Zuoyu Yan

Ph.D. Student ·

Wangxuan Institute of Computer Technology, Peking University

+86 18811318958 | yanzuoyu3@pku.edu.cn | https://pkuyzy.github.io/ | https://github.com/pkuyzy

Research Interest

My research interest lies in the intersection of machine learning and topological data analysis. In particular, I am interested in:

- Enhancing modern machine learning frameworks, e.g., graph neural networks, with topology/geometry.
- Applying these models in real-world scenarios, e.g., biomedicine, and recommendation systems.

Education

Peking University

Ph.D. Student, Computer Science *2019.09 till now*

Peking University

Undergraduate Degree, Data Science *2015.09 - 2019.06*

Intern Experience

Alibaba DAMO Academy

2021.05-2021.10

- Multimodal entity linking on Wikipedia and Wikinews data.

Publications

Zuoyu Yan, Tengfei Ma, Liangcai Gao, Zhi Tang, Yusu Wang, Chao Chen. Learning on Graphs Conference (**LoG**), 2023. (**Oral**)

Zuoyu Yan, Junru Zhou, Liangcai Gao, Zhi Tang, Muhan Zhang. Efficiently Counting Substructures by Subgraph GNNs without Running GNN on Subgraphs. arXiv, 2023.

Zuoyu Yan, Tengfei Ma, Liangcai Gao, Zhi Tang, Yusu Wang, Chao Chen. Neural Approximation of Graph Topological Features. Advances in Neural Information Processing Systems (**NeurIPS**, Top conference in machine learning.), 2022. (**Spotlight, acceptance rate 4%-5%**)

Zuoyu Yan, Tengfei Ma, Liangcai Gao, Zhi Tang, Chao Chen. Cycle Representation Learning for Inductive Relation Prediction. International Conference on Machine Learning (**ICML**, Top conference in machine learning.), 2022.(**Short talk, acceptance rate 21.9%**)

Zuoyu Yan, Tengfei Ma, Liangcai Gao, Zhi Tang, Chao Chen. Link Prediction with Persistent Homology: An Interactive View. International Conference on Machine Learning (**ICML**, Top conference in machine learning.), 2021.(**Short talk, acceptance rate 21.5%**)

Zuoyu Yan, Xinpeng Zhang, Liangcai Gao, Ke Yuan, Zhi Tang. ConvMath: A Convolutional Sequence Network for Mathematical Expression Recognition. International Conference on Pattern Recognition (**ICPR**), 2020.

Wenqi Zhao, Liangcai Gao, **Zuoyu Yan**, Shuai Peng, Lin Du, Ziyin Zhang. Handwritten mathematical expression recognition with bidirectionally trained transformer. International Conference on Document Analysis and Recognition (**ICDAR**), 2021. (**Best poster award, 2 out of 340 submissions**).

Liangcai Gao, Xiaohan Yi, Yuan Liao, Zhuoren Jiang, **Zuoyu Yan**, Zhi Tang. A deep learning-based formula detection method for PDF documents. International Conference on Document Analysis and Recognition (**ICDAR**) 2017.

Honors

- Won the 2021-2022th ”Outstanding Research Award” of Peking University. (15% out of 27 PhD students of Wangxuan Institute of Computer Technology of Peking University (**WICT**))
- Won the 2021-2022th ”Outstanding Student Award” of WICT. (7% out of 70 students of WICT)
- Won the 2018-2019th Excellent Project of Peking University President’s Fund (10% out of 97 projects)
- Won the third prize in the ”Schlumberger Cup” programming competition held in Peking University.
- Ranked 8th(out of 836 teams) in the Plant Seedlings Classification competition held on Kaggle website

Skills

- Languages: C/C++, Matlab, Python
- Tools: Torch, Tensorflow, PyTorch, OpenCV

Past projects

On one hand, classic machine learning methods leverage data samples without considering data topology/geometry, leading to their limited representation power. On the other hand, powerful machine learning models/topological features are usually computationally costly, and can hardly be applied in complex and large graphs which are common in real-world scenarios.

Graph Machine Learning. Structures represent meaningful components (e.g., rings, triadic closure property) in graph data and thus are essential for graph machine learning. However, classic graph machine learning models cannot detect the existence of many structures (e.g., cycles, cliques) and fail in many relevant tasks. To enhance the representation power of these models, I design strong and powerful topological features that contain important structural information and combine them with graph learning models organically (**ICML’ 2021, ICML’2022, arXiv’ 2023, LoG’ 2023**). Theoretical and empirical results on various benchmarks show the enhanced power of the proposed model.

Topological Representation Acceleration. To accelerate the computation of powerful topological features, I propose provably faster algorithm (**ICML’ 2021**) to accelerate their computation. Furthermore, I also develop machine learning models with strong algorithmic alignment to approximate these features (**NeurIPS’ 2022**). While achieving strong approximation results, the proposed model is nearly 100 times faster than conventional algorithms. With its observed transferability, the proposed model can be potentially applied to multiple real-world datasets.

Professional Service

Conference Reviewer for:

- Conference on Neural Information Processing Systems (NeurIPS)
- International Conference on Machine Learning (ICML)
- International Conference on Learning Representations (ICLR)
- Learning on Graphs Conference (LoG)

Reference

- **Zhi Tang**

Professor, Peking University
tangzhi@pku.edu.cn
https://www.wict.pku.edu.cn/cpdp/kydw/ggcy/1297369.htm

- **Chao Chen**
  Associate Professor, Stony Brook University
  chao.chen.1@stonybrook.edu
  https://chaochen.github.io/
- **Yusu Wang**
  Professor, University of California, San Diego
  yusuwang@ucsd.edu
  http://yusu.belkin-wang.org/
- **Tengfei Ma**
  Assistant Professor, Stony Brook University
  feitengma0123@gmail.com
  https://sites.google.com/site/matf0123/
- **Muhan Zhang**
  Assistant Professor, Peking University
  muhan@pku.edu.cn
  https://muhanzhang.github.io/