioral model of learning analytics[J]. Journal of Distance Education, 2013,(4):20-28. (In Chinese)

[9]. Hu Yi-ling, Gu Xiao-qing. Educational efficiency: from the perspective of learning analytics technology [J]. Modern Distance Education Research,2014,(6):41-48. (In Chinese)

[10]. Gu Xiao-qing,Hu Yi-ling,Cai Hui-ying. The localization of MOOCs and its Countermeasures [J]. Journal of Distance Education，2013,(5):3-9. (In Chinese)

[11]. Wang Quan-wang, Zhao Bing-chuan. Research on the application of data mining technology in Moodle course management system[J].China Educational Technology,2011,(11):69-74. (In Chinese)

[12]. Gu xiao. The application and practice of learning analytics technology in Senior High School [D]. Shanghai Normal University, 2012:30-39. (In Chinese)

[13]. Chen Chun-yan. Study on the application of learning analytics in Moodle online course [D]. Shanghai Normal University, 2014:43-58. (In Chinese)

[14]. Siemens, G. Learning and Knowledge Analytics-Knewton-the future of education [EB/OL].[2011-04-14]. http://www.learninganalytics.net/p=126. (In Chinese)

[15]. Zhu Zhi-ting, Shen De-hai. The new paradigm of educational technology research based on big data [J]. China Educational Technology, 2013,(10):9-13. (In Chinese)