So, H. J. et al. (Eds.) (2020). Proceedings of the 28th International Conference on Computers in Education.
Asia-Pacific Society for Computers in Education

Research trend and development process in learning analytics: a review of publications in selected journals from 2008 to 2019

Fuzheng ZHAO[a], Yoshiyuki TABATA[b], Chengjiu YIN[c*]
[a]*Graduate School of System Informatics, Kobe University, Japan*
[b]*Research Institute for Information Technology, Kyushu University, Japan*
[c]*Information Science and Technology Center, Kobe University, Japan*
*yin@lion.kobe-u.ac.jp

Abstract: We employed a Co-citation analysis by the Citespace software for the trend and development process in learning analytics. Through the clustering term time-zone view, it is clearly shown that 15 research clustering terms occurred from 2008 to 2019. Moreover, this paper proposed a clear review when these terms emerged and how they grew.

Keywords: Co-citation, Learning Analytics, Literature review

1. Introduction

With the development of technology, learning analytics (LA) have been improved (Yin & Hwang, 2018; Yin, Yamada, & Shimada, 2019). To better understand the development line of learning analytics, many researchers proposed literature reviews with different perspectives. For example, states, trends (Dawson, Gašević, & Siemens, et al., 2014), factors, nature, fundaments, applications (Peña-Ayala, 2018), limitations, methods, and key stakeholders (Leitner, Khallil, & Ebner, 2017). Although there was evidence of how many research terms occurred, it is not a clear outline of when they emerged and how they grew.

2. Method

2.1 data

14,035 records were gained on the Web of Science, by keywords: “learning analytics” or “education data mining” and timespan: 2008 to 2019. Then, from 14,035 records, selecting papers published in the top 20 publications by Google Scholar Metrics, and finally we got 496.

2.2 Analysis Method

Clustering in data mining is a process of aggregating and classifying data in complex networks based on similarities. In keyword clustering analysis, clustering reflects the similarity of nodes in a network (Chen, C. et al., 2010), which is helpful for identifying and detecting representative knowledge subgroups in a research field, i.e., hot topics in the research field. By setting the network nodes as “Keyword” in the Citespace software, and clustering them on the basis of keyword co-citation network. Finally, a keyword clustering network by time-zone is obtained, as shown in figure 1. The figure body is the historical development lines, which represent the trend of each clustered term. The label on the left is each clustering term name. The number on each line represents the order of each node.

3. Results

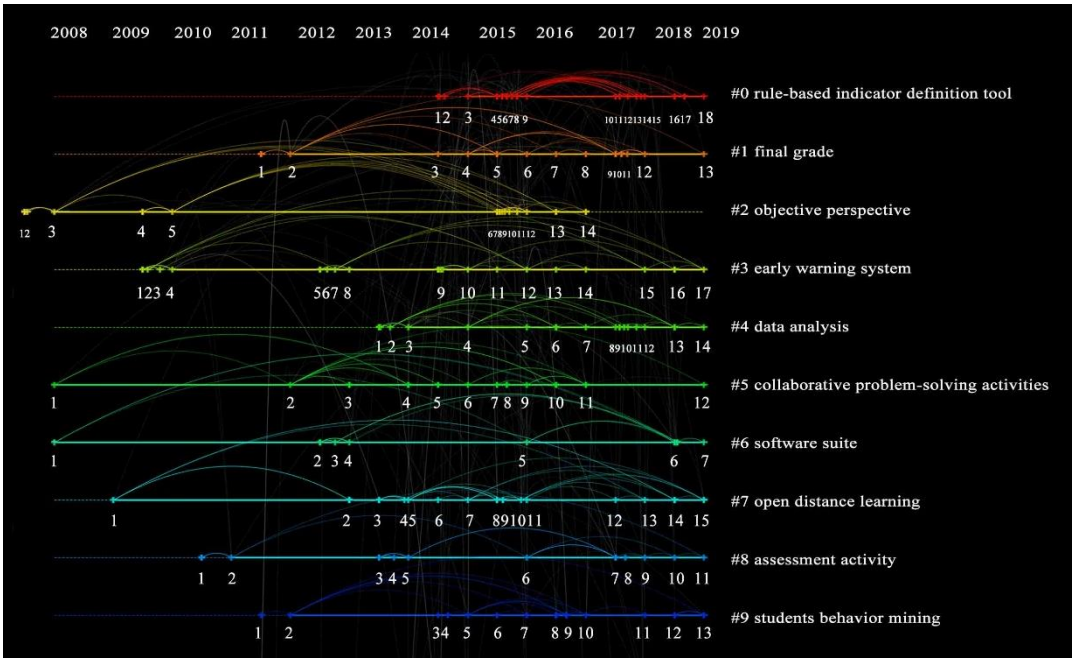

Figure 1. Keyword co-citation network clustering time-zone (Top 10).

In the past decade, the clustering term of LA includes 15 clustered terms, rule-based indicator definition tool, final grade, objective perspective, early warning system, data analysis, collaborative problem-solving activities, software suite, open distance learning, assessment activity, students' behavior mining, institutional strategical plan, managing cognitive load, emerging educational technologies, understanding social interaction, practical application. I selected the top 10 clustering terms for analysis

3.1 Rule-based indicator definition tool

"Actionable insights" merged in 2014. There have been two concentrated development stages. The first one consisted of "data collection" "learning context" and "self-regulated learning" within 2015. The second one included "multimodal data" "machine learning" and "student performance" within 2017. "Neural network" "case study" and "digital learning environment" are the latest research nodes.

3.2 Final grade

It has an even development line, starting with "combing education data mining" in 2011. "Considerable variability" "learning management system" "adult learners time management" "previous research" "learning analytics technique" "data source" "educational research" "5-year development" "open educational resources" and "google analytics" sequentially emerged from 2011 to 2016. Concentrated research has been formed by "applying learning analytics" and "undergraduate students" in 2017. "Course structure" became the latest research node in 2019.

3.3 Objective perspective

It started from 2008 to 2016. A concentrated development stage exited in 2015, including "learning behaviors" "educational technology" and "learning environment". It ended with "3D virtual laboratory" (2016) and "preliminary finding" (2016).

3.4 Early warning system

The initial point of this term concentrated with the "early warning system" "3at-risk students" and

"education institutions". After that, the second concentrated development stages followed by "earning analytics tool" "teacher inquiry" and "learning design". Since 2014, this term has begun to enter an even stage, the "knowledge gap" "educational institutions" "pedagogical model" "social network" "early identification" and "course materials" have emerged and developed in succession until 2017. From then, "blended learning" has been gained attention by 2019.

3.5 Data analysis

"Social learning" and "social network analysis" became the initial nodes in 2013. In the next three years, "massive open online course" "learning performance" "early detection" "foreign language" and "online environment" sequentially entered into the field of data analysis. The concentrated development stage occurred within 2017, focusing on "learning strategy" "online learning" "learning analytics dashboards" and "student engagement". In the past two years, it shifted to "demographic characteristic" "effective strategy" and "decision tree".

3.6 Collaborative problem-solving activities

It has been persistent from 2008 to 2019. Since the initial research node "educational data mining" merged in 2008, there has been limited attention to it until 2011. However, the next six years have seen the rapid development at an increasingly shorter interval. The development sequence is "learning analytics" (2011) "concept comprehension" (2012) "adaptive learning" (2013) "chronological framework" (2014) "student learning" (2014) "completion rates" (2015) "natural language processing" (2015) "collaborative learning" (2015) "guest editorial" (2016) "sequential analysis" (2016). "Artificial intelligence" became the recent research node in 2019.

3.7 Software suite

There were 7 research nodes in it, which presented in a way that is widely spaced over time. The first node was "educational context" in 2008. The second concentrated development stage consisted of "competence acquisition" "blended courses" and "course design" in 2012. The fifth node was "educational data". Both "students' behavior" and "empirical evidence" together constituted a research node in 2018. It followed that the predictive model became the latest research node in 2019.

3.8 Open distance learning

"Content analysis" was the starting point in 2009. After that, some research nodes continued to emerge until 2015, such as "authoring system subject" "critical reflection" "blog content" "analyzing large dataflows" "conceptual framework" "student data" "analytic approach" and "big data". There were two concentrated development stages before 2015, including "ethical considerations" "emerging field" and "learning analytics system". From 2017 to 2019, "classroom settings" "systematic review" "entire course" and "data science" sequentially emerged and formed the fourth concentrated development stage.

3.9 Assessment activity

There were 4 obvious development stages. The first one was composed of "complex chemical system" "conceptual understanding" and "complex system" in 2010. "Assessment activity" "supporting teachers" and "clustering analysis" sequentially emerged and constituted to be the second one in 2013. Compare with the former two stages, "learning experience" was the third one. From 2017 to 2019, "process mining" "course completion" "massive open online course" "available tool" and "theoretical framework" formed the fourth one.

3.10 Students behavior mining

The earliest emergence of this term was "communication technology", following by "learning activity"

in 2011. From 2014 to 2016, the increasing number has been accelerated. For example, “90th percentile” “final course grades” “learning outcome “disciplinary factor” “learning designs” “virtual learning environment” “student-facing learning analytics” and “control group”. The latest research nodes were “learning approach” “students’ engagement” and “e-learning environment” since 2017.

4. Conclusion

This study found that some terms have been persistent from 2008 to 2019, in which some nodes merged in a concentrated or separated way. For example, “collaborative problem-solving activities” and “software suite”. In addition, the majority of terms presented a shorter development line, and grown from a certain time to the present. Such as “rule-based indicator definition tool” “final grade” “early warning system” “data analysis” “open distance learning” “assessment activity” and “students behavior mining”. It is worth noting that “objective perspective” have continued to develop since they appeared, but they have not received attention in recent years.

References

Chen, C. et al. (2010). The structure and dynamics of co-citation clusters: A multiple-perspective co-citation analysis. *JASIST*, 61(7), 1386-1409.

Dawson, S., Gašević, D., Siemens, G., & Joksimovic, S. (2014, March). Current state and future trends: A citation network analysis of the learning analytics field. *In Proceedings of the fourth international conference on learning analytics and knowledge* (pp. 231-240).

Jivet, I., Scheffel, M., Specht, M., & Drachsler, H. (2018). License to evaluate: Preparing learning analytics dashboards for educational practice. *Proceedings of the 8th international conference on learning analytics & knowledge* (pp. 31–40). ACM.

Peña-Ayala, A. (2018). Learning analytics: A glance of evolution, status, and trends according to a proposed taxonomy. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(3), e1243.

Yin, C., & Hwang, G. J. (2018). Roles and strategies of learning analytics in the e-publication era. *Knowledge Management & E-Learning: An International Journal*, 10(4), 455-468.

Yin, C., Yamada, M., Oi, M., Shimada, A., Okubo, F., Kojima, K., & Ogata, H. (2019). Exploring the relationships between reading behavior patterns and learning outcomes based on log data from e-books: A human Factor Approach. *International Journal of Human-Computer Interaction*, 35(4-5), 313-322.