

Applying transformers and transfer learning for training and predicting topological features of AI citation networks

Speaker: Peter Tsung-Wen Yen

Division of Physics and Applied Physics, School of Physical and Mathematical Sciences,
Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798

Keywords: transformers, transfer learning, Science-of-science, bibliographic citation networks, Machine Learning

Extended Abstract

Motivation. Artificial intelligence (AI) and Machine Learning (ML) are currently the most strategic modern technologies being pursued in many countries, including the US, China, Southeast Asian countries, and European countries. Most AI scientists focus on how to design faster and more robust algorithms to perform ML tasks; others increase the number of neural network layers and the number of GPUs so as to scale up the performance of ML algorithms. Nevertheless, a more fundamental scientific question in AI is often neglected, and that is how AI and ML research fields evolved throughout the past 40 years. It is commonly believed that there have been three summers and two winters since the start of AI science at 1950s, and we are currently in the third AI summer [1, 2]. It strikes us that we should use real data and apply data-driven models to recreate such AI history figures, which will be less-biased and more convincing. From our most recent work [3], we identified persistent bibliographic citation networks (BCNs) in different years, and also constructed alluvial diagrams, from which we can recreate such AI history from data and models, and track the evolution of research topics of each BCN in different years.

However, as the bibliographic records start to accumulate after 2016, tracking such research topics becomes harder and slower, and eventually, we witness a bottleneck in 2022, when the bibliographic records had already surpassed 110,000 papers. In this regard, we propose to use the number of papers in each top BCNs and their Betti numbers β_n , to train a transformer model (Figure 1(A)) at weights $w = 5$, to see if the parameters and hyperparameters after training can give us an acceptable accuracy rate. Equipped with a recently developed algorithm, which can handle extremely large datasets in computing β_n by resorting to parallel computing and memory sharing, the β_n computation barrier can now be pushed forward till the end of 2023.

Approach and Methodology. The bibliographic records are first downloaded from the Web of Science website, using “artificial intelligence”, and “neural networks” as keywords, which consist of nearly 1.2 million records from 1985 to 2023. We then compute BCNs in each year at different weights w , where BCNs can be identified by using a simple partitioning algorithm. For each BCN in each year, we compute its size N and β_n , $n = 0, 1, 2$, using the new algorithm developed by Ian Yip Ho Yan, who is currently a Ph.D. student studying in our lab, which can slice larger BCNs into smaller lumps for computing its β_n s separately in a parallel fashion, and later being recombined to aggregate the final β_n results. BCNs can also be used to construct alluvial diagrams using R packages `networkflow`, `ggplot2`, and `ggalluvial`. For each

BCN, we collect the titles of all articles in it and ask ChatGPT-5.2 to identify the research topics for them. We train the transformer models using bibliographic data up to 2018 as training sets, and use subsequent years 2019 to 2023 as validation sets, to test their accuracy, before using the trained models to make predictions of the evolution of the size of top BCNs, and their β_n s in later years. Together with BCNs, alluvial diagrams, and also the identified research topics, we can now tell a coherent story regarding the evolution of research topics as well as information processing pathways in terms of splitting, merging, and continuation. For the parameters and hyperparameters obtained at $w = 5$, we apply transfer learning to predict the results of nearby weights $w = 4$, and 6, by only changing a small portion of parameters fitted to $w = 5$ BCNs (see Figure 1(B) for illustrations).

Results. We have computed the number of top 20 BCNs and their respective β_n s from 1991 to 2018, and used them for fitting the transformer model's weights. The accuracy rate after 100 epochs is 70%, and we are now trying to improve the accuracy rate up to 80% or more. After that, we will use the 2019 to 2023 BCN results for validation tests and also make predictions to benchmark the models' performance. Later, we will use the fitted model at $w = 5$ to predict nearby $w = 4$, and 6 BCNs results, before applying transfer learning to $w = 4$, and 6 to validate the performance and accuracy of applying transfer learning to our BCNs data.

Conclusions and Outlook. *Self-organization* and *complex phenomena* have multiscale characteristics in nature; in addition, *synergy* plays an essential role in connecting the two. As such, how these three layers mediate each other currently constitutes a research gap in complex systems. The transformer architecture, in particular, the attention layers can capture spatiotemporal characteristics that are inherent in our BCN data set; therefore, we want to exploit this supremacy to test the capability of the transformer and transfer learning to gain more insights into the causal relations of the three layers of AI citation networks, to elucidate complex systems' structure and dynamics more deep and thorough. We suspect transformers are promising in performing complex system-informed machine learning tasks, such as predicting the topological features of BCNs in later years, interpreting how information progresses within the dataset, or translating information contents in BCNs to human-readable languages, to name just a few.

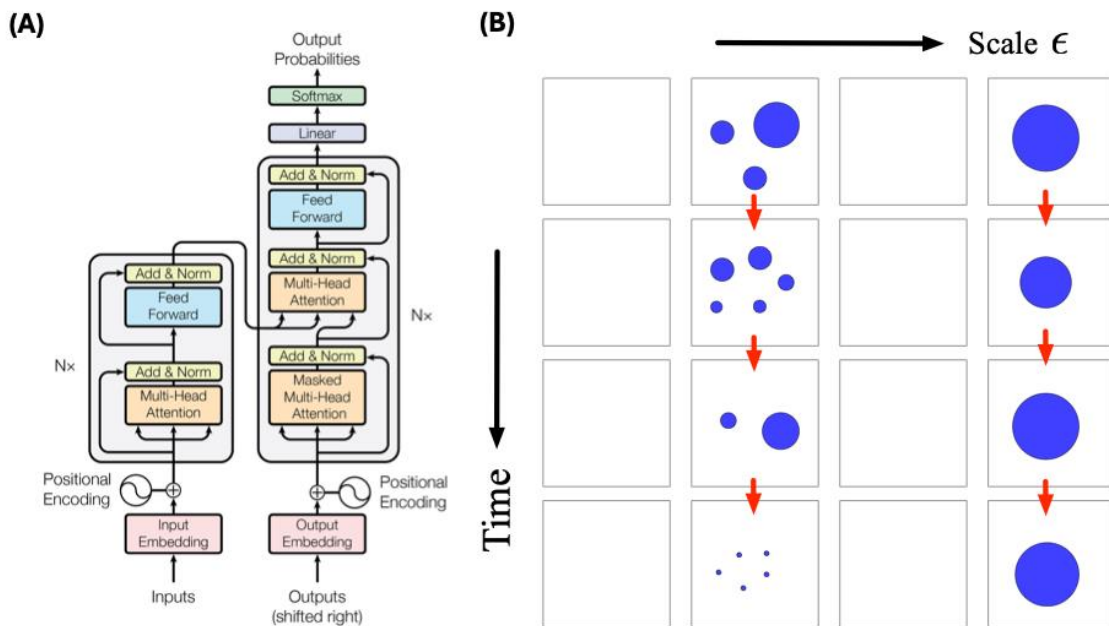


Figure 1. (A) A prototypic transformer model [4]. (B) A schematic diagram illustrating persistent structures emerging at different spatial length scales ϵ , and across different times. The Betti numbers of these persistent structures at ϵ_i can then be computed and fed into transformer models. The weights fitted at ϵ_i can then be stored, which can then be used in subsequent transfer learning tasks at fitting the weights of ϵ_{i+1} or ϵ_{i-1} , etc.

References

1. Lim, M. *History of AI Winters*. 2018 [cited 2026 02/11]; Available from: <https://www.actuaries.asn.au/research-analysis/history-of-ai-winters>.
2. Toosi, A., Bottino, A.G., Saboury, B., Siegel, E., and Rahmim, A., *A brief history of AI: how to prevent another winter (a critical review)*. *PET clinics*, **16**(4): p. 449-469, 2021.
3. Yen, T.-W., *SpatioTemporal Information Processing Pathways Identified in AI Citation Networks*, in *NetSciX 2026*. 2026: Auckland, New Zealand.
4. Google. *Attention Is All You Need - Encoder-decoder Architecture*. 2023 [cited 2026 03/10]; Available from: https://commons.wikimedia.org/wiki/File:Attention_Is_All_You_Need_-_Encoder-decoder_Architecture.png.