Yiyang Bian

(216)-3545035 • yxb227@case.edu • https://www.linkedin.com/in/yiyangbian/

EDUCATION

Case Western Reserve University Cleveland, Ohio, US

- *Master of Science in Computer Science* -- Artificial Intelligence 09/2021 - 12/2023

Central China Normal University Hubei, China

- *Bachelor of Science in Computer Science* 09/2017 - 06/2021

RESEARCH INTEREST

Deeply interested in graph and data-centric research, particularly in leveraging Graph Neural Networks (GNN) to unearth profound structures and patterns in data. This unique intersection offers a lens to explore intricate relationships within data, providing innovative perspectives for data management, data mining, and machine learning.

PUBLICATION

- CRUX: Crowdsourced Materials Science Resource and Workflow Exploration (Published) CIKM 2022
- Selecting Top-K Data Science Models by Example Dataset (Accepted, publishing) CIKM 2023
- Multi-Objective Data Discovery for Data Science Models (Submitting) SIGMOD 2024

RESEARCH EXPERIENCE

Case Western Reserve University - Database Lab (RA) 2022 - Present Cleveland, Ohio, US

- Design and implement **Pegasus** workflow management scripts for scientific workflow of material analysis.
- Design datasets crawling and preprocessing **Python** scripts to extract image classification datasets features from **Kaggle** and **HuggingFace** platform, and utilized **Keras** pre-trained **machine learning** models to train get model performance and environmental settings.
- Implemented the comparison between the experimental part baselines and applied to the new test dataset, compared with our new **graph-based** data discovery algorithm and completed the report of the experimental learning part.

PROJECT

Extended Collaborative Filtering for Machine Learning Model Recommendation (Master Graduation Thesis)

- Design and implement a novel knowledge graph-based collaborative filtering algorithm for the scientific machine learning model recommendation task. And proposed a new cold-start solution based on interactive similarity.
- In terms of new performance such as algorithm time complexity and recommendation performance NDCG, it has surpassed Microsoft azure and ModsNet's GNN-based model recommendation algorithm proposed by us before.

WORK EXPERIENCE

BIOINVISION INC - SDE Full Stack Intern 2023 - Present Cleveland, Ohio, US

- Implemented **Azure MongoDB** to optimize manage 3D medical image datasets, reduce data retrieval time by **40%**.
- Built a user-friendly front-end and back-end separation system using **Python Flask**, and **ReactJS**, focusing on email-based one-time password (OTP) authentication and permission management. Applied low-latency techniques:
  - Optimized **HTML**, **CSS**, and **JavaScript** code for faster loading.
  - Implemented **lazy loading** and prefetching for improved performance, and improved the loading speed by **30%**.

TEACHING

High Performance Computing (HPC) - TA 8/2022 - 12/2022 Cleveland, Ohio, US

- Led HPC instruction covering MPI, OpenMP, GPU programming with CUDA/OpenCL, utilized diagnostic tools for performance tuning, managed HPC resources using Slurm/Torque, introduced parallel file systems and cloud computing techniques, and collaborated on course improvements based on feedback.

TECHNICAL SKILLS

- Programming Languages: Java, JavaScript, CSS, Python, Swift (SwiftUI), MATLAB, SQL, C#, C++
- Database: MySQL, SQL Server, Oracle, MongoDB, Neo4j
- Frame / Tools: Spring, Spring Boot, Mybatis-Plus, Maven, Git, Linux Command, Azure, AWS
- ML Tools: TensorFlow, PyTorch, Keras, Scikit-learn, AutoML, Pandas, Numpy, Matplotlib
- Academic Paper: Zotero, LaTeX, Overleaf