Haoyuan Li

https://leehaoyuan.github.io/ | haoyuanl@cs.unc.edu |
https://scholar.google.com/citations?hl=en&user=BdfQ44MAAAAJ

EDUCATION

Renmin University of China, Beijing, CN Sep.2016–July.2020
Major: Statistics, Bachelor of Science, School of Statistics
- **Cumulative GPA:** 3.74/4.0

Harvard University, Cambridge, US Sep.2020-Jan.2022
Major: Health Data Science, Master of Science, School of Public Health
- **Cumulative GPA:** 3.79/4.0

University of North Carolina at Chapel Hill, Chapel Hill, US Sep.2022-Present
Major: Computer Science, Doctor of Philosophy, Department of Computer Science
- **Advisor: Snigdha Chaturvedi**

INTERNSHIP

Stanford University Jun.2019-Sep.2019
- **Position:** Research Assistant **Advisor:** Ram Rajagopal

Yale University Jun.2020-Sep.2020
- **Position:** Research Assistant **Advisor:** Dragomir R. Radev

Microsoft Research Asia Jan.2021-July.2021
- **Position:** Research Assistant **Advisor:** He Di

Dana-Farber Cancer Institute Sep.2021-Dec.2021
- **Position:** Data Science Trainee **Advisor:** Jacob Rosenthal

PUBLICATIONS

Wang, Z.*, **Haoyuan Li***, Rajagopal, R., (2019) Urban2Vec: Incorporating Street View Imagery and POIs for Multi-Modal Urban Neighborhood Embedding, AAAI 2020, (*contribute equally)

Fabbri, Alexander R., Simeng Han, **Haoyuan Li**, Haoran Li, Marjan Ghazvininejad, Shafiq Joty, Dragomir Radev, and Yashar Mehdad. "Improving Zero and Few-Shot Abstractive Summarization with Intermediate Fine-tuning and Data Augmentation.", NAACL 2021

Bi, Qiwei, **Haoyuan Li**, and Hanfang Yang. 2021. "Boosting Few-Shot Abstractive Summarization with Auxiliary Tasks.", CIKM 2021

Qiwei Bi, **Haoyuan Li**, Kun Lu, Hanfang Yang, Augmented Abstractive Summarization with Document-Level Semantic Graph, J. data sci. 19(2021)

Haoyuan Li, Somnath Basu Roy Chowdhury, Snigdha Chaturvedi, 'Aspect-aware Unsupervised Extractive Opinion Summarization', ACL 2023 Findings

Haoyuan Li, Snigdha Chaturvedi, 'Rationale-based Opinion Summarization', NAACL 2024

Haoyuan Li, Yusen Zhang, Rui Zhang, Snigdha Chaturvedi, 'Coverage-based Fairness in Multi-document Summarization', submitted to NAACL 2025

REASEARCH & PROJECTS

Urban2Vec: Incorporating Street View Imagery and POIs for Multi-Modal Urban Neighborhood Embedding
Advisor: Ram Rajagopal, Stanford University
Summary: We develop unsupervised methods to generate the embedding of neighborhoods based on the street view image and POIs. In the project, I explore the way to combine image and text and tune parameters and evaluate its performance.

Augmented Abstractive Summarization with Document-Level Semantic Graph
Advisor: Hanfang Yang, Asso Prof. School of Statistics, Renmin University of China
Summary: We generate a graph of entities in each document using distant supervision. Then, we apply GNN to generate the entity embedding and insert an attention sub-layer in each decoder layer to incorporate entity embedding into generation. In the project, I tune the hyper-parameter and explore the variant of generating the graph.

Improving Zero and Few-Shot Abstractive Summarization with Intermediate Fine-tuning and Data Augmentation
Advisor: Dragomir R. Radev, Yale University
Summary: We generate pseudo training pair from the wiki dataset based on the characteristics of the target dataset. We also introduce back-translation and consistency loss to further improve the performance. In the project, I explore the possibility of using consistency loss between original documents and augmented documents given pseudo-target summaries.

Boosting few-shot abstractive summarization with auxiliary tasks
Advisor: Hanfang Yang, Renmin University of China
Summary: We develop 3 auxiliary tasks, sentence extraction, object prediction and triple entailment and trained the abstractive summarization model with these tasks in a multi-task learning framework. In this project, I help build BART model for abstractive summarization, introduce adapter for multi-task learning, run experiment on xsum and write the paper.

Aspect-aware Unsupervised Extractive Opinion Summarization
Advisor: Snigdha Chaturvedi, University of North Carolina at Chapel Hill
Summary: We introduce a method for unsupervised extractive opinion summarization that automatically identifies the aspects described in the review sentences and then extracts sentences based on their aspects. It identifies the underlying aspects of the review sentences using the roots of noun phrases and adjectives appearing in them.

Rationale-based Opinion Summarization

Advisor: Snigdha Chaturvedi, University of North Carolina at Chapel Hill
Summary: We propose a new paradigm for opinion summarization, where each represented opinion is paired with a rationale that provides more details. To extract rationales for a representative opinion, we extract sentences based on 4 criteria: relatedness, specificity, popularity and diversity.

Coverage-based Fairness in Multi-Document
Advisor: Snigdha Chaturvedi, University of North Carolina at Chapel Hill
Summary: We propose a new summary-level fairness measure, Equal Coverage, which is based on coverage of documents with different social attribute values and considers the redundancy within documents. Equal Coverage addresses the redundancy issue faced by previous measures like Proportional Representation. We also propose a new corpus-level measure, Coverage Parity.

HONORS

Scholarship for Academic Achievement, 2017,2019(**Top 20%**), Prize of National English Contest of University Students,2017(**Top 1%**), First Prize of China Undergraduate Mathematical Contest in Modeling, 2018(**Top 1%**)