The Role of Big Data and Artificial Intelligence in Education and Education Research: A Literature Mapping

Si-Pan Sun[1] Han-Teng Liao[1,2] Xue Wu[1,2*]

[1] *School of Literatures and Communications, Sun Yat-sen University Nanfang College, Guangzhou, Guangdong, 510970, China*
[2] *New Media Research Centre, Sun Yat-sen University Nanfang College, Guangzhou, Guangdong, 510970, China*
Corresponding author. Email: wuxue_newmedia@outlook.com

ABSTRACT

Artificial intelligence (AI) and big data projects have been the focus of discussion to improve the learning experiences and outcomes. Based on a mapping of 980 articles, this article conducts a systematic review of educational research related to big data and AI by using VOSviewer. This article mainly examines three aspects: main sources, disciplines, and keywords. It identifies several research clusters (e.g. multidisciplinary, education technology, and information) and main research topics such as learning analytics, intelligent tutoring systems, and collaborative learning, higher education, etc. The systematic mapping of the literature contributes to the groundwork for educators, researchers, and policymakers for further research, curriculum and policy work.
Keywords: learning analytics, intelligent tutoring systems, AI, big data, education

1. INTRODUCTION

As our everyday life involves multiple digital and networked services, our learning environments and experiences are expected to undergo a digitalization and digital transformation processes, including those advanced by the latest big data and artificial intelligence (AI) technologies and applications. For example, the issues of education technology, such as personalized learning[1], are expected to be discussed with the emergence of intelligent learning systems that are designed to improve cooperation ability and learning efficiency[2,3]. For another, the development of machine learning and big data technology also brought more possibilities for medical treatment and health[4], for predictions and interventions[5]. Such AI- and data-driven processes for better learning experiences are expected to be critical for education, when the COVID-19 outbreak has affected 91% of students worldwide[6].

Related research in the education field has been conducted, with the aim to better understand the relationships between AI, data analytics, and education. Such research development includes work on educational data infrastructure[7], data institutional framework[8,9], curriculum development[10], literacy education[11], and ethical issues surrounding using machine learning[12]. However, there is currently a little systematic review of this research area. Thus, it is essential to systematically review the latest educational research results on big data and AI. The purpose of this article is to provide such a roadmap for researchers, application designers, curriculum and course designers, and policymakers to identify the latest scientific development and discussions on the subject. Such a groundwork is essential to guide big data and AI development for learning research and practices.

2. DATA AND METHODS

In order to retrieve the relevant education research from the Web of Science (WoS) database, we used the query:
- (SU="Education & Educational Research" OR WC= ("Education & Educational Research" OR "Education, Scientific Disciplines" OR "Education, Special" OR "Psychology, Educational")) AND TS = ("artificial intelligence" OR "machine learning" OR "big data")

In addition to the topics (see field tag TS above), we used the Web of Science Category (field tag WC) and research area (field tag SU) to limit the scope within education. Our search included SCI-EXPANDED, SSCI, A&HCI, ESCI, resulting in 980 articles collected on February 5, 2020. VOSviewer was used for clustering and visualizing results.

3. RESEARCH FINDINGS

As shown in Figure 1, the number of publications has largely steadily grown before 2014
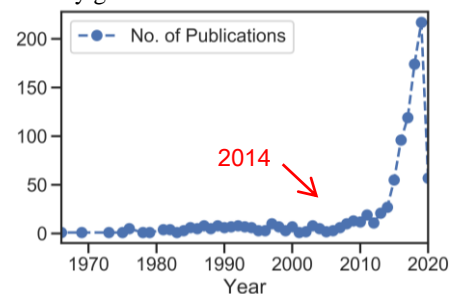

Figure 1 The number of publications

Since then, the number of publications has increased rapidly. Our datasets show that in 2019 alone, 217 papers have been published on the topic.

3.1. *Main clusters of sources and disciplines*

To explore the latest development of AI and Big Data in the field of education, the key publications, disciplines, and keywords are presented as follows.

3.1.1. *Clusters of publication sources*

Based on bibliographic coupling analysis of publication sources, we have found 7 clusters, as shown in Figure 2. Providing more detailed information on the clusters and their respective member journals Table 1 summarizes also the respective disciplines (based on WoS research areas) and their respective ranking based on the number of articles each of them has published. Before going into the details, the top three clusters #1, #2, #3, in red, green and blue, respectively, clearly take up the main part of Figure 2, showing their strong presence of the overall network.

Table 1 Detailed information of top publication sources based on bibliographic coupling relationship network

Cluster	Journals	Disciplines
#1	(1) International Journal of Emerging Technologies in Learning, (2) Computers & Education, (4) Computer Applications in Engineering Education, (5) IEEE Transactions on Learning Technologies, (6) International Journal of Engineering Education, (7) Education and Information Technologies, (8-1) IEEE Transactions on Education, (9) Interactive Learning Environments, (10) Educational Technology & Society, (12) Education Sciences, (13) Journal of Computer Assisted Learning, (14-2) International Journal of Electrical Engineering Education, (18-2) ACM Transactions on Computing Education, (19-1) International Journal of Computer-Supported Collaborative Learning, (20-1) Measurement in Physical Education and Exercise Science, (21-4) Journal of Educational Computing Research	Computer Science, Engineering, Information Science & Library Science, Social Sciences - Other Topics, Sport Sciences, Education & Educational Research*
#2	(3) British Journal of Educational Technology, (14-1) Learning Media and Technology, (17-3) International Review of Research in Open and Distributed Learning, (18-1) Technology Knowledge and Learning, (18-4) Theory and Research in Education, (18-5) Journal of E-Learning and Knowledge Society, (20-3) Tec trends, (21-5) International Journal of Educational Technology in Higher Education, (21-6) Journal of Education Policy	Education & Educational Research*
#3	(14-3) Eurasia Journal of Mathematics Science and Technology Education, (18-3) Education for Information, (20-2) Decision Sciences-Journal of Innovative Education, (21-7) Journal of Education for Business	Information Science & Library Science, Education & Educational Research*
#4	(11) Academic Medicine, (16-2) Medical Teacher, (21-1) Medical Education	Health Care Sciences & Services, Education & Educational Research*
#5	(17-5) Instructional Science, (20-4) Programmed Learning & Educational Technology	Psychology, Education & Educational Research*
#6	(17-1) Journal of Chemical Education	Chemistry, Multidisciplinary, Education & Educational Research*
#7	(19-2) International Journal of Engineering Pedagogy	Education & Educational Research*

The first and main "multidisciplinary" cluster #1 can be seen at the center of Figure 2, in red color, with top journals such as (1)*International Journal of Emerging Technologies in Learning*, (2)*Computers & Education*, etc. Cluster #1 is also between the cluster #2(in green) and cluster #3(in blue).
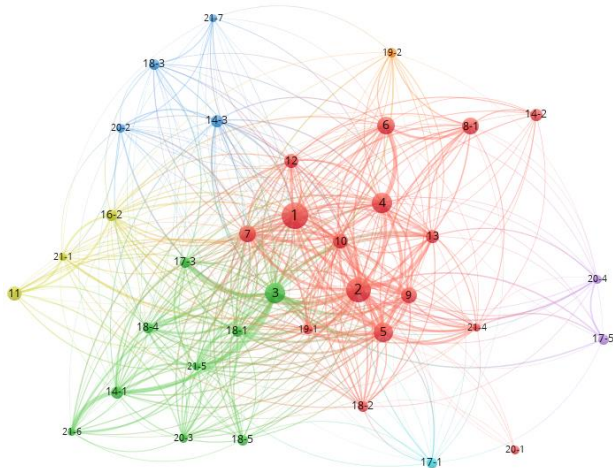
Figure 2 Visualization of top publication sources based on bibliographic coupling relationship network.

The second main "education technology" cluster #2 can be seen at the lower-left part of Figure 2, in green color, with top journals such as (3)*British Journal of Educational Technology*. In contrast to the multidisciplinary cluster #1, cluster #2 is more concentrated on education research. Note that education research appears in all clusters by design.

The third "information science" cluster #3 can be seen at the upper-left corner of Figure 2, in blue color, including innovation and business education journals such as (14-3) *Eurasia Journal of Mathematics Science and Technology Education,* (18-3)*Education for Information*, (20-2)*Decision Sciences-Journal of Innovative Education*, (21-7)*Journal of Education for Business, etc.*

The fourth "health care" cluster #4 can be seen at the left corner in yellow, including (11) *Academic Medicine*, (16-2) *Medical Teacher*, (21-1) *Medical Education*. The fifth "psychology" cluster #5 can be seen in purple on the right side of Figure 2. For the rest, specific single-item clusters include cluster #6 (chemical education) and cluster #7 (engineering pedagogy).

It is noteworthy that overall the IEEE and ACM journals take up the main space of cluster #1, whereas the education technology ones while being central, are relatively secondary in cluster #2.

3.2. The main cluster of author keywords

To analyze the emerging topics in this field, Figure3 shows the research findings based on author keywords, showing five clusters of topics. The top three clusters, in red, green and blue, respectively, take up the central space.
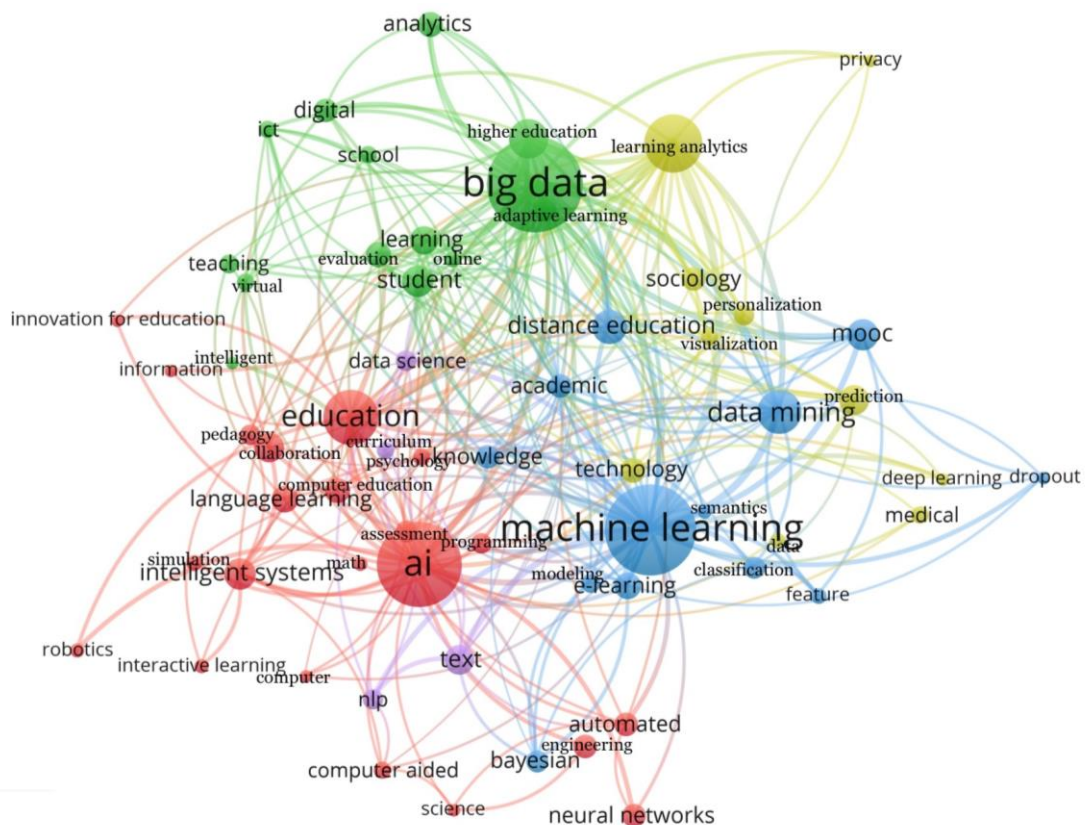

Figure 3 A map based on bibliographic keyword co-occurrence (author keywords)

The green "Big Data" cluster contains topics such as evaluation, adaptive learning. Note that it is close to the yellow "learning analysis" cluster that includes topics such as sociology, prediction, privacy.

The blue "machine learning" cluster relates to topics such as data mining, online education, academic, classification.

The red "AI" cluster includes topics such as intelligent systems, interactive learning, neural network, automated, collaboration, etc.

In-between the "Big Data" cluster, the "machine learning" cluster and the "AI" cluster lies the purple cluster of "data science", including topics such as NLP (natural language processing), and text analysis.

3.3. Emerging topics in details

Based on the above clustering outcomes, we further used the Python scripts we developed to identify emerging topics in each cluster.

The first and main "multidisciplinary" cluster #1 discusses a wide range of topics such as learning analytics, intelligent tutoring systems, collaborative learning, interactive learning environments, cloud computing, Massive Open Online Courses (MOOCs), etc.

The second main "education technology" cluster #2 also includes topics such as learning analytics, MOOCs as cluster #1, but its focus on personal learning is relatively more significant. Apart from that, the remaining clusters have much more sparse research subjects.

4. CONCLUSION

The purpose of this article is to provide a roadmap to visualize the current research fronts of big data and AI research in education. Based on our systematic mapping of 980 articles in the WoS database, we have found that there has been a significant increase in publications since 2014, and specific research fronts have been described.

Among the 7 clusters of publication outlets, chief among them are multidisciplinary, education technology, and information science clusters. Other smaller specialty outlets such as health care, psychology, chemical education, and engineering pedagogy. The main research topics include learning analytics and MOOCs. The "multidisciplinary" cluster contains work on intelligent tutoring systems, collaborative learning, interactive learning environments, etc., whereas the second "education technology" cluster features topics such as personal learning. The findings above indicate that educators and education researchers expect the increasing impacts of AI and big data on learning. In particular, the "data science" cluster, including topics such as NLP and text analysis appear to be important in bridging these keyword clusters. Future work can look at the development of social or human-centered data science.

The data-driven and digital transformation aspects of our everyday life present both possibilities and risks for learning. Especially when the outbreak of COVID-19 has led to the closure of schools, affecting over 91% students worldwide[6], the AI and big data applications in education must contribute to the digital efforts such as distance education solutions and recommendations as effective and meaningful response to the situation, so as to improve the learning environments and outcomes.

Although the findings presented here may be limited in-depth, they nonetheless lay the groundwork with a systematic mapping, so that educators, researchers, and policymakers can use as a roadmap to identify both opportunities and risks in the recent rapid development of the research fronts of education-related AI and big data.

ACKNOWLEDGMENT

REFERENCES

[1] G. Dishon, New data, old tensions: Big data, personalized learning, and the challenges of progressive education, Theory Res. Educ. 15 (2017) 272. https://doi.org/10.1177/1477878517735233.

[2] I. Magnisalis, S. Demetriadis, A. Karakostas, Adaptive and Intelligent Systems for Collaborative Learning Support: A Review of the Field, IEEE Trans. Learning Technol. 4 (2011) 5–20. https://doi.org/10.1109/TLT.2011.2.

[3] R.A. Calvo, S.T. O'Rourke, J. Jones, K. Yacef, P. Reimann, Collaborative Writing Support Tools on the Cloud, IEEE Transactions on Learning Technologies. 4 (2011) 88–97. https://doi.org/10.1109/TLT.2010.43.

[4] C.N. Roy, Anemia of Inflammation, Hematology Am Soc Hematol Educ Program. 2010 (2010) 276–280. https://doi.org/10.1182/asheducation-2010.1.276.

[5] L. Shaughnessy, S.R. Doshi, S.E. Jones, Attempted suicide and associated health risk behaviors among Native American high school students, J Sch Health. 74 (2004) 177–182. https://doi.org/10.1111/j.1746-1561.2004.tb08217.x.

[6] UNESCO, Global Education Coalition, Global Education Coalition UNESCO. (2020). https://en.unesco.org/covid19/educationresponse/global coalition (accessed April 8, 2020).

[7] P. Johanes, C. Thille, The heart of educational data infrastructures = Conscious humanity and scientific responsibility, not infinite data and limitless experimentation, Br J Educ Technol. 50 (2019) 2959–2973. https://doi.org/10.1111/bjet.12862.

[8] B. Williamson, Policy networks, performance metrics and platform markets: Charting the expanding data infrastructure of higher education, Br J Educ Technol. 50 (2019) 2794–2809. https://doi.org/10.1111/bjet.12849.

[9] K. Kitto, S. Knight, Practical ethics for building learning analytics, Br J Educ Technol. 50 (2019) 2855–2870. https://doi.org/10.1111/bjet.12868.

[10] H.-T. Liao, Z. Wang, X. Wu, Developing a Minimum Viable Product for Big Data and AI Education: Action Research Based on a Two-Year Reform of an Undergraduate Program of Internet and New Media, in: Proceedings of the 2019 4th International Conference on Big Data and Computing, ACM, Guangzhou, China, 2019: pp. 42–47. https://doi.org/10.1145/3335484.3335509.

[11] Z. Wang, Exploring different notions of literacy: A literature review analysis of literacy research related to Artificial Intelligence and Big Data applications, in: AIBDA 2019, Guangzhou, China, 2019.

[12] J. Saltz, M. Skirpan, C. Fiesler, M. Gorelick, T. Yeh, R. Heckman, N. Dewar, N. Beard, Integrating Ethics within Machine-learning Courses, ACM Trans. Comput. Educ. 19 (2019) 1–26. https://doi.org/10.1145/3341164.