Preface to the 1st Workshop onAI + Informetrics: Multi-disciplinary Interactions on the Era of Big Data

Yi Zhang[1[0000-0001-5263-5823]], Chengzhi Zhang[2[0000-0001-9522-2914]], Philipp Mayr[3[0000-0002-6656-1658]], Arho Suominen[4[0000-0001-9844-7799]]

[1] Australian Artificial Intelligence Institute, University of Technology Sydney, Australia.
[2] Department of Information Management, Nanjing University of Science and Technology, China.
[3] GESIS – Leibniz Institute for the Social Sciences, Germany.
[4] VTT Technical Research Centre of Finland, & Tampere University, Industrial Engineering, Finland.
yi.zhang@uts.edu.au; zhangcz@njust.edu.cn;
Philipp.Mayr@gesis.org; Arho.Suominen@vtt.fi

1 Introduction

Driven by the big data boom, *informetrics*, known as the study of quantitative aspects of information, has gained great benefits from artificial intelligence (AI) [1] – including a wide range of intelligent agents through techniques such as neural networks, genetic programming, computer vision, heuristic search, knowledge representation and reasoning, Bayes network, planning and language understanding. With its capacities in analyzing unstructured scalable data and streams, understanding uncertain semantics, and developing robust and repeatable models, "*AI + Informetrics*" has demonstrated enormous success in turning big data into big value and impact by handling diverse challenges raised from multiple disciplines and research areas. For example, bibliometric-enhanced information retrieval [2], science mapping with topic models [3], streaming data analytics for tracking technological change [4], and entity extraction with unsupervised machine learning techniques [5]. Such endeavors with broadened perspectives from machine intelligence would portend far-reaching implications for science [6], but how to effectively cohere the power of AI and informetrics to create cross-disciplinary solutions is still elusive from neither theoretical nor practical perspectives.

The 1st Workshop on AI + Informetrics (Virtual Meeting) on March 17, 2021 has successfully constructed a collaborative platform to gather more than 50 researchers and practical users for exchanging ideas, sharing pilot studies, and scoping future directions on this cutting-edge venue. We particularly highlight "AI + Informetrics" as endeavors in constructing fundamental theories, developing novel methodologies, bridging conceptual knowledge with practical uses, and creating real-word solutions.

2 Submission Overview

In this workshop, we have accepted 17 papers for presentation and inclusion in the proceedings – including 5 long papers and 12 short papers. We also invited two keynotes touching the frontiers and pilot studies in line with AI + Informetrics. All workshop contributions are documented in the workshop website (https://ai-informetrics.github.io/). The following section briefly lists the various contributions.

2.1 Keynotes

Keynote 1: *AI and Science of Science* by Prof. Ying Ding, School of Information, University of Texas at Austin

Artificial Intelligence (AI) has fundamentally changed every aspect of our lives. In AI, the half-life of a paper could be less than one year which means that new algorithms have been developed and become out of date within just one year, sometimes could be just few months. Computer vision leads the newly development of fascinating AI algorithms, which are then diffused to natural language processing and graph mining. AI has challenged our understandings of collaboration, knowledge diffusion, and even citing behavior. The new concepts of human-machine teaming, cognitive computing in knowledge diffusion, citing future rather past are all happening right now at the AI era. This talk will highlight several research practices and share the thoughts about the current and future of AI and Science of Science.

Keynote 2: *Global Models of Science and Their Applications* by Dr. Kevin Boyack, SciTech Strategies

Global models based on full literature databases are far more accurate at characterizing the structure of science than local models that are based on keyword or journal sets. However, local models continue to form the basis of most scientometrics studies, likely due to the lack of access to full databases by most practitioners. In this presentation we 1) examine the relative accuracies of global and local models, 2) explain the applications and benefits of using global models, and 3) show how global models can be created by nearly anyone using today's open access resources.

2.2 Research Papers

The 17 accepted papers were organized in four sessions.

Session 1: AI + Informetrics for Scholarly Recommendation

This session highlights the use of novel AI techniques in developing recommender systems for academic activities, such as book recommendation and collaborator recommendation.

Xin Zhang, Yi Wen, and Haiyu Xu proposed a network model for collaboration prediction, incorporating heterogeneous bibliographical information with network embedding techniques.

Jaeyoung Choi et al. touched the issues of book recommendation in university libraries and an embedding-based neural network model was developed.

Hongshu Chen and Xinna Song predicted the collaborative patterns between universities and industry sectors by taking collaboration and knowledge networks into account.

Xiaowen Xi, Ying Guo, and Weiyu Duan exploited word embedding and network embedding techniques for recommending academic collaborators.

Session 2: AI + Informetrics for Knowledge Extraction

This session collects contributions on developing novel AI-empowered models for knowledge and entity extraction from bibliometric documents.

Bolin Hua and Youngkug Shin concentrated on the conclusion section of academic papers and developed a model of extracting sentences describing the originality of the work.

Chuan Jiang et al. held interests in software entities mentioned in the full text of research articles and employed a BERT model to particularly extract these entities.

Shiyun Wang, Jin Mao, and Hao Xie developed an automatic model for identifying and classifying integrated knowledge contents in interdisciplinary fields.

Zekun Deng et al. created a solution of automatically generating a Related Work section by using sentence extraction and reordering techniques.

Session 3: AI + Informetrics for Information Studies

This session provides examples of incorporating AI techniques and informetric approaches in investigating broad information studies, such as interdisciplinary measurements, research evaluation, and job detection.

Asta Back, Arash Hajikhani, and Arho Suominen exploited text mining techniques to analyze job advertisement data for job detection.

Sha Yuan et al. targeted to the long-term scientific impacts of research papers and developed a citation-based prediction model to foresee such an impact. Interestingly, this paper highlights that the limited attention can better stand on the shoulders of giants.

Lu Huang et al. introduced a hybrid model combining citation statistics and semantic to measure interdisciplinary interactions.

Ruiyuan Li, Pin Tian, and Shenghui Wang glanced over the English literature published in the past 150 years and observed the drift of semantics over time.

Session 4: AI + Informetrics for Science, Technology & Innovation

This session demonstrates research approaches and examples with AI + informetrics for science, technology & innovation studies, e.g., tech mining and technology opportunity analysis, and topics in science of science.

Xin An, Xin Sun, and Shuo Xu proposed a semi-supervised classification model for identifying important citations.

Jingwen Luo et al. conducted emerging technology opportunity analysis by assembling network analytics and burst detection.

Arash Hajikhani and Arho Suonimen applied a machine learning approach to investigate the interrelations of sustainable development goals in scientific publications and patents.

Junwan Liu and Rui Wang introduced LSTM neural networks to predict the research performance of scientists in terms of research productivity and citations.

Xuefeng Wang, Shuo Zhang, and Yuqin Liu developed a system tool called ITGInsight for knowledge discovery and visualization, particularly competitive technological intelligence.

3 Outlook and Further Reading

The 1st Workshop on AI + Informetrics has achieved great success with attentions from the research community and outcomes on either novel technological development or insightful empirical practices. A special issue “AI + Informetrics” associated with the bibliometric venue *Scientometrics* is calling for papers, and please find more information on the website: https://link.springer.com/collections/ebfiegeiie.

References

1. Nilsson, N. J., Nilsson, N. J.: Artificial intelligence: a new synthesis. Morgan Kaufmann (1998).
2. Mayr, P., Scharnhorst, A., Larsen, B., Schaer, P. Mutschke, P.: Bibliometric-enhanced information retrieval. In: European Conference on Information Retrieval, pp. 798-801. Springer, Cham (2014).
3. Suominen, A. Toivanen, H.: Map of science with topic modeling: Comparison of unsupervised learning and human‐assigned subject classification. Journal of the Association for Information Science and Technology, 67(10), pp. 2464-2476 (2016).
4. Zhang, Y., Zhang, G., Zhu, D. Lu, J.: Scientific evolutionary pathways: Identifying and visualizing relationships for scientific topics. Journal of the Association for Information Science and Technology, 68(8), pp.1925-1939 (2017).
5. Zhang, Y. Zhang, C.: Unsupervised keyphrase extraction in academic publications using human attention. In: 17th International Conference of the International Society for Scientometrics and Informetrics, pp. 2483-2484, Italy, Rome (2019).
6. Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., ..., Barabási, A. L.: Science of science. Science, 359(6379), eaao0185 (2018).