PEI CHEN

https://brickee.github.io/
+1 979-739-4507 ⋄ chenpei.net@gmail.com

EDUCATION

Ph.D. in Computer Science 2019.8 - 2024.5
· Research Areas & Interests: Natural Language Processing (NLP); Large Language Models (LLMs) Pretraining, (Parameter-efficient) Fine-tuning, Efficient Inference, and Prompting; Information Extraction including Named Entity Recognition, Event Extraction, and Knowledge Base Completion.
· Overall GPA: 4.0/4.0.
Texas A&M University

MS in Finance 2016.9 – 2018.6
· Thesis: Does News Sentiment Predict the Stock Market? An Example on Chinese Growth Market;
· Received 2017 National Scholarship for Graduate Student;
· Overall GPA: 3.9/5.0, ranking 1/178.
Southwestern University of Finance and Economics

B.Engr. in Simulation Engineering 2010.9 – 2014.6
· Thesis: Analyze and reconstruct the multi-resolution modeling technology of a simulation system written by millions of lines of C++ code;
· Overall GPA 88.61/100, ranking 1/45.
National University of Defense Technology

EXPERIENCE

Stores Foundational AI Team, Amazon 2024.1 – now
Applied Scientist (full-time) *Palo Alto, CA*

· Doing research and applications of LLMs in helping Amazon's customers.

Open Source AI Team, AWS AI 2023.5 – 2023.8
Applied Scientist (Intern) *Santa Clara, CA*

· Doing research on designing cutting-edge multi-agent prompting methods for LLMs;
. Proposed using collaborative multi-agent prompting methods for LLMs, achieving new SOTA performance for complex science problems on both zero-shot settings and few-shot settings. The related **paper is accepted to NAACL 2024.**

Bedrock Team (Amazon Titan Model), AWS AI 2022.6 – 2023.1
Applied Scientist (Intern) *Santa Clara, CA*

· Proposed a novel tabular language model that models tables as hypergraphs to capture the table structures, achieving new SOTA tabular data representation learning performance (accepted to **NeurIPS** 2023 as a **spotlight**);
· Scale the preprocessing of 27 million tabular data for pretraining using parallel multiprocessing;
· Successfully pretrained the proposed model with the large-scale tabular data on AWS EKS clusters with multiple machines and GPUs.

Tencent AI Lab 2021.6 – 2021.8
NLP Researcher (Intern) *office in Seattle, WA, remote from College Station, TX*

Proposed a comprehensive benchmark for zero-shot knowledge base completion (KBC) tasks, covering state-of-the-art KBC methods and broad knowledge source data.

Department of Computer Science & Engineering, Texas A&M University 2019.9 – 2023.12
Research Assistant and Teaching Assistant *College Station, TX*

Research: Improved domain-specific named entity recognition task by modeling non-sequential entity mention relations using Graph Neural Networks; Improved fine-grained opinion mining task; Built fine-grained named location recognition benchmark, etc.
Teaching: CSCE 636 Deep Learning; CSCE 313 Computer Systems

National Lab of Pattern Recognition, Chinese Academy of Sciences 2018.1 – 2019.8
Research Engineer (full-time) *Beijing, China*

Improved the event extraction and causality detection tasks from financial domain texts.

Innovation Lab of Global Exchange, State Street 2017.7- 2018.1
Data Analyst (Intern) *Hangzhou, China*

Working on data cleaning, analysis, visualization and database construction for innovative financial applications.

PUBLICATIONS

"Don't Know How to Express Emotion? LAMB - A Learning Augmented Emotion Booster for Your Plain Text", 3rd co-author, submitted to ACL 2024.

Ming Li, **Pei Chen**, Chenguang Wang, Hongyu Zhao, Yijun Liang, Yupeng Hou, Fuxiao Liu, Tianyi Zhou. "Mosaic IT: Enhancing Instruction Tuning with Data Mosaics", submitted to NeurIPS 2024.

Wangtao Sun, Haotian Xu, Xuanqing Yu, **Pei Chen**, Shizhu He, Jun Zhao, Kang Liu. "ItD: Large Language Models Can Teach Themselves Induction through Deduction", accepted to ACL 2024.

Yuling Xia, Teague McCracken, Tong Liu, **Pei Chen**, Andrew Metcalf, Chao Fan. "Understanding Disparities of PM2.5 Air Pollution in Urban Areas via Deep Support Vector Regression.", Environmental Science and Technology, 2024 May 2.

Pei Chen, Shuai Zhang, Boran Han. "CoMM: Collaborative Multi-Agent, Multi-Reasoning-Path Prompting for Complex Problem Solving.", accepted to NAACL 2024.

Teague McCracken, **Pei Chen**, Andrew Metcalf, Chao Fan. "Quantifying the Impacts of Canadian Wildfires on Regional Air Pollution Networks.", Science of The Total Environment. 2024 Apr 12:172461.

Pei Chen, Haibo Ding, Jun Araki, and Ruihong Huang. "DOMAIN-SPECIFIC NAMED ENTITY RECOGNITION VIA GRAPH NEURAL NETWORKS." U.S. Patent Application 17/877,543, filed February 1, 2024.

Pei Chen, Soumajyoti Sarkar, Leonard Lausen, Balasubramaniam Srinivasan, Sheng Zha, Ruihong Huang, and George Karypis. "HYTREL: Hypergraph-enhanced Tabular Data Representation Learning.", NeurIPS 2023 (**spotlight**), acceptance rate: 5%.

Pei Chen, Wenlin Yao, Hongming Zhang, Xiaoman Pan, Dian Yu, Dong Yu, and Jianshu Chen. "ZeroKBC: A Comprehensive Benchmark for Zero-Shot Knowledge Base Completion." ICDM-2022, KG workshop.

Pei Chen, Haotian Xu, Cheng Zhang, and Ruihong Huang. "Crossroads, Buildings and Neighborhoods: a Dataset for Fine-grained Location Recognition". NAACL-2022, long paper, acceptance rate: 21.96%.

Pei Chen, Haibo Ding, Jun Araki, and Ruihong Huang. "Explicitly Capturing Relations between Entity Mentions via Graph Neural Networks for Domain-specific Named Entity Recognition." ACL-2021, short paper, acceptance rate: 21.2%.

Pei Chen, Kang Liu, Yubo Chen, Taifeng Wang, and Jun Zhao. "Probing into the Root: A Dataset for Reason Extraction of Structural Events from Financial Documents." EACL-2021, short paper, acceptance rate: 24.7%.

Pei Chen, Hang Yang, Kang Liu, Ruihong Huang, Yubo Chen, Taifeng Wang, and Jun Zhao. "Reconstructing Event Regions for Event Extraction via Graph Attention Networks." AACL-2020, long paper, acceptance rate: 28.3%.

SKILLS

Competent: Python, PyTorch, PyTorch Lightning, Git, Docker

Familiar: C/C++, TensorFlow, Keras, SQL

Experienced: Parallel pretraining and fine-tuning with multiple GPUs and machines.

PROFESSIONAL SERVICE

2024: Program Committee/Reviewer for ACL Rolling Review

2023: Program Committee/Reviewer for ACL, EMNLP, ACL Rolling Review

2022: Program Committee/Reviewer for EMNLP, ACL Rolling Review, NLPCC

2021: Program Committee/Reviewer for EMNLP, ACL Rolling Review, NLPCC