LIANG YAO (姚亮)

(+86) 131-7369-4368 ◇ yaoliang@zju.edu.cn
https://scholar.google.com/citations?user=45HM9IgAAAAJhl=en
https://yao8839836.github.io/

RESEARCH INTERESTS

Natural Language Processing
Data Mining
Medical Informatics
Large Language Models
Graph Neural Networks
Probabilistic Graphical Models

Google Scholar citations: 3708, H-index: 21

WORK EXPERIENCE

Tencent Inc. *Sep 2019- Present*
Senior Researcher
Shenzhen, China

EDUCATION

Northwestern University *Jan 2018- Sep 2019*
Postdoctoral Fellow
Chicago, USA

Zhejiang University *Sep 6th, 2012- Sep 30th, 2017*
Ph.D. in Computer Science
Hangzhou, China

Sichuan University *Sep 2008- June 2012*
B.Eng. in Computer Science and Technology
Minor in Finance
Chengdu, China

PUBLICATIONS

Liang Yao, Chengsheng Mao, Yuan Luo. Graph Convolutional Networks for Text Classification. *33rd AAAI Conference on Artificial Intelligence* (**AAAI 2019**), pp. 7370-7377. (citations: **2012**, GitHub star: **1341**)

Liang Yao, Chengsheng Mao, Yuan Luo. KG-BERT: BERT for Knowledge Graph Completion. Submitted to *IEEE Transactions on Pattern Analysis and Machine Intelligence* (**TPAMI**, citations: **582**, GitHub star: **659**)

Liang Yao. Large Language Models are Contrastive Reasoners. arXiv preprint arXiv:2403.08211 (2024). Submitted to NeurIPS 2024.

Liang Yao, Jiazhen Peng, Chengsheng Mao, Yuan Luo. Exploring Large Language Models for Knowledge Graph Completion. arXiv preprint arXiv:2308.13916 (2023). Submitted to EMNLP 2024.

Liang Yao, Jiazhen Peng, Shenggong Ji, Qiang Liu, Hongyun Cai, Feng He, Xu Cheng. Friend Ranking in Online Games via Pre-training Edge Transformers. Accepted by **SIGIR 2023**.

Liang Yao, Yin Zhang, Baogang Wei, Wenjin Zhang, Zhe Jin. A Topic Modeling Approach for Traditional Chinese Medicine Prescriptions. *IEEE Transactions on Knowledge and Data Engineering* (**TKDE**) 30.6 (2018): 1007-1021. (SCI, IF:8.9)

Liang Yao, Yin Zhang, Baogang Wei, Zhe Jin, Rui Zhang, Yangyang Zhang, Qinfei Chen. Incorporating Knowledge Graph Embeddings into Topic Modeling. In *31st AAAI Conference on Artificial Intelligence* (**AAAI 2017**) pp. 3119-3126.

Xu Cheng, **Liang Yao***, Feng He, Chenhui Zhang, Wenzheng Feng, Jie Tang. LPS-GNN: Deploying Graph Neural Networks on Graphs with 100-Billion Edges. Submitted to **TKDE**. *Corresponding author.

Liang Yao, Zhe Jin, Chengsheng Mao, Yin Zhang and Yuan Luo. Traditional Chinese Medicine Clinical Records Classification with BERT and Domain Specific Corpora. *Journal of the American Medical Informatics Association* (**JAMIA**). 26, no. 12 (2019): 1632-1636. (SCI, IF: 6.4)

Chengsheng Mao, **Liang Yao**, and Yuan Luo. "Imagegcn: Multi-relational image graph convolutional networks for disease identification with chest x-rays." IEEE Transactions on Medical Imaging 41, no. 8 (2022): 1990-2003. (SCI, IF: 10.6)

Chengsheng Mao, **Liang Yao** and Yuan Luo. "MedGCN: Medication recommendation and lab test imputation via graph convolutional networks." Journal of Biomedical Informatics 127 (2022): 104000. (SCI, IF:4.5)

Liang Yao, Chengsheng Mao, Yuan Luo. Clinical Text Classification with Rule-based Features and Knowledge-guided Convolutional Neural Networks. *BMC Medical Informatics and Decision Making*, 9(3) (2019), p.71. SCI, IF: 3.5

Liang Yao, Yin Zhang, Baogang Wei, Zherong Li. Traditional Chinese Medicine Clinical Records Classification using Knowledge-Powered Document Embedding. In *2016 IEEE International Conference on Bioinformatics and Biomedicine* (**BIBM 2016**).

Liang Yao, Yin Zhang, Baogang Wei, Lei Li, Fei Wu, Peng Zhang, Yali Bian. Concept over time: the combination of probabilistic topic model with Wikipedia knowledge. *Expert Systems with Applications* 60 (2016): 27-38. (SCI, IF:8.5)

Liang Yao, Yin Zhang, Qinfei Chen, Hongze Qian, Baogang Wei, Zhifeng Hu. Mining Coherent Topics in Documents using Word Embeddings and Large-scale Text Data. *Engineering Applications of Artificial Intelligence* 64 (2017): 432-439. (SCI, IF:8.0)

Liang Yao, Yin Zhang, Baogang Wei, Wei Wang, Yuejiao Zhang, Xiaolin Ren, and Yali Bian. Discovering treatment patterns in Traditional Chinese Medicine clinical cases by exploiting supervised topic model and domain knowledge. *Journal of Biomedical Informatics* 58 (2015): 260-267. (SCI, IF:4.5)

Yang Zhao, **Liang Yao**, Yin Zhang. Purchase prediction using Tmall-specific features. *Concurrency and Computation: Practice and Experience* 28.14 (2016): 3879-3894. (SCI, IF:2.0)

Liang Yao, Yin Zhang, Baogang Wei, Hongze Qian, and Yibing Wang. Incorporating probabilistic knowledge into topic models. In *19th Pacific-Asia Conference on Knowledge Discovery and Data Mining* (**PAKDD 2015**).

Shansong Yang, Weiming Lu, Dezhi Yang, **Liang Yao**, and Baogang Wei. Short text understanding by leveraging knowledge into topic model. In *The 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (**NAACL 2015**)

Xiangzhou Huang, Yin Zhang, Baogang Wei, and **Liang Yao**. A question-answering system over Traditional Chinese Medicine. In **BIBM 2015**.

Liang Yao, Yin Zhang, Baogang Wei, Wei Wang, Yuejiao Zhang, and Xiaolin Ren. Discovering treatment pattern in traditional Chinese medicine clinical cases using topic model and domain knowledge. In **BIBM 2014**.

Liang Yao, Yin Zhang. and Baogang Wei, 2014. An Evolution System for Traditional Chinese Medicine Prescription. In *Knowledge Engineering and Management* (pp. 95-106). Springer Berlin Heidelberg.

PATENTS

Partition-based Graph Neural Networks Framework for Minor Detection (Outstanding Patents Award of Tencent)
A knowledge-guided method for multi-lingual Game User Name Generation.
A time-aware heterogeneous network embedding algorithm for recommendation.

PROFESSIONAL ACTIVITIES

Conference PC Member/Reviewer: AAAI, IJCAI, NeurIPS, ACL, EMNLP, KDD, ECML-PKDD.
Journal Reviewer: TKDE, Neural Networks, ACM Computing Surveys, JBI, etc.

RESEARCH AND ENGINEERING EXPERIENCE

Large Language Models *Feb 2023- Present*

- Introducing Contrastive Prompting (CP) which improves reasoning tasks (e.g., 35.9% to 88.8% on GSM8K)
- Fine-tuning LLaMA and ChatGLM with KG structure, SOTA results, outperforms GPT-4 and ChatGPT.
- Knowledge-enhanced LLMs (GPT-4 + KB) for question answering, No.1 in OpenBookQA challenge.
- Applications in Tencent Games.

Large-scale Graph Neural Networks (GNN) *Sep 2019- Present*

Applying GNN to Tencent Game social networks

- 0.9 billion nodes, 100 billion edges
- graph partition + subgraph GNN
- Link prediction, propose Edge Transformer and pre-training with masked auto-encoders
- applications in friend ranking, advertisement, and minor recognition.
- 10%+ improvements in many online A/B tests, 50+ times faster
- No.1 in OGB ogbl-collab challenge
- No.1 in OGB ogbl-wikikg2 challenge

BERT for Knowledge Graph (KG) Completion *July 2019- Sep 2019*

Predicting plausibility of a triple with entity/relation descriptions and pre-trained language models

- Turning KG completion into a sequence classification problem.
- state-of-the-art results in triple classification, relation prediction and link prediction.

BERT for TCM clinical records classification *Apr 2019- June 2019*

- fine-tuning BERT with unlabeled traditional Chinese medicine (TCM) clinical text。
- initializing text classifier with TCM BERT
- state-of-the-art results

Modeling text and knowledge graphs with Graph Convolutional Networks *Apr 2018- Mar 2019*

Text classification with Graph Convolutional Networks (GCN):

- Building a text graph with word co-occurrence and document word relations
- GCN for word/document embedding
- Classifying unlabeled documents nodes

Jointly learning embeddings of words and knowledge graphs with:

- Variational autoencoders
- Graph convolutional networks

Clinical text classification *Jan 2018- Apr 2018*

Obesity challenge, predicting patients' obesity and its 15 comorbidities

- Rule-based features with regular expressions.
- word embeddings and medical entity embeddings.
- Knowledge-guided convolutional neural networks (CNN) with TensorFlow.

Knowledge-based topic models *Sep 2013- Sep 2017*

Incorporating external knowledge into topic models for better topic modeling. The external knowledge is from:

- knowledge graph (e.g., Microsoft's Probase and YAGO).
- Wikipedia.
- big data (encoded by word embedding).

Knowledge discovery in traditional Chinese medicine data *Sep 2013- Dec 2017*

- Discovering patterns in Chinese medical prescriptions, recommending herbs for given symptoms.
- Discovering treatment pattern in traditional Chinese medicine clinical records.
- Classifying traditional Chinese medicine clinical records.

Purchase prediction using Tmall data *Nov 2014- Dec 2014*

- Designed Tmall specific features.
- Explored the performance of different classifiers.

Professional knowledge service system for Chinese herbal medicine *Sep 2012- Sep 2015*

- http://zcy.ckcest.cn/tcm/
- Contributed 8000 lines+ codes to the system especially in Analysis System for Medicines, Prescriptions, Diseases and Syndromes.

COMPUTER LANGUAGES AND TOOLS

Computer Languages	Python, Java, Scala, C/C++, Matlab, R, LaTeX, Javascript
Databases	MySQL, PostgreSQL, Microsoft SQL
Tools	SVN, GitHub, TensorFlow, PyTorch, Keras, Stanford CoreNLP, NLTK

AWARDS AND HONORS

World's Top 2% Scientists, by Stanford (2022-2023)

AI 2000 Most Influential Scholar Award Honorable Mention, by Tsinghua (2022-2023)

No.1 in OpenBookQA Leaderboard (2023)

No.1 in OGB ogbl-wikikg2 challenge (2023)

No.1 in OGB ogbl-collab challenge (2022)

Outstanding Contributor of Tencent (2022)

Overseas High-Caliber Personnel in Shenzhen (2020)

Oulin Scholarship (2016)

Excellent Ph.d Student Scholarship of Zhejiang University (2015, 2016)

Excellent Graduate Student of Zhejiang University (2013, 2015, 2016)

China National Scholarship for Encouragement (2010)

Outstanding Student of Sichuan University (2009)