Chenxiao Yang

✉ chr26195@sjtu.edu.cn | github.com/chr26195 | 🎓 Google Scholar

Education

Shanghai Jiao Tong University   *Shanghai, China*

M.S. in Computer Science   *2021.09 - 2024.03 (expected)*

- Advised by Prof. Junchi Yan
- Research Interests: Machine Learning on Graphs, Geometric Deep Learning, Generalization

Shanghai Jiao Tong University   *Shanghai, China*

B.S. in Computer Science   *2017.09 - 2021.07*

- Overall GPA: 89.6 / 100, Linear Algebra: 100 / 100, Probability and Statistics: 99 / 100
- Worked with Prof. Xiaofeng Gao in the last and a half years
- Research Interests: Data Mining, Recommender Systems, Online Advertisement

Selected Publications

MACHINE LEARNING ON GRAPHS / RELATIONAL DATA

Graph Neural Networks are Inherently Good Generalizers: Insights by Bridging GNNs and MLPs
Chenxiao Yang, Qitian Wu, Jiahua Wang, Junchi Yan
International Conference on Learning Representations (ICLR), 2023

DIFFormer: Scalable (Graph) Transformers Induced by Energy Constrained Diffusion
Qitian Wu, Chenxiao Yang, Wentao Zhao, Yixuan He, David Wipf, Junchi Yan
International Conference on Learning Representations (ICLR spotlight), 2023

Energy-based Out-of-Distribution Detection for Graph Neural Networks
Qitian Wu, Yiting Chen, Chenxiao Yang, Junchi Yan
International Conference on Learning Representations (ICLR), 2023

Geometric Knowledge Distillation: Topology Compression for Graph Neural Networks
Chenxiao Yang, Qitian Wu, Junchi Yan
Advances in Neural Information Processing Systems (NeurIPS), 2022

Towards Out-of-Distribution Sequential Event Prediction: A Causal Treatment
Chenxiao Yang, Qitian Wu, Qingsong Wen, Zhiqiang Zhou, Liang Sun, Junchi Yan
Advances in Neural Information Processing Systems (NeurIPS), 2022

Towards Open-World Feature Extrapolation: An Inductive Graph Learning Approach
Qitian Wu, Chenxiao Yang, Junchi Yan
Advances in Neural Information Processing Systems (NeurIPS), 2021

DATA MINING

Cross-task Knowledge Distillation in Multi-Task Recommendation
Chenxiao Yang, Junwei Pan, Xiaofeng Gao, Tingyu Jiang, Dapeng Liu, Guihai Chen
AAAI Conference on Artificial Intelligence (AAAI), 2022

Trading Hard Negatives and True Negatives: A Debiased Contrastive Collaborative Filtering Approach
Chenxiao Yang, Qitian Wu, Jipeng Jin, Xiaofeng Gao, Junwei Pan, Guihai Chen
International Joint Conference on Artificial Intelligence (IJCAI), 2022

Seq2bubbles: Region-based Embedding Learning for User Behaviors in Sequential Recommenders
Qitian Wu, Chenxiao Yang, Shuodian Yu, Xiaofeng Gao, Guihai Chen
ACM International Conference on Information & Knowledge Management (CIKM spotlight), 2021

Selected Research Projects

Understand GNN Generalization and Accelerate GNN Trainin.

(Lead all parts of the work. Accepted to ICLR 2023.)   *2022.05 - 2022.10*

- Propose an intermediate model class between GNNs and MLPs called PMLP that is identical to MLP in training but adopts GNNs' architectures in testing. PMLPs are very efficient to train, more robust to structural noises, and can perform on par with GNNs in many experimental settings.
- Reveal GNNs' inherent generalization ability implicit in the message passing based architecture itself, which is shown to be the major source of their superior performance in node-level prediction tasks, in contrast with conventional wisdom.
- Analyze MLP, PMLP and GNN in over-parameterized regime, and derive explicit formulas for computing GNTK and the corresponding feature map in node regression tasks. Extrapolation analysis reveals common advantages of GNNs and PMLPs in generalizing to out-of-distribution testing samples near the training data support.
- **Keywords:** Efficient GNNs, Generalization, Extrapolation Analysis, Overparameterization.

Out-of-Distribution Detection on Non-IID Data for Trustworthy GNNs.

(Conceptualization and theoretical analysis. Accepted to ICLR 2023.)   *2022.07 - 2022.10*

- Formulate out-of-distribution detection problem on non-IID data with graph structure, and uncover OoD discriminators naturally present in every GNN classifier trained with standard learning objective from the perspective of energy models.
- Propose energy-based belief propagation scheme to boosting the detection consensus over graph topology. Theoretical analysis shows the method is provably effective for yielding distinguishable scores for in-distribution and OoD nodes.
- **Keywords:** OoD Detection, Energy Model, Trustworthy AI.

Diffusion-Based Graph Transformer with Linear Complexity.

(Conceptualization, modeling and writing. Accepted to ICLR 2023 spotlight with average score 8.)   *2022.02 - 2022.07*

- Propose a new theoretical framework to uncover data interdependencies via a diffusion process describing microscopic behaviors of feature evolution, and an energy function as macroscopic quantification of global consistency. The framework can yield popular models such as MLP, GCN and GAT up to choices of the diffusivity function.
- Introduce a new class of (graph) Transformers called DIFFormer as natural derivatives under the proposed framework. It includes an instantiation with linear complexity w.r.t. number of instances enabling it to scale to large datasets.
- Experiments on a wide variety of tasks such as node and image classification, spatial-temporal dynamics prediction show the superior performance of DIFFormer.
- **Keywords:** Efficient Graph Transformers, Diffusion on Graphs, Geometric Deep Learning, Energy Unfolding.

Characterize Geometric Property of Latent Graph Manifold and Knowledge Transfer.

(Lead all parts of the work. Published in NeurIPS 2022)   *2021.10 - 2022.02*

- Investigate on a new paradigm of knowledge transfer that aims at encoding graph topological information into GNNs such that they can effectively and efficiently operate on smaller or sparser graphs.
- Introduce a new notion called *Neural Heat Kernel (NHK)*, an extension of conventional heat kernel to graph neural diffusion, for uncovering geometric property of the latent graph manifold associated with GNNs.
- Propose a flexible and principled distillation framework for encoding and transferring geometric knowledge by aligning NHKs behind GNN models. The approach can also be applied for conventional KD purposes.
- **Keywords:** Efficient GNNs, Diffusion on Graphs, Geometric Deep Learning, OoD Generalization (under Topological Shift).

Inductive (Unseen) Feature Embedding Learning on Tabular Data by GNNs.

(Conceptualization and theoretical analysis. Published in NeurIPS 2021)   *2021.02 - 2021.10*

- Formulate open-world feature extrapolation problem where input space would change during inference and models are supposed to infer new (discrete) feature embeddings in an inductive manner without retraining.
- Proposed a graph-enhanced feature extrapolation approach and derived its generalization error bound.
- Conducted experiments on six UCI datasets and two large-scale advertisement click prediction datasets.
- **Keywords:** Tabular Data Learning, Generalization, Feature Extrapolation, Inductive Learning.

Industry Experience

DAMO Academy, Alibaba (U.S.) Inc.   *Remote*

Research Intern. (Leading all parts of the work. Published in NeurIPS 2022)   *2021.06 - 2021.10*

- Research on temporal distribution shifts on sequential data, and attempt to improve models' out-of-distribution generalization performance in future new environments.
- Develop a new framework based on causal intervention and variational inference, called *Variational Context Adjustment*, to cancel out the confounding effect of latent contexts to facilitate models to explore the true causality between random variables.
- Propose a flexible framework as instantiation, which could be freely combined with most off-the-shelf sequence backbone models.
- **Keywords:** Causal Intervention, Variational Inference, OoD Generalization (under Temporal Shift).

Technology and Engineering Group, Tencent Inc.   *Shenzhen, China*

Research Intern. (Conceptualized together with Junwei Pan. Published in AAAI 2022)   *2020.12 - 2021.06*

- Research on transferring user preference information/knowledge across different platforms or tasks for accurate online advertisement.
- Propose a new knowledge distillation framework to distill ranking-based knowledge across tasks with considerations of task conflict, prediction alignment and synchronous optimization.
- The framework is deployed on the WeChat platform (i.e., the most popular Chinese social media) and improved the AUC performance of its recommender system.
- **Keywords:** Recommender Systems, Online Advertisement, CTR Prediction, Multi-Task Learning, Knowledge Transfer.

Academic Service

I served as program committee or reviewer in ICML (2022, 2023), NeurIPS (2022), AAAI (2023), IJCAI (2023). I also write blogs (currently in Chinese but will be extended to English in the near future) to share my recent works on platforms such as Zhihu (i.e., a Chinese knowledge-sharing platform).

Selected Awards

2022	**Top 1%,** National Scholarship for Graduate Students	*China*
2019	**Honorable Mention**, Mathematical Contest in Modeling	*U.S.*
2018	**First Prize, Top 5%,** Spring Semester Course Project Exhibition	*China*
2016	**First Prize**, National Mathematics Competition (Zhejiang)	*China*

Skills

Programming	Python (NumPy, Matplotlib, PyTorch, PyG, etc.), C/C++, HTML/CSS, JavaScript, etc.
Languages	Mandarin Chinese (Native proficiency), English (Professional proficiency).
Miscellaneous	Linux, Shell, LaTeX(Overleaf), Markdown, Microsoft Office, Git, etc.