Hongchao Fang

535 Pontius Ave N, Seattle, WA, 98109
Mobile: (425)245-6502

LinkedIn: Hongchao Fang
Google Scholar: Hongchao Fang
Email: colefang.hongchao@gmail.com

Education

- **Northeastern University** — Seattle, US
  M.S. in Computer Science ; GPA: 3.94 — *Sep 2021 - Dec 2023*
- **Central University of Finance and Economics** — Beijing, China
  B.S. in Information Security — *Sep 2017 - Jun 2021*

Selected Work Experience

- **Paul G. Allen Center, UW** — Seattle, WA
  Research Assistant (Natural Language Processing) — *Jan 2023 - Present*
  - Applied contrastive learning method including Byol, MOCO, Simclr on few-shot and zero-shot tasks.
  - Analysed the meaning of self-supervised learning for text classification tasks with inadequate training sentences.
  - Proposed new self-supervised frame for prompt-based tuning which improves the few-shot performance on GLUE benchmark by 5%.
- **Amazon** — Seattle, WA
  Software Engineer Intern — *May 2022 - Aug 2022*
  - Designed and implemented a Java API for sellers to fetch billing and invoice information from **DynamoDB**.
  - Optimized frond-end **Ember.JS** UI to show more detailed invoices information.
  - Designed for a new pub/sub system for users to get real-time invoice change using native **AWS** tools.
  - Scalied the search API to handle 200% more queries under real-time load tests and query events.
- **AI4H Lab, UCSD** — Remote
  Research Assistant (Natural Language Processing) — *Apr 2020 - Oct 2020*
  - Established a large-scale medical dialogue dataset: MedDialog with 3.4 million conversations between patients and doctors, 11.3 million utterances, 660.2 million tokens, covering 172 specialties of diseases, which is the largest medical dialogue dataset so far .
  - Proposed a special self-supervised method to solve the common problem that large pre-trained language models cannot perfectly represent the features of new sentences.
  - Applied our special method to different language models including BERT, ERNIE, and Robert. The results show that our model brings about 3% improvements to various language models today.
- **Chinese Academy of Sciences** — Beijing, China
  Research Assistant (Computer Vision) — *Apr 2019 - Aug 2021*
  - Trained appropriate image classification models (ResNet, AlexNet, InceptionV3) on our own dataset, and used data augmentation to improve the accuracy from 0.84 to 0.91.
  - Optimized the image classification algorithm and saved the time and energy costs on mobile devices by 50%.
  - Implemented the function that mobile devices can automatically generate different classification models adapting to their environment, under the instructions of the server.

Selected project

- **Order Data Analysis System** — Beijing, China
  Large Scale Distributed Systems, Apache Hadoop, Apache Spark — *Apr 2020 - Jul 2020*
  - Crawled the data of orders from Taobao using**Requests** and **BeautifulSoup**.
  - Analysed the preference of different people from orders data in real-time with **Spark**.
  - Saved the results and data on **HDFS** for future analysis using **Hadoop**.

Publications

- **CERT: Contrastive Self-supervised Learning for Language Understanding**
  (accepted by TACL)(210 citations)
- **MedDialog: Large-scale Medical Dialogue Datasets**
  (accepted by EMNLP)(75 citations)

Honors & Awards

- **Academic Excellence Award at Central University of Finance and Economics** — Beijing, China
  undergraduate student — *Oct 2018*

Skills Summary

- **Programming Languages** : Python, Java, C++, C, JavaScript, SQL, Shell
- **Back-end Frameworks**: Django, Spring, Spring MVC, Spring Boot, Node.js
- **Front-end Technologies**: HTML, CSS, React.js, Angular.js
- **Database Technologies**: MySQL, Redis, DynamoDB, MongoDB
- **Platform & Tools**: AWS, Git, Docker, Kubernetes, Maven, Nginx
- **Big Data Technologies**: Hadoop, HBase, Spark, MapReduce, Hive, Zookeeper
- **Machine learning Technologies**: Tensorflow, Pytorch, OpenCV, sklearn, pandas, karas