

KANGDA WEI

2701 Homestead Rd, APT 901, Chapel-Hill, NC, 27516
(919) 904-5280 ◇ kangda@live.unc.edu ◇ hansonwei666@gmail.com

RESEARCH INTEREST

Natural Language Processing, Machine Learning, Deep Learning

EDUCATION

The University of North Carolina at Chapel-Hill

Master of Science in Computer Science

Chapel-Hill, NC

Aug 2022 - Present

- **Graduate Teaching Assistant:** Hold office hour for COMP431 Internet Service and Protocol. Grade assignments and exams. Write auto-grader for grading programming assignments. Help course instructor organize course materials and agenda.

The University of North Carolina at Chapel-Hill

B.S. in Computer Science, B.S. in Statistics and Operational Research

Chapel-Hill, NC

Aug 2019 - May 2022

- **GPA:** 3.812/4.0
- **Relevant coursework:** Foundation of Programming, Data Structures, Algorithms and Analysis, Computational Photograph, Optimization, Structured Prediction, Machine Learning & Deep Learning, Programming Language, Time Series, Simulation Analysis, Stochastic Modeling, Linear Algebra, Computer Organization

Online courses

- Stanford CS224n (NLP), Machine Learning, Deep Learning Specialization Series taught by Andrew Ng on Coursera

RESEARCH EXPERIENCE

Prompting Multi-hop Question Answering Task With Pre-trained Language Model

Johns Hopkins University

Baltimore, MD

Visiting Research Assistant, Mentor: Prof. Benjamin van Durme

May 2022 - present

- Propose a new training and testing procedure that incorporate hard and soft prompts using T-5 Model. Train and test the model on Multi-hop Question Answering task, BREAK.
- Improve the QA performance compared to prior work: 5% higher in Exact Match score and 10% higher in F1 score.
- Paper in preparation.

Learning Task-Specific Zero-Shot Classifiers through Data Programming

The University of North Carolina at Chapel Hill

Chapel Hill, NC

Research Assistant, Mentor: Prof. Shashank Srivastava

April 2022 - present

- Present an unsupervised learning system that combines pre-trained language model (LM) with data programming technique to generate labels for unlabeled data under a zero-shot learning setting.
- Achieved 8% higher classification accuracy by utilizing label aggregation with LM, and three times better accuracy comparing to baselines without zero-shot learning.
- Simplify and improve the label aggregation technique by replacing hand-written labeling functions with LM.
- Paper in preparation.

Compositional Generalization for Kinship Prediction through Data Augmentation

The University of North Carolina at Chapel Hill

Chapel Hill, NC

Research Assistant, Mentor: Prof. Shashank Srivastava

February 2021 - March 2022

- Evaluated empirically the utility of data augmentation and intermediate structured representations towards compositional generalization for the task of kinship prediction from a story.
- Boosted generalization performance by around 20% on average relative to the baseline model without data augmentation. Found that using intermediate kinship graphs led to a deterioration in the models generalization ability.
- Tested the impact of incorporating data augmentation and intermediate structured data on model's performance. Data augmentation boosted generalization performance by around 20% on average relative to a baseline model from prior work. Predicting and using intermediate kinship graphs led to a deterioration in the generalization of kinship prediction.

- Published paper at *Proceedings of the 4th Workshop of Narrative Understanding (WNU2022)*.

A Multilingual COVID-19 Question Answering System

University of California, Santa Barbara

Visiting Research Assistant, Mentor: Prof. William Wang

Santa Barbara, CA

May 2021 - August 2021

- Established a multilingual COVID-19 Question Answering system using mBERT and XLM-Roberta models.
- Focused primarily on building and training the Reading Comprehension part of the QA system, including collecting and processing the data, and performing machine translation on large-scale natural language dataset.
- Performed large-scale neural machine translation on entire COVID-19 dataset to acquire data in Chinese and French to alleviate the data scarcity problem in foreign languages.
- Reached an F1 score of 60.5 for the reading comprehension model of the final QA system.

Temporal and Spatial Analysis on U.S. Domestic Mobility

Research Assistant

Mentor: Prof. Weizi Li at the University of Memphis, Prof. Lei Lin at University of Rochester

Chapel Hill, NC

June 2020 - October 2020

- Investigated the impact of COVID-19 on the United States mobility through temporal and spatial analysis with network models.
- Collected relative datasets from GitHub, Kaggle, New York Times and Google Search Trends; Preprocessed data with Python and implemented the model with Networkx package.
- Published paper in the 20th and 21st Joint COTA International Conference of Transportation Professionals.

PUBLICATIONS

- **Kangda Wei**, Orion Weller, Yunmo Chen, Benjamin van Durme. Prompting Multi-hop Question Answering Task With Pre-trained Language Model. *in preparation*
- **Kangda Wei**, Sayan Ghosh, Rakesh Menon, and Shashank Srivastava. Learning Task-Specific Zero Shot Classifiers through Data Programming using Explanations. *in preparation*
- **Kangda Wei**, Sayan Ghosh, and Shashank Srivastava. 2022. Compositional Generalization for Kinship Prediction through Data Augmentation. *In Proceedings of the 4th Workshop of Narrative Understanding (WNU2022)*, pages 1319, Seattle, United States. Association for Computational Linguistics.
- Songhe Wang, ***Kangda Wei**, Lei Lin, Weizi Li. Spatial-temporal Analysis of COVID-19's Impact on Human Mobility: the Case of the United States, in the *20th and 21st Joint COTA International Conference of Transportation Professionals*. *Co-author: equal contribution

NOTABLE PROJECTS

Email Plug-in Startup

January 2022 - May 2022

- Built a classification model from scratch for email classification using PyTorch and Huggingface Transformer.
- Increased the model accuracy from 0.32 to 0.84 with limited amount of annotated data.

Cyber-infrastructure: Web-based Application for Sharing Neural Data

September 2020 - January 2021

- Built a website where users can register and login to access the neural data of patients.
- Practiced the basis of front-end and back-end engineering using Python Flask and Firebase.

Mathematical Contest in Modeling (MCM)

February 2020 & February 2021

- Implemented the majority of my team's solution. Processed and analyzed given data via R and MATLAB, and presented the mathematical model in the comprehensible report for the 2020 MCM contest.
- Utilized ResNet to preprocess image data and implement Neural Network to classify the true and false Asian Giant Hornet report for the 2021 MCM contest.

SKILLS

- **Programming Languages/Frameworks:** Python, Java, JavaScript, C, HTML, CSS, MATLAB, R, PyTorch, Linux
- **Software & Tools:** PyCharm, VS Code, Spyder, Jupyter Notebook, Tableau, Latex, Anaconda, RStudio
- **Language:** Chinese (native), English (proficient)