Zhenjie Sun
Google Scholar
http://szj.ai

Email : zjsun@umich.edu
Mobile : +1-734-596-7477

RESEARCH INTEREST

Large Language Models (LLMs) for Tabular Data, Table Representation Learning, AI Agent

Education

- **University of Michigan** — *Ann Arbor, U.S*
  B.S.E. Computer Science — *Aug. 2022 - Dec. 2024 (Expected)*
  - GPA: 3.93 / 4.00
  - Relevant Coursework: Natural Language Processing (A), Computer Vision (A+), Machine Learning (CSE) (A)

- **Shanghai Jiao Tong University** — *Shanghai, China*
  B.S.E. Electrical and Computer Engineering — *Sept. 2020 - Aug. 2025 (Expected)*
  - GPA: 3.5 / 4.0

Publications

[1] Naihao Deng*, **Zhenjie Sun***, Ruiqi He, Aman Sikka, Yulong Chen, Lin Ma, Yue Zhang, and Rada Mihalcea. Tables as Texts or Images: Evaluating the Table Reasoning Ability of LLMs and MLLMs. *In ACL 2024.* [pdf]

[2] Tianji Cong*, **Zhenjie Sun***, Paul Groth, H. V. Jagadish, and Madelon Hulsebos. Introducing the Observatory Library for End-to-End Table Embedding Inference. *In NeurIPS 2023 Second Table Representation Learning Workshop.* [pdf]

[3] Tianji Cong, Madelon Hulsebos, **Zhenjie Sun**, Paul Groth, and H. V. Jagadish. Observatory: Characterizing Embeddings of Relational Tables [Experiments, Analysis & Benchmark]. *In VLDB 2024.* [pdf]

***** Equal Contribution.

Research Experience

- **LIT Group - University of Michigan** — *Ann Arbor, U.S*
  Advisor: Rada Mihalcea — *Sept. 2023 - Present*
  - **Exploring the Strengths and Limitations of LLMs on Multimodal Representations of Tabular Data [1]**:
    - **Objective:** Evaluate Large Language Models' ability to understand tabular data using varied prompts and formats, introducing a novel assessment of image-based tables and providing insights for improving LLMs in table-centric tasks.
    - **Responsibility:** Co-led the project. Came up with various table representations and designed all the prompts. Aggregated and analyzed the experiment results.

- **Database Research Group - University of Michigan** — *Ann Arbor, U.S*
  Advisor: H. V. Jagadish — *April 2023 - April 2024*
  - **Characterizing Embeddings of Relational Tables [3]**:
    - **Objective:** Develop a comprehensive understanding of the strengths and weaknesses of language models and table embedding models. Introduce a formal framework for systematically analyzing the embedding representations of relational tables to improve efficiency and success in downstream applications.
    - **Responsibility:** Developed a comprehensive pipeline to transform tables into table, row, column and cell embeddings for language models like BERT and T5, and table models such as TURL and Tapas. Integrated eight distinct properties to characterize table embeddings and implemented corresponding quantitative measures.
  - **The Observatory Library [2]**:

- **Objective:** Encapsulate our research findings and methodologies from [3] into a user-friendly, end-to-end pipeline for assessing models' proficiency in generating table embeddings. The library is now available in the link.
- **Responsibility:** Played a pivotal role in the design and development of the library's core features. Involved in conducting thorough tests to ensure the stability and performance of the library across various use cases.

Programming Skills

- **Languages**: Python, C++, SQL, Java, Javascript
- **Technologies**: PyTorch, Prompt Engineering, MATLAB, LaTeX, Mathematica, Origin, Git

Activities and Awards

University Honors, 2023 & 2024 *Ann Arbor, U.S.*
Dean's List, Dec. 2022 & April 2023 & Dec. 2023 & April 2024 *Ann Arbor, U.S.*
Undergraduate Excellence Scholarship, 2021 *Shanghai, China*
Director, Rights Center of SJTU Student Union *Shanghai, China*
Leader, UM-SJTU JI People Interview Group *Shanghai, China*