

Jiaxin Huang

NATURAL LANGUAGE PROCESSING · DATA MINING

Siebel Center for Computer Science, 201 N. Goodwin Ave, Urbana, IL 61801

☎ (+1) 217-550-2778 | ✉ jjaxinh3@illinois.edu | 🏠 teapot123.github.io | 🎓 Google Scholar

Research Interests

My research aims to design text mining approaches for **label-efficient knowledge acquisition and utilization**, from constructing fundamental **knowledge representations for concepts, entities and documents**, to improving **complex reasoning process** for language models. Specifically, my work tackles the label scarcity issues for a variety of knowledge-intensive applications, including concept ontology construction, entity extraction, text classification and commonsense reasoning.

Education Background

University of Illinois at Urbana-Champaign

PH.D. IN COMPUTER SCIENCE

Urbana, IL, USA

2018 - 2023 (expected)

- Advisor: Jiawei Han
- Thesis Committee: Chengxiang Zhai, Tarek Abdelzaher, Jianfeng Gao

Tsinghua University

B. ENG. IN ELECTRONIC INFORMATION SCIENCE AND TECHNOLOGY

Beijing, China

2014 - 2018

- Advisor: Yong Li

Honors & Awards

2021-2023 **Microsoft Research Ph.D. Fellowship**

2022 **C.W. Gear Outstanding Graduate Award** (awarded one student for excellence in research in UIUC CS dept.)

2020, 2022 **KDD Student Travel Award**

2018 **Chirag Foundation Graduate Fellowship in Computer Science**, UIUC

2018 **Outstanding Graduates**, Tsinghua University

2015-2017 **Academic Excellence Scholarship in EE dept. Tsinghua**

2016 **China National Scholarship** (highest scholarship given by Chinese government, top 0.1%)

Publications

- (* indicates equal contribution.)
- Google Scholar: <https://scholar.google.com/citations?hl=en&user=0Ttkrj0AAAAJ>

PEER-REVIEWED CONFERENCE AND JOURNAL PUBLICATIONS

[P16] **FineSum: Target-Oriented, Fine-Grained Opinion Summarization**

Suyu Ge, **Jiaxin Huang**, Yu Meng, and Jiawei Han
WSDM 2023

[P15] **Few-Shot Fine-Grained Entity Typing with Automatic Label Interpretation and Instance Generation**

Jiaxin Huang, Yu Meng, and Jiawei Han
KDD 2022

[P14] **Topic Discovery via Latent Space Clustering of Language Model Embeddings**

Yu Meng, Yunyi Zhang, **Jiaxin Huang**, Yu Zhang and Jiawei Han
WWW 2022

[P13] **Few-Shot Named Entity Recognition: An Empirical Baseline Study**

Jiaxin Huang, Chunyuan Li, Krishan Subudhi, Damien Jose, Shobana Balakrishnan, Weizhu Chen, Baolin Peng, Jianfeng Gao and Jiawei Han
EMNLP 2021

[P12] **Distantly-Supervised Named Entity Recognition with Noise-Robust Learning and Language Model Augmented Self-Training**

Yu Meng, Yunyi Zhang, **Jiaxin Huang**, Xuan Wang, Yu Zhang, Heng Ji and Jiawei Han
EMNLP 2021

- [P11] **Weakly-Supervised Aspect-Based Sentiment Analysis via Joint Aspect-Sentiment Topic Embedding**
Jiaxin Huang, Yu Meng, Fang Guo, Heng Ji and Jiawei Han
EMNLP 2020
- [P10] **Text classification using label names only: A language model self-training approach**
Yu Meng, Yunyi Zhang, **Jiaxin Huang**, Chenyan Xiong, Heng Ji, Chao Zhang and Jiawei Han
EMNLP 2020
- [P9] **CoRel: Seed-Guided Topical Taxonomy Construction by Concept Learning and Relation Transferring**
Jiaxin Huang, Yiqing Xie, Yu Meng, Yunyi Zhang and Jiawei Han
KDD 2020
- [P8] **Hierarchical Topic Mining via Joint Spherical Tree and Text Embedding**
Yu Meng, Yunyi Zhang, Jiaxin Huang, Yu Zhang, Chao Zhang and Jiawei Han
KDD 2020
- [P7] **Minimally Supervised Categorization of Text with Metadata**
Yu Zhang, Yu Meng, **Jiaxin Huang**, Frank F. Xu, Xuan Wang and Jiawei Han
SIGIR 2020
- [P6] **Discriminative Topic Mining via Category-Name Guided Text Embedding**
Yu Meng*, **Jiaxin Huang***, Guangyuan Wang, Zihan Wang, Chao Zhang, Yu Zhang and Jiawei Han
WWW 2020
- [P5] **Guiding Corpus-based Set Expansion by Auxiliary Sets Generation and Co-Expansion**
Jiaxin Huang*, Yiqing Xie*, Yu Meng, Jiaming Shen, Yunyi Zhang and Jiawei Han
WWW 2020
- [P4] **Spherical Text Embedding**
Yu Meng, **Jiaxin Huang**, Guangyuan Wang, Chao Zhang, Honglei Zhuang, Lance Kaplan and Jiawei Han
NeurIPS 2019
- [P3] **Unsupervised Word Embedding Learning by Incorporating Local and Global Contexts**
Yu Meng, **Jiaxin Huang**, Guangyuan Wang, Zihan Wang, Chao Zhang, Jiawei Han
Frontiers in Big Data 2020
- [P2] **Understanding Motivations behind Inaccurate Check-ins**
Fengli Xu, Guozhen Zhang, Zhilong Chen, **Jiaxin Huang**, Yong Li, Diyi Yang, Ben Y. Zhao, Fanchao Meng
CSCW 2018
- [P1] **Big Data Driven Mobile Traffic Understanding and Forecasting: A Time Series Approach**
Fengli Xu, Yujun Lin, **Jiaxin Huang**, Yong Li
IEEE Transactions on Services Computing 2016

MANUSCRIPTS AND PRE-PRINTS

- [P3] **Large Language Models Can Self-Improve**
Jiaxin Huang, Shane Shixiang Gu, Le Hou, Yuexin Wu, Hongkun Yu, Jiawei Han
2022
- [P2] **Tuning Language Models as Training Data Generators for Augmentation-Enhanced Few-Shot Learning**
Yu Meng, Martin Michalski, **Jiaxin Huang**, Yu Zhang, Tarek Abdelzaher, Jiawei Han
2022
- [P1] **All Birds with One Stone: Multi-task Text Classification for Efficient Inference with One Forward Pass**
Jiaxin Huang, Tianqi Liu, Jialu Liu, Adam D. Lelkes, Cong Yu, Jiawei Han
2022

Internship Experience

Google Core-ML

RESEARCH INTERN

May. 2022 - Aug. 2022

- Hosts: Le Hou and Yuexin Wu
- Project: Reasoning with large language models through chain-of-thought self-improving

Google Research

RESEARCH INTERN

May. 2021 - Aug. 2021

- Hosts: Tianqi Liu and Jialu Liu
- Project: Efficient Multi-task Learning for Pre-trained Language Models

Microsoft Research

RESEARCH INTERN

Jun. 2020 - Sept. 2020

- Host: Chunyuan Li
- Project: Few-shot Named Entity Recognition

Invited Talks

CONFERENCE TUTORIALS

- [T5] **Adapting Pretrained Representations for Text Mining**
In KDD 2022. Instructors: Yu Meng, Jiaxin Huang, Yu Zhang, Jiawei Han
- [T4] **Pre-Trained Language Representations for Text Mining**
In AAAI 2022. Instructors: Yu Meng, Jiaxin Huang, Yu Zhang, Jiawei Han
- [T3] **On the Power of Pre-Trained Text Representations: Models and Applications in Text Mining**
In KDD 2021. Instructors: Yu Meng, Jiaxin Huang, Yu Zhang, Jiawei Han
- [T2] **Embedding-Driven Multi-Dimensional Topic Mining and Text Analysis**
In KDD 2020. Instructors: Yu Meng, Jiaxin Huang, Jiawei Han
- [T1] **TextCube: Automated Construction and Multidimensional Exploration**
In VLDB 2019. Instructors: Yu Meng, Jiaxin Huang, Jingbo Shang, and Jiawei Han

INVITED SEMINARS

- [T8] **From Pre-trained Language Representations to Human Actionable Knowledge**
Symposium on Frontiers of ML and AI, USC, Nov 2022
- [T7] **Self-Improving Reasoning Ability of Large Language Models**
Google Brain, Sept 2022
- [T6] **Efficient Multi-task fine-tuning for Pre-trained Language Models**
Google Research, Aug 2021
- [T5] **Few-shot Named Entity Recognition**
Microsoft Research, Aug 2020
- [T4] **Controlled Language Generation**
UIUC Data Mining Seminar, Mar 2020
- [T3] **Multi-Dimensional Cube-based Analysis**
ARL Report, Mar 2020
- [T2] **Topical Taxonomy Construction**
DARPA KAIROS Site Visit Report, Mar 2020
- [T1] **Towards Understanding Word Embeddings**
UIUC Data Mining Seminar, Oct 2019

Teaching & Mentoring

TEACHING ASSISTANT

Spring 2019 CS 412 (Introduction to Data Mining), UIUC

MENTORED STUDENTS

Suyu Ge , 2021-Present	BS student at Tsinghua EE → PhD student at UIUC CS
Yunyi Zhang , 2020-Present	MS → PhD student at UIUC CS
Yiqing Xie , 2019-2021	MS student at UIUC CS → PhD student at CMU LTI
Fang Guo , 2019-2021	MS student at UIUC CS → PhD student at Westlake University
Zihan Wang , 2019-2020	BS student at UIUC CS → PhD student at UCSD CS

References

Jiawei Han (Ph.D. Advisor)

MICHAEL AIKEN CHAIR PROFESSOR

- Computer Science Department, University of Illinois at Urbana-Champaign
- hanj@illinois.edu

Chengxiang Zhai

DONALD BIGGAR WILLETT PROFESSOR

- Computer Science Department, University of Illinois at Urbana-Champaign
- czhai@illinois.edu

Jianfeng Gao

DISTINGUISHED SCIENTIST & VICE PRESIDENT

- Microsoft Resesarch
- jfgao@microsoft.com

Chunyuan Li

PRINCIPAL RESEARCHER

- Microsoft Resesarch
- Chunyuan.Li@microsoft.com