Mingchen Li

Researcher, Division of Computational Health Sciences, University of Minnesota Twin Cities

Email: li003378@umn.edu
Mobile: 470-815-7917

Please Click My Blog
Please Click My Github

SUMMARY

My primary research focus on natural language processing and data mining, encompassing key areas such as **Language Model**, **Knowledge Graph**, **Information Extraction**, **Question Answering**, **Text Classification**, and **Link Prediction**.

Education

- **Georgia State University (GSU)** — ATL, US
  M.S. in Computer Science — *01.2021-12.2022*
- **Qufu Normal University (QFNU)** — Shandong, China
  B.S. in Internet of Things — *08.2015-06.2019*

Experience and Projects

- **University of Minnesota Twin Cities** — 01.2022-Present
  Researcher
  - **Outcome1**: Developed a **Retrieval-based Large Language Model** with Chain-of-Thought, **RT**, for biomedical name entity recognition.
  - **Outcome2**: Developed an enhanced **Retrieval-based Large Language Model** utilizing the Tailored Chunk Scorer, **PETAILOR**, for Biomedical Triple Extraction. In this study, PETAILOR stands as the pioneering model to integrate external knowledge (such as chunked relation descriptions and relation types) into a Language Model (LM). A customized chunk scorer is introduced to adapt the LM's requirements by utilizing the LM output as a signal. Additionally, a biomedical triple extraction dataset, GM-CIHT, is introduced, featuring high-quality annotations and a comprehensive range of relation types.
  - **Ongoing**: **Graph with Large Language model**, Retrieval-based Language Model
- **Georgia state university-ARCTIC** — 08.2022-12.2022
  Research Assistant
  - **Outcome**: Developed a **spatio-temporal model** employing an **attention network** for predicting COVID-19 pandemic trends. Integrated graph embedding into a gated recurrent unit (GRU) network to capture temporal features.
- **Prescio Consulting** — 06.2022-08.2022
  Internship
  - **Outcome**: This project aims to instruct the client in comprehending the interpretability of ML models more easily. It involved the following: 1) A comparison of interpretability, using methods such as **SHAP**, among various machine learning frameworks (**PyCaret and PIML**) across diverse models: **KNN, SVM, Decision Tree, Cluster, Random Forest**, etc. 2) The implementation of **data cleaning** and **data normalization** as part of the data processing stage. The dataset used is derived from real-world **financial data**.
- **Natural Language Processing Center, Virginia Tech** — 01.2022-12.2022
  Research Assistant
  - **Outcome**: Developed an End-to-End Knowledge Informed framework, **KIEST**, for open domain entity state tracking framework. In this work, 1) KIEST incorporates the external entity and attribute knowledge to inform the model to better generate entity state changes with higher coverage. 2) A novel **Dynamic Knowledge Grained Encoder-Decoder** approach is proposed to dynamically incorporate the external knowledge and autoregressively generate all the entity state changes. 3) A novel **Constrained Decoding** strategy, coupled with an **Automatic Reward** function, has been introduced to assess the coherence of entity state changes. This work has been accepted by ***SIGIR 2023.***
- **Computer Science, Georgia State University (GSU)** — 01.2021-05.2021
  Graduate Research Assistant
  - **Outcome1**: Developed a Semantic Structure-based Query Graph Prediction framework, **SSKGQA**, for knowledge graph based question answering (KBQA). This project introduced a novel classifier called **StructureBERT** aimed at predicting the **Semantic Structure** for each question and then denoising the candidate query graph. Additionally, an improved BERT model has been proposed to rank query graphs to obtain the most relevant one for each question. This work has been accepted by ***COLING 2022.***
  - **Outcome2**: Developed a Hierarchical N-Gram framework, **HNZSLP**, for zero-shot link prediction. HNZSLP utilizes a novel GramTransformer to effectively model n-gram information derived from relation surface names, contributing significantly to enhancing task performance. This work has been accepted by ***EMNLP 2022.***

Publications and Patents

1). **Mingchen Li**, M.Chen, Huixue Zhou, Rui Zhang. **PeTailor: Improving Large Language Model by Tailored Chunk Scorer in Biomedical Triple Extraction**. Preprint 2023. [PDF], [Code].

2). **Mingchen Li**, Lifu Huang. **Understand the Dynamic World: An End-to-End Knowledge Informed Framework for Open Domain Entity State Tracking**. SIGIR 2023. [PDF], [Code].

3). **Mingchen Li**, Rui Zhang. **How far is Language Model from 100% Few-shot Named Entity Recognition in Medical Domain**. Preprint 2023. [PDF], [Code].

4). **Mingchen Li**, Junfan Chen, Samuel Mensah, Nikolaos Aletras, Xiulong Yang, Yang Ye. **A Hierarchical N-Gram Framework for Zero-Shot Link Prediction**. EMNLP 2022 Findings. [PDF], [Code]

5). **Mingchen Li**, Shihao Ji. **Semantic Structure based Query Graph Prediction for Question Answering over Knowledge Graph**. COLING 2022. [PDF], [Code].

6). **Mingchen Li**, Zili Zhou, Yanna Wang. **Multi-Fusion Chinese WordNet (MCW): Compound of Machine Learning and Manual Correction**. CICLing 2019. [PDF].

7). Ying Liu, Haozhu Wang, Huixue Zhou, **Mingchen Li**, Yu Hou, Sicheng Zhou, Fang Wang, Rama Hoetzlein, Rui Zhang. **A Review of Reinforcement Learning for Natural Language Processing, and Applications in Healthcare**. Preprint 2023. [PDF], [Code].

8). **Mingchen Li**, Yang Ye, Huixue Zhou, Jeremy Yeung, Huaiyuan Chu, Rui Zhang. **W-PROCER: Weighted Prototypical Contrastive Learning for Medical Few-Shot Named Entity Recognition**. Preprint 2023. [PDF], [Code].

9). **Mingchen Li**, Zili Zhou, Yanna Wang. **Solving the Chinese Physical Problem Based on Deep Learning and Knowledge Graph**. ICITE 2019. [PDF].

10). Yanna Wang, Zili Zhou, **Mingchen Li**, Yantian Hu, Zheng Su, Dezhi Rong and Ning Zhang. **An intelligent collection system for testing paper**. Publication number: CN107908752A. (**1st** student author).

11). Zili Zhou, Yanna Wang, Jinghu Zhang, Ning Zhang, **Mingchen Li**, Dezhi Rong. **An intelligent hardware control method driven by knowledge graph**. Publication number: CN107272521B. (**2st** student author).

Skills Summary

- **Languages**: Python, Java, SQL(Oracle), Unix, Linux, JSON, C++, C
- **Frameworks**: TensorFlow, Pytorch, MySQL, MongoDB, Neo4j, Transformers, Scikit-Learn, Pandas, PyCaret, PiML, Vaex