Comparison Research of Hotspots and Trends of Learning Analytics from 2011-2021: A Visualization Analysis Based on CiteSpace

Yimin Chen¹, Yuxuan Liu²*

¹School of Education Science, Nanjing Normal University, Nanjing, China
²College of Arts and Sciences, Case Western Reserve University, Cleveland, USA
Email: 210621040@njnu.edu.cn, *yxl2460@case.edu

Abstract

With the development of big data and social informatization, learning analytics became a popular topic in education and received more and more attention from scholars. In order to explore the similarities and differences of learning analytics between home and abroad, CiteSpace was used to compare and analyze 178 articles from the CNKI database and 1056 articles from the WOS database from 2011 to 2021. Hotspots, trends, frontiers of domestic and foreign research in learning analytics were visually analyzed. The findings show that: 1) The number of international publications was higher than that of Chinese publications from 2011 to 2021; 2) While more teacher education universities focus on the field of learning analytics in China, comprehensive universities pay more attention to the topic from the result of international publications. 3) Online learning is a key research area for learning analytics research, both in China and in other countries. 4) The research frontier of learning analytics in China mainly focused on learning prediction, while the research frontier in the international articles mainly focused on “educational data mining”, “big data”, and “the design of learning environments and tools”. The analysis captures the hotspots, trends, frontiers in the field of learning analytics and provides a reference for further research by scholars at home and abroad.

Keywords

Learning Analysis, Visualization Analysis, CiteSpace, Hotspots, Trend

1. Introduction

Since the Horizon Report in 2011 came out, learning analytics has become a new hot topic for research and application in education technology (He, 2016). More and more scholars from other countries are also gradually focusing on learning
analytics. It was early defined as the use of data and models to predict student gains and behaviors that had the ability to process this information (Siemens & Long, 2011). A team led by Gu defined learning analytics as “a tool for extracting implicit, unknown and potentially application-worthy information or patterns from the vast amount of data in the field of education, as well as a decision-making tool” (Hu et al., 2014).

With the application of the Internet in various fields, online learning began to develop as a new supplement to learning methods, and the data left behind by learners using online learning platforms to learn made it easier to analyze their learning, having an in-depth exploration of the learning process from the behavioral data left by the learners (O’Halloran, 2011). Building analytical models and displaying and interpreting data helped teachers and educational administrators to do their jobs better (Gu et al., 2012). On one hand, big data mindsets and technological innovations also presented opportunities and challenges for learning analytics. Learning analytics can be said to be a product of further development and integration of web analytics, academic analytics, educational data mining, behavioral analytics, etc. (Elias, 2011), acquiring data and forming an educational database were the foundation of educational data decision research (Gu, 2010). Learning analytics subsequently became a new wave in education informatization (Wu et al., 2013). On the other hand, some scholars also sorted out the leading edge and trends in learning analytics-related fields in China. For example, Wang and Yu (2015) reviewed learning analytics from the perspective of big data. He (2016) published The New Development of “Learning Analytics Technology” in China. Mei et al. (2021) studied the research path of international learning analytics and its inspiration. All of them provided some useful insights for the further development of learning analytics research. However, the general trends, similarities and differences in the development of learning analytics at home and abroad from 2011 to 2021 have not been explored. Therefore, this study focuses on comparing the similarities and differences between Chinese and international high-impact research scholars and institutions, hotspots, trends, and leading edge related to learning analytics.

Based on this, this paper aims to figure out the current status, research hotspots and trends of learning analytics through CiteSpase for future research. Through a comparative analysis of Chinese and international learning analytics from 2011 to 2021, the following questions are explored: 1) What are the similarities and differences in the authors and institutions of Chinese and international authors and institutions in learning analytics; 2) What are the similarities and differences in Chinese and international research hotspots and trends in learning analytics research; 3) What are the similarities and differences in Chinese and international research frontiers in learning analytics research.

2. Data Sources and Processing

2.1. Source of the Sample

The data samples for this study were obtained from CNKI (China National
Knowledge Infrastructure) and WOS (Web of Science), and information such as news and conference announcements contained in the search results were removed during the data collection process. CNKI is the most complete research database in China, and the language is predominantly Chinese, representing data resources from China. WOS is the largest comprehensive database of academic information resources in the world, covering most disciplines, and the language is predominantly English, representing data resources from all over the world. In the advanced search interface of CNKI, we entered “learning analysis” for title search, set the publication period to run from 2011 to 2021, and the source category as China Social Science Citation Index (CSSCI), and the search date was February 26, 2022. 178 core articles were retrieved, including authors, titles, keywords, abstracts, and citations. The search mode in WOS was set to advanced search, the search formula was as follows ((TS = (“learning analysis”)) OR TS = (“learning analytics”)) AND DT = (Article). The time span was set to 2011-2021. The language was set to English. The source category of data was set to Web of Science™ Core Collection. The search date was February 26, 2022. A total of 2600 records were retrieved. The record content was Full record and Cited Reference, selected Education Educational Research in the category, and 1056 valid results were obtained after de-duplication. Finally, 178 articles related to learning analytics from CNKI database and 1056 articles related to learning analytics from WOS database were analyzed.

2.2. Data Processing

The search results on CNKI were exported in Refworks format and converted to identifiable data suitable for CiteSpace analysis, and the search results on WOS were exported in text format and imported into the CiteSpace software developed by Dr. Chen.

In this study, the visualization analysis related to the learning analytics knowledge graph of CNKI and WOS from 2011 to 2021 was performed by CiteSpace 5.8, using visualizations present the structure, patterns, and distribution of scientific knowledge. In this process, the files were restricted to the period from 2011 to 2021; the time slice was set to 1 year; the selected literature sources were “title”, “abstract”, “author keyword (DE)” and “Keyword+ (ID)”; the threshold was set to “Top N % = 50”; the visualization options were “Cluster View-Static” and “Show Merged Network”. The knowledge maps were drawn from three aspects, including authors, research institutions and keywords, and the research themes were sorted out by combining with literature analysis.

3. Data Results and Analysis

3.1. Number of Articles Issued Per Year

The data results from the analysis of the annual volume of publications from CNKI and WOS about learning analytics from 2011 to 2022 are shown in Figure 1. From the figure, it can be seen that the number of publications on CNKI
about learning analytics showed a slight increase from 2011 to 2016, reached the peak number of publications in 2016 (35 articles), while the overall number of publications fluctuated slightly from 2016 to 2021, but the difference was not significant. The figure showed a trend of year-by-year increase in the annual number of posts for WOS from 2011 to 2020, with the number of posts maintaining a small increase from 2011 to 2016, and a larger increase from 2016 to 2020, with the annual number of posts reaching a peak (244 articles) in 2020 and a slight decrease in 2021. Overall, the annual number of articles published on WOS is more than that on CNKI, and this number gap reached its maximum in 2020 (difference of 221 articles) (see Figure 1).

3.2. Author of the Paper

To search the high-impact authors in CNKI and WOS, the node type is set as “Author” in CiteSpace (see Table 1). Among the authors with the most publications in CNKI, Zhao, W. ranked first (17), followed by Gu, X. Q. (12) and Jiang,
Q. (12). In terms of publications in WOS, Gasevic, D. ranked the first place with 37 articles on learning analytics, followed by Pardo, A. and Rienties, B., both with 21 articles on learning. In general, the number of articles published by scholars from other countries in the field of learning analytics was higher than the number of articles published by scholars in CNKI (see Table 1).

The study generated a co-occurrence map of authors of learning analytics research (see Figure 2 and Figure 3) to identify the core academic groups of learning analytics research. CiteSpace provides two metrics, module value (Q-value) and average profile value (S-value), based on the clarity of network structure and clustering, as a basis for judging the effectiveness of mapping. In general, Q-value is generally within the interval [0, 1), and if Q ≥ 0.3 indicates that the structure of the delineated societies is significant; S value ≥ 0.5 indicates reasonable clustering, and if it is above 0.7, it indicates significant clustering. 320 nodes and 475 links in the CNKI author collaboration network, with an overall network density of 0.0093, Q = 0.956 > 0.3, and S = 1, indicated that the social group of authors formed by the learning analysis research is significant, as shown in Figure 2, the learning analysis research formed three cohesive research teams with Wu, Y. H. scholars, Gu, X. Q. scholars, and Zhao, W. scholars as its core, respectively (see Figure 2). 2407 nodes and 5237 connections are in the WOS author collaboration network. The overall density of the network was 0.0018, and the Q value was 0.9535 > 0.3, and S = 1, it indicates significant clustering. In Figure 3, the learning analytics research at WOS formed a cohesive scientific team with Gasevic, D., Pardo, A., and Martinez-Maldonado, R. as its core (see Figure 3).

Figure 2. The network of authors in CNKI.
3.3. Research Institutions

In the node type of CiteSpace, select "Institution", and treat the publishing institutions of different colleges of the same school as the same publishing school, among which the high-impact research institutions in CNKI and WOS are shown in Table 2. Among the most published institutions in CNKI, East China Normal University ranked first, followed by Beijing Normal University (32 articles, ranking the second) and Northeast Normal University (25 articles, ranking the third). Among the institutions with the most papers published by WOS, Monash University ranked first, followed by The Open University and Edinburgh University. In comparison, more Chinese institutions focus on the field of learning analytics are teacher education universities, and among the many colleges that pay more attention to the field related to learning analytics are the Department of Educational Technology and the School of Information Technology, International institutions that focus more on the field of learning analytics are comprehensive universities (see Table 2).

The study generated a learning analytics research institution co-occurrence map (see Figure 4 and Figure 5) to identify the core research institutions for media literacy research. The CNKI collaborative network has 186 nodes and 193 nodes, with an overall network density of 0.0112, a Q value of 0.9027, and an S value of 1. Three core research institutions, namely, East China Normal University, Northeast Normal University, and Beijing Normal University are shown in Figure 4. The institutional cooperation network of WOS has 899 nodes and 1705 nodes. The overall density of the network is 0.0042, the Q value is 0.7894, and the S value is 0.9402. A core research collaboration group with Monash University, Open University, University of Edinburgh, University of Technology...
Sydney, and University of South Australia is formed in Figure 5. The figure shows that there are also a number of collaborations between different institutions, where the Chinese university, Central China Normal University, has formed a research collaboration with universities in other countries in the field of learning analytics.
Table 2. Top 10 institutions in CNKI and WOS (2011-2021).

| CNKI        | WOS            |
|-------------|----------------|
| Count       | Institutions   | Count           | Institutions               |
| 1           | 40 East China Normal University | 40 Monash University |
| 2           | 32 Beijing Normal University | 40 Open University |
| 3           | 25 Northeast Normal University | 28 University of Edinburgh |
| 4           | 21 Central China Normal University | 20 University of Technology Sydney |
| 5           | 10 Beijing University of Posts and Telecommunications | 20 University of South Australia |
| 6           | 7 Jiangnan University | 19 University of Sydney |
| 7           | 6 South China Normal University | 19 Kyoto University |
| 8           | 6 Jilin University | 16 Carlos III University of Madrid |
| 9           | 6 Minzu University of China | 14 Tallinn University |
| 10          | 6 Changchun College of Education | 13 University of Florida |

3.4. Analysis of Keywords

3.4.1. The Network of Keywords

The keywords with higher frequency and centrality in the analysis results are the research hotspots. Keywords are the core summary of an article, and there is some relationship between keywords in a text, the more occurrences of a word pair in the same document, the stronger the association between the two topics. CiteSpace can measure the literature of a specified field to explore the research hotspots and development trends of a discipline or field, and its keyword co-occurrence can directly reflect the research hotspots and frontier trends of a research field (Chen, 2006). The higher the frequency of keyword co-occurrences is, the higher the point centrality is, and the more important the node is in its field. Except for the basic keyword “learning analytics”, the top 10 keyword co-occurrences of learning analytics from 2011 to 2021 are shown in Table 3. The top three most frequent keywords in CNKI learning analytics research are “big data” with 20 occurrences and a centrality of 0.05, followed by “data mining” with 12 occurrences and a centrality of 0.02, and “learning behavior” with 8 occurrences and a centrality of 0.04. The top three most frequent keywords in WOS learning analytics research are: “learning process” with 93 occurrences and a centrality of 0.02, “learning outcomes” with 55 occurrences and a centrality of 0.04. “Learning outcomes” appears 55 times with a centrality of 0.05, and “learning management system” appears 50 times with a centrality of 0.02.

As shown in Table 3, the research directions of learning analytics at international publications and Chinese publications are mainly “online learning” and “MOOC”. In China, the main research objects are “big data” and “data mining”, while in other countries, the research mainly analyzes the learning process and the related learning results. As can be seen from the table, “online learning” and “MOOC” are common keywords in Chinese and international research, which reveal the research focus in the field of learning analytics in the world (see Table 3).
Table 3. Top 10 keywords in CNKI and WOS (2011-2021).

| Rank | CNKI |  |  | WOS |  |  |
|------|------|----|----|-----|----|----|
|      | Count | Centrality | Keywords | Count | Centrality | Keywords |
| 1    | 20    | 0.05 | big data  | 93    | 0.02 | learning process  |
| 2    | 12    | 0.02 | data mining | 55    | 0.05 | learning outcomes |
| 3    | 8     | 0.04 | learning behavior | 50    | 0.02 | learning management system |
| 4    | 8     | 0.02 | visualization | 46    | 0.01 | learning design |
| 5    | 8     | 0.02 | online Learning | 43    | 0.01 | massive open online courses |
| 6    | 5     | 0.02 | smart education | 43    | 0.03 | learning activity |
| 7    | 5     | 0.01 | MOOC | 41    | 0.06 | online learning |
| 8    | 4     | 0.00 | smart learning | 41    | 0.06 | student engagement |
| 9    | 4     | 0.01 | machine Learning | 39    | 0.01 | self-regulated learning |
| 10   | 4     | 0.00 | multimodal | 37    | 0.02 | learning environment |

3.4.2. Keywords Cluster

In CiteSpace, the timeline view shows the publication and peak times of articles and terms, while the clustering view provides node and linkage graphs where nodes indicate details of authors, institutions, countries, terms, keywords, cited literature, cited journals, etc (Chen et al., 2010). The keyword clustering results are shown in Figure 6 and Figure 7. The Q values are 0.5641 and 0.8413, and the S values are 0.9392 and 0.8748, respectively, indicating that the clustering results have reference value. The smaller the serial number is, the more keywords are included in the clustering results and the larger the category is. 26 categories are available in the keyword clustering results of CNKI (the first 12 clustering results are shown in the figure). It can be seen that “Big Data”, “Data Mining” and “Learning Behavior” rank the top three among the 26 categories in CNKI (see Figure 6). The WOS keyword clustering results in a total of 176 entries, with the top three being “MOOC”, “learning outcome” and “uncertainty” (see Figure 7).

3.4.3. Timeline of Keywords

The keyword timeline view (Figure 8 and Figure 9) shows the development of hotspots in Chinese and international learning analytics research from 2011 to 2021. From 2011 to 2021, the scope of learning analytics research in China has undergone a more expansive development, and the early hotspots of learning analytics are mainly focused on “big data”, “data mining”, “learning behavior”, and “decision support”. “Data mining” and “learning behavior” gradually declined from 2020 to 2021. “Decision support” has been declining in popularity since 2017, and it is obvious that the application of big data on learning was the most important research hotspot of learning analytics. “Optimization of learning” and “learning evaluation” were relatively recent research topics Overall, the scope of learning analytics research has gradually expanded and increased from 2011 to 2021 (see Figure 8). “MOOC” was the key area of research in learning
Figure 6. The network of keywords from CNKI.

Figure 7. The network of keywords from WOS.
analytics in WOS, and there was a wide range of research around MOOC and the research area of learning analytics has been expanding year by year since 2011. The range of research areas in learning analytics had expanded year by year since 2011, and the years 2011 to 2017 had been a high period of hotspots in
learning analytics, with research such as “achievement goal theory”, “evaluation facilitation strategies”, “learning analytics dashboard” and “web mining”. After 2016, research on “learning outcome”, “evaluation facilitation strategies”, and “learning analytics dashboard” began to decrease gradually (see Figure 9).

### 3.4.4. Keywords Burst Terms

The flourishing of a research frontier inevitably leads to an increase in the number of its keywords in a short period of time. Burst terms refer to words that appear or used more frequently in a short period of time, and the frontiers and trends of research fields can be judged based on the word frequency changes of burst words. **Figure 10** shows the burst mapping of learning analytics keywords in CNKI from 2011 to 2021. The early topics of interest in the field of learning analytics are “data mining (2012)”, “learning behavior (2012)”. “Data mining (2.31)”, “visualization (1.26)” and “MOOC (Massive open online courses, 1.13)” are among the topics that were highlighted with high intensity. “Learning prediction”, “empirical studies”, and “literature review” were the topics of more interest in the field of learning from 2019 to 2021, and “empirical studies”, and “literature review” were also the main research methods in the field of learning analytics in China this year (see **Figure 10**). **Figure 11** shows the learning analytics burst terms mapping in WOS from 2011 to 2021. It can be seen that “learning process”, “learning environment”, “big data”, “educational data mining”, and “learning environment” are the five hot topics in learning analytics research. Among them, “learning process” was a hot topic from 2019 to 2021, “learning environment” was a hot topic from 2017 to 2021, and “big data” was a

![Top 15 Keywords with the Strongest Citation Bursts](image)

**Figure 10.** The burst of keywords in CNKI (2011-2021).
hot topic from 2017 to 2021. “Big data” was a hot topic from 2017 to 2019, “educational data mining” was a hot topic from 2019 to 2021, and “instructional design” was a hot topic from 2019 to 2021 (see Figure 11).

The research frontier of learning analytics in China is mainly focused on learning prediction, and the research approach is mostly based on empirical studies and literature reviews and the international publications are mainly focused on educational data mining, big data, and the design of learning environments and tools. More focus on the learner’s learning process.

4. Discussion and Conclusion

This study aims to explore the development trends and frontiers of learning analytics from 2011 to 2021, using CiteSpace as a research tool to visualize and analyze data from CNKI and WOS. This paper visualizes the similarities and differences of learning analytics research on Chinese publications and international publications in a graphical way, and also provides a reference for researchers in the field of learning analytics.

4.1. Similarities and Differences between Authors and Institutions

From the analysis of the volume of author publications and publication institutions, it can be concluded that international research on learning analytics is expanding year by year, while the development of Chinese research in this area is still at a plateau. Both Chinese and international research on learning analytics
have established collaborations and have cohesive research groups. The number of articles published by normal universities in China far exceeds that of comprehensive universities, what’s more, the normal schools focus more on learning analysis in comparison. In other countries, comprehensive universities pay more attention to the topic of learning analytics. In the statistics of number of international publications, Central China Normal University, still has a notable performance.

In the future, Chinese scholars and institutions can form more collaborations with scholars and institutions from other countries in the field of learning analytics.

4.2. Similarities and Differences in Research Hotspots and Trends

From the network of keywords map and the timeline of keywords map, it can be found that the research hotspots of learning analytics in China focus on “big data”, “data mining” and “learning behavior”, while the hotspots of international learning analytics focus on the “learning process”, “learning outcomes”, and “learning management systems”. From the research results, it can be seen that Chinese research on learning analytics is mostly focused on data analysis of the learning process, while the main object of international learning analytics research is learning-related aspects, from data back to the learning behavior itself. This also provides a new research idea for the next stage of localization of learning analytics development in China.

4.3. Similarities and Differences in Research Frontiers

From the keyword Burst terms tables, we can conclude that the research frontiers of learning analytics in China are mainly focused on research on “learning prediction” and “smart education”, and the research area is dominated by online learning. In addition, it can be inferred that the research on the construction of models based on learning analytics in smart education and learning analytics in online learning is likely to be the focus of the future study. The frontiers of international learning analytics mainly focus on “learning process”, “learning environment” and “education data mining”, and education technology-related majors are more concerned with the development of learning analytics. Combining learning analytics with various factors in the learning process is also of greater significant relevance.

4.4. Limitations and Future Research

This study still has some limitations that the study data only encompasses the CNKI database and the WOS database, with a small sample base of data, which may not cover all the studies in the fields. Future studies may consider obtaining more comprehensive data from more databases. Additionally, CiteSpace can be used in combination with other literature analysis softwares in the future to provide a more comprehensive and in-depth study of relevant topics.
Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Chen, C. (2006). Citespace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature. *Journal of the American Society for Information Science and Technology, 57*, 359-377. [https://doi.org/10.1002/asi.20317](https://doi.org/10.1002/asi.20317)

Chen, C., Ibekwe-SanJuan, F., & Hou, J. (2010). The Structure and Dynamics of Cocitation Clusters: A Multiple-Perspective Cocitation Analysis. *Journal of the American Society for Information Science and Technology, 61*, 1386-1409. [https://doi.org/10.1002/asi.21309](https://doi.org/10.1002/asi.21309)

Elias, T. (2011). *Learning Analytics: The Definitions, the Processes, and the Potential*. The Landing Athabasca University. [https://landing.athabascau.ca/file/download/43713](https://landing.athabascau.ca/file/download/43713)

Gu, X. Q., Huang, J. Q., Zhu, Y. K., & Yuan, C. J. (2010). Let Data Speak: Educational Application of Data-Driven Decision Making System. *Open Education Research, 16*, 99-106.

Gu, X. Q., Zhang, J. L., & Cai, H. Y. (2012). Learning Analytics: The Emerging Data Technology. *Journal of Distance Education, 30*, 18-25.

He, K. K. (2016). The New Development of “Learning Analytics Technology” in China. *e-Education Research, No. 7*, 5-13.

Hu, Y. L., Gu, X. Q., Luo, J. T., & Chen, J. Y. (2014). An Analysis of Education-Decision Support Oriented by Educational Effectiveness: From the Perspective of Learning Analytics. *Modern Distance Education Research, No. 6*, 41-48.

Mei, P. F., Zhang, L. N., & Qi, J. (2021). International Research Paths of Learning Analytics and Their Insights. *Chinese Journal of ICT in Education, No. 6*, 7-13. (In Chinese) [https://doi.org/10.3969/j.issn.1673-8454.2021.06.003](https://doi.org/10.3969/j.issn.1673-8454.2021.06.003)

O’Halloran, K. (2011). Investigating Argumentation in Reading Groups: Combining Manual Qualitative Coding and Automated Corpus Analysis Tools. *Applied Linguistics, 32*, 172-196. [https://doi.org/10.1093/applin/amq041](https://doi.org/10.1093/applin/amq041)

Siemens, G., & Long, P. (2011). Penetrating the Fog: Analytics in Learning and Education. *TD Tecnologie Didattiche, 22*, 132-137.

Wang, L. Z., & Yu, W. H. (2015). A Review of Learning Analytics from a Big Data Perspective. *Distance Education in China, No. 3*, 31-37. (In Chinese) [https://doi.org/10.3969/j.issn.1009-458X.2015.03.004](https://doi.org/10.3969/j.issn.1009-458X.2015.03.004)

Wu, Y. H., Chen, D., Ma, X. L., Cao, P., Feng, X., & Zhu, Z. T. (2013). Learning Analytics: The New Coming Wave of Education Information. *Journal of Distance Education, 31*, 11-19.