Jiarui Xu

+61 04 1269 6563 Email LinkedIn XavierSpycy

Summary

- Studying Master of Data Science (Research) at the University of Sydney, with a specialization in Machine Learning;
- Deeply passionate about research and desired to explore the realms of computer vision and natural language processing;
- Experienced in object-oriented programming, especially Python, with good coding practices.

Technical Skills

Languages/Database: Python, MySQL
Data Analytics Libraries: Numpy, Pandas, Scipy, Matplotlib
Machine learning & Deep Learning Frameworks: SciKit Learn, PyTorch, TensorFlow, MLForce(self-developed)

Education

University of Sydney, Sydney **Feb 2023 – Jun 2024**
Master of Data Science *WAM: 80.5*
Specialization: Machine Learning
Coursework: Deep Learning, Machine Learning and Data Mining (HD), Computational Statistical Methods (HD), etc.

Nanjing Tech University, Nanjing **Sep 2017 – Jun 2021**
Bachelor of Industrial Engineering *GPA: 3.77/4*
Specialization: Logistics and Supply Chain Management

Projects

Robust Trainer for Noisy Labels | *PyTorch Git* **Oct – Nov 2023**

- **Robust Trainers**: Implemented and deployed four distinct robust training strategies, significantly enhancing the performance of classifiers.
- **Transition Matrix Estimator**: Implemented the Dual-T estimator achieving reliable and accurate estimations.
- **Performance**: Achieved over 90% and around 80% accuracy on datasets with 50% and 60% noise levels, respectively, showcasing robust resilience to data noise.

NumPy-Based Machine Learning Framework Development | *NumPy* **Aug – Nov 2023**

- **Keras-Style Multilayer Perceptron (MLP) using NumPy** *Git*
  - **Activation Functions**: Implemented all activation functions available in PyTorch, along with their derivatives.
  - **Layers**: Implemented Dense with various initialization strategies, Batch Normalization, and Dropout layers.
  - **Optimizers & Callbacks**: Developed popular optimizers, multiple learning rate schedulers, and early stopping.
  - **MLP**: Engineered regression and classification loss functions and integrated other advanced techniques.
  - **Results**: Achieved satisfactory performance on various datasets efficiently.
- **Non-negative Factorization (NMF) using NumPy** *Git*
  - **Algorithms**: Implemented eight effective and efficient NMF algorithms, each with distinct loss functions.
  - **Experiments**: Conducted extensive experiments on two facial image datasets under ten different noisy conditions.
  - **Framework**: Designed a comprehensive framework to facilitate the easy creation of new algorithms for academic research.

Multi-Modal Integration for Text-Image Classification | *PyTorch Git* **Apr – Jul 2023**

- **Extraction**: Deployed distilled BERT models for text encoding and utilized ResNet or DenseNet for image encoding.
- **Integration**: Implemented self-attention and cross-attention mechanisms to effectively integrate features extracted from multimodal sources.
- **Outcomes**: Achieved a top 22% (31/142) ranking with an F1 score of 87.5% in an internal Kaggle competition.

Convolutional Neural Networks for Handwritten Character Recognition | *PyTorch Git* **Apr – Jun 2023**

- **Classifiers**: Implemented variations of six unique CNNs, encompassing both traditional and contemporary designs.
- **Performance**: Attained nearly 90% across all metrics on training and test sets by pre-training and fine-tuning.
- **Generalization**: Exhibited generalization capabilities on the entire dataset as well as downstream datasets.