

An editorial of “AI + Informetrics”: Multi-disciplinary interactions in the era of big data

Yi Zhang¹, Chengzhi Zhang², Philipp Mayr³, Arho Suominen⁴

¹ Australian Artificial Intelligence Institute, Faculty of Engineering and Information Technology, University of Technology Sydney, NSW, Australia.

² Department of Information Management, Nanjing University of Science and Technology, Nanjing, China

³ GESIS-Leibniz Institute for the Social Sciences, Germany

⁴ Tampere University, Finland

Email: yi.zhang@uts.edu.au; zhangcz@njust.edu.cn; Philipp.Mayr@gesis.org; arho.suominen@tuni.fi

ORCID: 0000-0002-7731-0301 (Yi Zhang); 0000-0001-9522-2914 (Chengzhi Zhang); 0000-0002-6656-1658 (Philipp Mayr); 0000-0001-9844-7799 (Arho Suominen)

Introduction

The Pandora’s Box, opened by big data and artificial intelligence (AI), has released incredible powers to reshape the thinking patterns and behaviors of the modern world. Embracing AI’s extraordinary capabilities in analyzing unstructured and scalable data streams, understanding the uncertainty of scientific semantics, and learning robust models with enhanced reliability and adaptability, the bibliometrics community has already demonstrated impressive passion on introducing this powerful tool to broad information studies and achieved enormous success in turning big data into big value and impact. Recent influential studies on the topic include bibliometrics-enhanced information retrieval (Mayr et al., 2014) and the extraction and evaluation of knowledge entities from scientific documents (C. Zhang et al., 2020). Notably, intelligent bibliometrics (Y. Zhang et al., 2020), emphasizing the development and application of computational models that incorporate AI and data science techniques with bibliometric indicators, raises a particular interest in promoting a research agenda on “AI + Informetrics” (AII). These endeavors with broadened perspectives from machine intelligence have portended far-reaching implications for science (Fortunato et al., 2018). However, how to effectively cohere the power of AI and informetrics to create cross-disciplinary solutions in line with this big data boom is still elusive from neither theoretical nor practical perspectives.

The 1st Workshop on AI + Informetrics was held online co-located with the iConference2021 on March 17, 2021. The goal of this workshop series¹ is to create an interactive cohort for global researchers to exchange ideas, share pilot studies, and scope future

¹ The website of the AII Workshop series: <https://ai-informetrics.github.io/>

directions on this cutting-edge venue. In this workshop, we highlighted efforts on constructing fundamental theories and concepts, developing novel methodologies, bridging conceptual knowledge with practical uses, and creating solutions for real-world needs.

This special issue collected articles presented at the 1st Workshop on AI + Informetrics and relevant external submissions. This collection contains 11 articles, contributed by 43 authors from 8 countries (e.g., China, Greece, Pakistan, South Korea, Australia, the US, Germany, and Finland). Specifically, we summarized the 11 articles and introduced this special issue through the following four categories: Prediction and scholarly recommendation, citation evaluation, tools for science, technology, and innovation (ST&I) studies, and interdisciplinary measures.

AI + Informetrics for Prediction and Scholarly Recommendation

The bibliometrics community has long been interested in broad prediction tasks in information studies, such as predicting research outcomes and impact, and scholarly recommendations for publication venues, peer reviewers, and collaborators.

Within an interesting angle in the science of science – gender bias, Kuppler (2022) analyzed the bibliographical data of 111,156 computer science researchers and facilitated random forests and gradient boosting machines to predict their h-index. The results suggested the under-representation of female researchers in the computer science discipline.

Aiming to develop a recommender system for patent applicants and examiners to efficiently locate and cite prior patents, Choi et al. (2022) enlarged the use of patent features, including textual information, metadata, and patent examiner citation information, and developed a deep learning-based system for embedding, searching, and ranking.

Ali et al. (2022) draw attention to the unignorable connections between authors and their research interests. Document embedding techniques and a memory network were integrated to represent paper contents, capture their relationships with author preferences, and conduct personalized recommendations.

Concentrating on collaborator recommendations, Xi et al. (2022) highlighted the similarity between the research interests of scholars and the topology of their collaboration networks, and developed a recommender system using word embedding and node embedding techniques for knowledge representation.

AI + Informetrics for Citation Evaluation

Understanding the content of references and evaluating their value could be one fundamental task of broad citation analyses. Identifying core features from a reference's limited information and constructing evaluation models to measure references are among the top challenges.

Zhang et al. (2022) emphasized the value of a reference's native information (e.g., citation context and section name), and integrated these features into different neural text representation models. Based on a dataset with labeled citations, the improved classification performance endorsed the effectiveness of their efforts on feature selection and representation.

Driven by the task of identifying important citations, An et al. (2022) utilized a semi-supervised self-training technique, including a module of full text-based feature engineering

with six key features (e.g., authors, semantics, and citations) and a self-training strategy using support vector machine and random forest techniques.

AI + Informetrics for ST&I Studies

Bibliometric data have been recognized as a core ST&I data source for profiling research R&D, analyzing emerging technologies, and understanding scientific activities and behaviors. The involvement of AI and data science techniques further strengthens such analytical capabilities.

Focusing on collaborations between universities and enterprises, Chen et al. (2022) constructed a dual-layer network with the two parties and their research topics, depicted network dynamics through network topology, individual characteristics, and knowledge proximity, and developed a stochastic actor-oriented model to understand the evolution of such collaborations.

With a distinctive angle of mapping the sustainable development goals (SDGs) in ST&I studies, Hajikhani and Suominen (2022) compiled a classification model using the SDG classification of scientific articles. They applied it to classify the SDG relevancy in patent families registered in the European Patent Office.

Motivated by developing a one-stop analytical tool for discovering and visualizing research fronts in scientific articles, Wang et al. (2021) integrated a set of models within a systematic toolkit called ITGInsight, covering functionalities of extracting domain terms, clustering topics, and visualizing topic evolution.

AI + Informetrics for Interdisciplinary Measures

The bibliometrics community has been dedicated to measuring interdisciplinary interactions through diverse approaches (e.g., science maps) for decades. Still, new angles and solutions have been added with cutting-edge AI techniques.

To measure the strength of the interdisciplinary interactions between two disciplines, Huang et al. (2022) developed an analytical framework incorporating citation analysis and semantic analysis. Specifically, citation-based relationships were calculated through direction citations and bibliographical couplings, while topic models and word embedding techniques were applied to calculate semantic similarities.

To demystify the characteristics of knowledge integration in interdisciplinary research from a perspective of knowledge content, Wang et al. (2022) examined the best performance of the BERT model in an eHealth dataset when representing integrated knowledge phrases shared between citation sentences and cited references and classifying them into given knowledge categories.

Conclusions

In the AI + Informetrics workshop and this special issue, we noted the enthusiasm from the global community, demonstrating their strong willingness and interests in practicing this challenging but creative combination and recombination – developing and applying AI-empowered computational models for information studies. Interestingly, in this special issue, we are glad to showcase some successful pilot studies, e.g., various embedding techniques for

knowledge representation, heterogeneous network analytics for discovering in-depth social connections among scientific researchers, and task-driven classifier selection and comparisons. We anticipate AI + Informetrics will be further strengthened with disruptive methodological innovation through AI's vertical involvement, and create impactful and insightful empirical evidence for broad decision-making scenarios.

Acknowledgments

Yi Zhang acknowledges the Discovery Early Career Researcher Award granted by the Australian Research Council (Grant No. DE190100994). Chengzhi Zhang acknowledges the National Natural Science Foundation of China (Grant No. 72074113).

References

- Ali, Z., Qi, G., Kefalas, P., Khusro, S., Khan, I., & Muhammad, K. (2022). SPR-SMN: scientific paper recommendation employing SPECTER with memory network. *Scientometrics*, 1-23. <https://doi.org/10.1007/s11192-022-04425-3>
- An, X., Sun, X., & Xu, S. (2022). Important citations identification with semi-supervised classification model. *Scientometrics*, 1-23. <https://doi.org/10.1007/s11192-021-04212-6>
- Chen, H., Song, X., Jin, Q., & Wang, X. (2022). Network dynamics in university-industry collaboration: a collaboration-knowledge dual-layer network perspective. *Scientometrics*, 1-24. <https://doi.org/10.1007/s11192-022-04330-9>
- Choi, J., Lee, J., Yoon, J., Jang, S., Kim, J., & Choi, S. (2022). A two-stage deep learning-based system for patent citation recommendation. *Scientometrics*, 1-22. <https://doi.org/10.1007/s11192-022-04301-0>
- Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., & Uzzi, B. (2018). Science of science. *Science*, 359(6379), eaao0185.
- Hajikhani, A., & Suominen, A. (2022). Mapping the sustainable development goals (SDGs) in science, technology and innovation: application of machine learning in SDG-oriented artefact detection. *Scientometrics*, 1-33. <https://doi.org/10.1007/s11192-022-04358-x>
- Huang, L., Cai, Y., Zhao, E., Zhang, S., Shu, Y., & Fan, J. (2022). Measuring the interdisciplinarity of Information and Library Science interactions using citation analysis and semantic analysis. *Scientometrics*, 1-29. <https://doi.org/10.1007/s11192-022-04401-x>
- Kuppler, M. (2022). Predicting the future impact of Computer Science researchers: Is there a gender bias? *Scientometrics*, 1-38. <https://doi.org/10.1007/s11192-022-04337-2>
- Mayr, P., Scharnhorst, A., Larsen, B., Schaer, P., & Mutschke, P. (2014). Bibliometric-Enhanced Information Retrieval. In M. de Rijke, T. Kenter, A. P. de Vries, C. Zhai, F. de Jong, K. Radinsky, & K. Hofmann, *Advances in Information Retrieval* 36th European Conference on Information Retrieval, Amsterdam, the Netherlands.
- Wang, S., Mao, J., Cao, Y., & Li, G. (2022). Integrated knowledge content in an interdisciplinary field: identification, classification, and application. *Scientometrics*, 1-34. <https://doi.org/10.1007/s11192-022-04282-0>
- Wang, X., Zhang, S., & Liu, Y. (2021). ITGInsight—discovering and visualizing research fronts in the scientific literature. *Scientometrics*, 1-23. <https://doi.org/10.1007/s11192-021-04190-9>

- Xi, X., Wei, J., Guo, Y., & Duan, W. (2022). Academic collaborations: A recommender framework spanning research interests and network topology. *Scientometrics*.
- Zhang, C., Mayr, P., Lu, W., & Zhang, Y. (2020). Extraction and evaluation of knowledge entities from scientific documents: EEKE2020. Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020,
- Zhang, Y., Porter, A., Cunningham, S. W., Chiavetta, D., & Newman, N. (2020). Parallel or intersecting lines? Intelligent bibliometrics for investigating the involvement of data science in policy analysis. *IEEE Transactions on Engineering Management*, 68(5), 1259-1271.
- Zhang, Y., Zhao, R., Wang, Y., Chen, H., Mahmood, A., Zaib, M., Zhang, W. E., & Sheng, Q. Z. (2022). Towards employing native information in citation function classification. *Scientometrics*, 1-21. <https://doi.org/10.1007/s11192-021-04242-0>