Haopeng Zhang

Email: haopeng@ifmlab.org
Tel: 404-376-1886
Website: hpzhang94.github.io
Google Scholar

Research Interests

- **Natural Language Processing**
  Graph4NLP, Extractive Text Summarization, Large Language Model
- **Graph Mining**
  Graph Neural Network

Education

University of California, Davis Expected: Jun. 2024
Ph.D. Candidate in Computer Science Advisor: Dr. Jiawei Zhang
Transferred from Florida State University (FSU)
Thesis: Leveraging structures for natural language processing.

Georgia Institute of Technology Dec. 2018
M.S. in Computational Science and Engineering
M.S. in Electrical and Computer Engineering

University of Illinois, Urbana and Champaign (UIUC) May 2016
B.E. in Electrical Engineering

Professional Experience

Amazon Web Services (AWS) AI Lab Santa Clara, CA
Applied Scientist Intern (Mentors: Dr. Lan Liu, Dr. Peng Qi, Dr. Zhiheng Huang) Summer 2023
Worked on exploring different optimization strategies of retrieval-augmented generation systems for large-scale industry search engines.

Megagon Lab Mountain View, CA
Research Intern (Mentors: Dr. Hayate Iso, Dr. Sairam Gurajada) Spring 2023
Proposed and created the first large-scale benchmark for explainable fine-grained text editing task. The research paper is under review.

Tencent AI Lab Seattle, WA
Research Intern (Mentors: Dr. Sangwoo Cho, Dr. Kaiqiang Song) Summer 2022
Worked on graph-based unsupervised multi-document summarization; proposed a holistic framework that balances summary salience and diversity. The research paper is under review.

Salesforce Research Remote
Research Intern (Mentors: Dr. Semih Yavuz, Dr. Yingbo Zhou) Summer 2021
Proposed a new approach to address the entity-level summary hallucination problem by controlling entity coverage and domain transfer for abstractive summarization with intermediate data. The research paper is published in NAACL 2022.

Kidswant Company Nanjing
Data Mining Research Intern (Mentor: Dr. Hang Liu) Summer 2018
Worked on extracting lifecycle rules for different classes of products and improved the recommendation system; Developed tool with Django and Pyecharts for dynamic data visualization.

Publications

[1] **H. Zhang**, X. Liu, and J. Zhang, "Extractive summarization via chatgpt for faithful summary generation", *Foundations and Applications in Large-scale AI Models Workshop at KDD 2023 conference (**KDD 2023 LLM4AI**)*, 2023.

[2] **H. Zhang**, X. Liu, and J. Zhang, "DiffuSum: Generation enhanced extractive summarization with diffusion", in *Findings of the Association for Computational Linguistics: **ACL 2023***, Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 13 089–13 100.

[3] **H. Zhang**, X. Liu, and J. Zhang, "Contrastive hierarchical discourse graph for scientific document summarization", in *Proceedings of the 4th Workshop on Computational Approaches to Discourse (**ACL CODI 2023**)*, Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 37–47.

[4] **H. Zhang**, X. Liu, and J. Zhang, "HEGEL: Hypergraph transformer for long document summarization", in *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (**EMNLP 2022**)*, Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, Dec. 2022, pp. 10 167–10 176.

[5] **H. Zhang**, S. Yavuz, W. Kryscinski, K. Hashimoto, and Y. Zhou, "Improving the faithfulness of abstractive summarization via entity coverage control", in *Findings of the Association for Computational Linguistics: **NAACL 2022***, Seattle, United States: Association for Computational Linguistics, Jul. 2022, pp. 528–535.

[6] **H. Zhang** and J. Zhang, "Text graph transformer for document classification", in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (**EMNLP 2020**)*, Online: Association for Computational Linguistics, Nov. 2020, pp. 8322–8327.

[7] Z. Li, **H. Zhang**, and J. Zhang, "A revisit of fake news dataset with augmented fact-checking by chatgpt", under review, 2023.

[8] **H. Zhang**, H. Iso, S. Gurajada, and N. Bhutani, "Xatu: A fine-grained instruction-based benchmark for explainable text updates", under review, 2023.

[9] **H. Zhang**, X. Liu, and J. Zhang, "Summit: Iterative text summarization via chatgpt", *arXiv preprint arXiv:2305.14835*, 2023.

[10] **H. Zhang**, S. Cho, K. Song, X. Wang, J. Zhang, and Y. Dong, "Unsupervised multi-document summarization with holistic inference", under review, 2022.

[11] **H. Zhang** and J. Zhang, "Centrality meets centroid: A graph-based approach for unsupervised document summarization", *arXiv preprint arXiv:2103.15327*, 2021.

[12] J. Zhang, **H. Zhang**, L. Sun, and C. Xia, "Graph-bert: Only attention is needed for learning graph representations", *arXiv preprint arXiv:2001.05140*, 2020.

Scholarships and Awards

- UC Davis Graduate Group in Computer Science (GGCS) Research Fellowship 2023
- FSU Travel Award 2020
- FSU Adelaide D. Wilson Graduate Fellowship Endowment Fund 2019
- UIUC Russell E. Berthold Scholarship 2015

Service

- **Journal Reviewer**

  IEEE Transactions on Big Data
  IEEE/ACM Transactions on Audio, Speech, and Language Processing
  Artificial Intelligence Review
  Computer Speech & Language

- **Conference Reviewer**

  EMNLP 23, IJCNN 23, ACL 23, EACL 23, KDD 23, AAAI 22/21

- **Program Committee**

  The 4th New Frontiers in Summarization Workshop @EMNLP 2023

Teaching Experience

- **Teaching Assistant@UC Davis**

  Deep Learning (ECS189G) Winter 2022

- **Teaching Assistant@FSU**

  Advanced Database Systems (COP5725) Fall 2021, Spring 2021

  Complexity and Analysis of Data Structures and Algorithms (COP4531) Spring 2020

  Theory and Structure of Databases (COP4710) Fall 2020, Fall 2019

- **Teaching Assistant@Georgia Tech**

  Semiconductor Devices (ECE3450) Fall 2017, Spring 2017

Mentorship

- **Xiao Liu** (M.S. student → Ph.D. student, UC Davis)

  I have been mentoring Xiao for two years since he was a M.S. student and working with him through the Ph.D. program. I guided and helped him in several research projects on extractive text summarization, structural document representation, and large language model applications. We collaborated and published four papers in top-tier NLP conferences and workshops.

- **Zizhong Li** (Ph.D. student, UC Davis)

  I have been mentoring Zizhong for a year on her research projects involving fake news detection with large language models and retrieval augmented generation. I helped and provided suggestions on the designs of projects and experiments. Our collaborative research paper is under review.