Taiwei Wu

nz-666.com · taiwei.wu@polytechnique.edu

Education

Institut Polytechnique de Paris, Data Science, *M2* 2025.9 - Now

Institut Polytechnique de Paris, Data and Artificial Intelligence, *M1* 2024.9 - 2025.8

Shenzhen Technology University, Computer Science and Technology, *Bachelor* 2020.9 - 2024.6

President's Award *(the highest scholarship at SZTU, only 0.3% of the undergraduate students were awarded)*

Project Experience

A MIDI Retrieval System Based on MG2V 2023.3 - 2024.5

- Project Background: Representing words as embeddings is commonly used in NLP tasks. As music shares properties with text, representing music as embeddings could also enable retrieval systems. However, training music embeddings in common methods like skip-gram would significantly increase training time as data scales. An algorithm Music-Graph2Vec (MG2V) aims to reduce this time. Based on this algorithm, we plan to develop a system allowing users to upload MIDI files and receive song metadata through a web interface.
- Role: As the project lead, I initially collected relevant materials and selected the topic. I conducted the core development work and programming for the MG2V algorithm, and directed my group members to do experiments related to MG2V. We have had a paper accepted to ACM MM Asia 2023 conference and have been invited to attend the event in Taiwan. Currently, I am now building a web interface based on Flask to utilize MG2V for music retrieval.

An Outline On Wastewater-Based Epidemiology 2023.12 - 2024.4

- Project Background: Epidemic surveillance relies heavily on conventional clinical diagnostic test(CDT) , but CDT' s capacity largely lags behind the demand and the presence of asymptomatic virus carriers undiagnosed by clinical surveillance. The reports of SARS-CoV-2 viral genome detection in sewage networks highlights the potential of wastewater-based epidemiology (WBE) as a promising global solution complementary to CDT. In this project, we summarized the technological evolution of WBE, currently used WBE techniques for the detection and quantification of the virus and the effectiveness of WBE in monitoring environmental viruses and predicting epidemics.
- Role: I am in charge of summarizing the use of WBE in epidemic prediction. I investigated past studies predicting epidemics based on WBE. I summarized the findings and scenarios of these modeling efforts and wrote a part of our review article.

Cross-modal Language Model for Botanists 2025.4 - 2025.8

- Project Background: Multimodal large language models that can process information from various sources, such as images and text, are becoming increasingly prevalent. The LLaVA (Large Language-and-Vision Assistant) framework, which connects a vision encoder with a large language model, is a prominent example of such a model. This project aimed to leverage the LLaVA architecture to develop a specialized multimodal language model for botany, capable of understanding and reasoning about plants using their genetic, visual, and textual data.
- Role: My primary responsibility was to conduct a comprehensive survey of existing multimodal large model frameworks to determine the most suitable architecture for our project. I was also in charge of collecting and organizing the diverse dataset required for training, which included plant-related genetic sequences, images, and textual descriptions (flora). Using this curated dataset, I trained a 14B parameter multimodal language model for botanists based on the LLaVA framework. The model was trained on the Jean Zay supercomputer, a high-performance computing resource at IDRIS, the national computing center for the CNRS (France's National Centre for Scientific Research).

Publication

Taiwei Wu, Jianhao Zhang, Lian Duan, and Yuanzhe Cai. 2024. Music-Graph2Vec: An Efficient Method for Embedding Pitch Segment. In Proceedings of the 5th ACM International Conference on Multimedia in Asia (MMAsia '23).

Leadership Experience

ACM-ICPC Club of SZTU, *Co-founder & Inaugural President* 2020.10 - 2022.10

Our club mainly focuses on training members to take part in collegiate programming competitions, such as ICPC (International Collegiate Programming Contest). In 2020, I gathered a small group of 20 fellows to launch the club. Now we have nearly **800 members**, with around 200 people regularly attending each training session.
Through our collective efforts, the club has built a strong training program. It includes introductory coding workshops for newcomers as well as personalized training plans for more advanced members based on their competition goals. During the 2021-2022 academic year alone, members collectively earned 38 awards at the national collegiate programming contests. During my time leading the club, we organized six programming competitions across campus successfully. Participation grew from around 60 people initially to over 300 (unfortunately we had to hold an online qualifying round due to space issues). In the annual reviews by our university, our club has consistently been rated "Excellent" since we started.

Work Experience

ShenZhen Wemed Medical Technique Co., Ltd, *Algorithm Development Engineer* 2023.3 - 2024.4

For over a year, I have been employed part-time in this role, during which I led a research team of four individuals to develop the Music-Graph2Vec algorithm.

IRD-UMMISCO, *Intern* 2025.3 - 2025.8

I worked at French National Research Institute for Sustainable Development - Unité Mixte de Modélisation mathématique et informatique de systèmes complexes as an internship student. During the internship, I was involved in Cross-modal Language Model for Botanists project.

Awards

The First Prize of the National Final of RoboCom-CAIP Programming * 2022

RoboCom is a national competition held by Talent Exchange Center of Ministry of Industry and Information Technology (MIITEC), recognized by the Ministry of Education of China.
A total of 1126 competitors advanced to the the National Final, where I ranked 58th.

The Second Prize of the National Final of Lanqiao Cup C++ Programming Competition * 2023

"Lanqiao Cup" is one of the longest-running collegiate programming competitions in China.
It is also a national competition recognized by the Ministry of Education of China.

President's Award ‡ 2022

The highest scholarship at SZTU (¥20000). Only 43 out of 13,000 undergraduate students received it.

The First Prize of Research and Innovation Award ‡ 2022

Granted by SZTU for my outstanding performance in the competitions.

Outstanding Graduate of Shenzhen Technology University ‡ 2024

*National award
‡University-level award