Tianzhang Cai

Phone: +447786484588 ❖ Email: jultcai@gmail.com ❖ Github: tztsai ❖ LinkedIn: tianzhang cai

WORK EXPERIENCE

Research Software Engineer & Research Associate *Nov. 2023 – Present*
University of Cambridge Cambridge, United Kingdom
- ✧ Collaborated with researchers and software engineers to develop ML models that couple with climate science software. Ran model simulation and ML training on the GPU cluster in the University's HPC environment (CSD3).
- ✧ Automated testing and deployment of models through CI/CD pipelines using Github Workflows.
- ✧ Provided two workshops on machine learning and scientific computing at the ICCS Summer School and APSCI.

Research Engineer *Nov. 2022 – Mar. 2023*
KTH Royal Institute of Technology Stockholm, Sweden
- ✧ Conducted cellular traffic data analysis and user behavior modelling of different categories based on a large-scale (~1TB) DPI dataset provided by the Swedish mobile operator Tele2.
- ✧ Designed and implemented a multi-agent RL algorithm (MAPPO) in PyTorch. Ran distributed training on Google Cloud.
- ✧ Developed an advanced multi-cell massive MIMO 5G network simulation environment in Python.

Machine Learning Engineer Intern *Sep. 2021 – Dec. 2021*
DeepWisdom Remote
- ✧ Conducted tabular and time-series data collection, analysis, preprocessing, and feature engineering.
- ✧ Time series AutoML (AutoSeries) pipeline refactoring, testing, and benchmarking. Improved product sales prediction by 15% for a client company using our AutoML solution.
- ✧ Paper review and reproduction of Deep Feature Synthesis.

EDUCATION

KTH Royal Institute of Technology **Stockholm, Sweden**
M.Sc. Machine Learning *Sep. 2021 – Oct. 2022*
Artificial Intelligence (A), Machine Learning (A), Data Mining (A), Scalable ML and Deep Learning (A).

Shanghai Jiao Tong University **Shanghai, China**
B.E. Electrical and Computer Engineering *Sep. 2017 – Aug. 2021*
Honors Math II (A+), Honors Math III (A), Discrete Math (A), Programming & Data Structures (A).

PROJECTS

Pyrealm - an Integrated Toolbox for Plant Modelling *Cambridge, 2024*
Pyrealm is a Python package that provides an integrated toolbox for modelling plant productivity, growth, and demography. I developed a data-driven model for plant productivity, wrote unit tests, regression tests, and profiling tools, and incorporated them into the CI/CD pipeline. I also implemented a data visualization module to plot the profiling and benchmark results.

Sythcity – a Synthetic Data Library *Cambridge, 2023*
Synthcity is an ML-based Python library for generating and evaluating synthetic tabular, time series, and survival analysis data used for training AI healthcare models, maintained by the van der Schaar Lab at the Cambridge Centre for AI in Medicine (CCAIM). I am one of the major contributors of this project and my work consists of implementing ML models like TabNet and TabDDPM, refactoring and improving the data processing pipeline, introducing hyperparameter optimization, and more.

Traffic Analysis and Multi-Agent RL for Energy-Efficient Base Station Control *KTH, 2022*
BS-Sleeping-RL is a part of the EU Celtic Plus Project AI4Green. In this project, I independently implemented a multi-agent deep RL algorithm (MAPPO) using PyTorch for cooperative decision-making of base stations in multi-cell 5G networks, to optimize energy efficiency while preserving quality of service, achieving a 19% improvement in energy efficiency in comparison to today's symbol-level BS sleeping mechanism. The results were published on ICMLCN 2024.

Fine-tuned Sentence Transformer for Semantic Search System *KTH, 2022*
In this project, I implemented a PyTorch sentence transformer using a pre-trained BERT as the base model, and fine-tuned it using Natural Language Inference (NLI) and Semantic Textual Similarity (STS) tasks. I evaluated the model performance using the STS benchmark and it obtained 85% accuracy. Then I used the model to generate sentence embeddings of IMDB reviews and built a vector database in which the user can query the most similar reviews given an input string.

Metal Artifact Reduction in CT Images using CycleGAN *KTH, 2022*
In this project, we implemented and trained a CycleGAN model in PyTorch with Feature Pyramid Network (FPN) to reduce metal artifacts in CT images of human brain (CT-MAR). Our novel introduction of FPN improved the SSIM (structure similarity) score by 7.5% in comparison to the vanilla CycleGAN approach.

SKILLS

Programming Languages
Python (proficient), C/C++ (intermediate), Julia (intermediate), MATLAB, C#, Java, JavaScript.

Machine Learning
Scikit-learn (proficient), PyTorch (proficient), TensorFlow, Hugging Face, WandB, LangChain, Optuna, Stable Baselines.

Data Analysis & Data Engineering
NumPy (proficient), Pandas & Polars (proficient), SQL (intermediate), Xarray, Dask, PySpark, Matplotlib, Plotly, Dash.

DevOps & Cloud Computing
Git, Docker, CMake, Pytest, Poetry, Github Workflows, Google Cloud (Compute Engine, Vertex AI), Kubernetes, Slurm.

Languages
English (fluent), Chinese (native).

Interests
Reading, Tennis, Hiking, Programming Puzzles (my full solutions of AoC 2023).

PUBLICATIONS

Cai, Tianzhang, et al. **Multi-Agent Reinforcement Learning for Energy Saving in Multi-Cell Massive MIMO Systems**. arXiv:2402.03204, arXiv, 5 Feb. 2024. arXiv.org, https://doi.org/10.48550/arXiv.2402.03204.