Jiarui Xu

+61 04 1269 6563 Email LinkedIn Profile 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 Libraries: 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, Computational Statistical Methods, etc.

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

Projects

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

- **Data augmentation and normalization**: For image, random flipping/rotation, contrast adjustment, etc, were employed. For testing input, a normalization method was used involving stretching and padding. For text data, an augmentation method involving word swapping was utilized, and the text was encoded into binary vectors based on attention masks.
- **Feature fusion with attention mechanisms**: Image classification models, as well as lightweight language models like TinyBert, were deployed and trained. Attention mechanisms were employed to construct multi-modal models.
- **Training Process**: Architecture optimization involved combining different base classifiers and gradually unfreezing the feature extraction layers. A learning rate scheduler was used for warm-up, and the AdamW optimizer was used.
- **Team Competition**: In the Kaggle Competition, I achieved a top 22% (31/142) ranking with an F1 score of 87.5%.

CNNs for EMNIST Handwritten Character Recognition | *PyTorch Git* **Apr – Jun 2023**

- **Implementation**: Variants of six convolutional neural network models, including VGGNet, ResNet, etc, were trained on the EMNIST ByClass dataset. The training process utilized only around 12% of the full dataset as training data.
- **Pre-training and Fine-tuning**: Hyperparameter combinations were obtained through random search. The models were pre-trained using the SGD optimizer on 90% of the training data and fine-tuned using the Adam optimizer on the full training data.
- **Performance**: All metrics the models achieved are close to 90%. The models also demonstrated strong generalization ability. Testing the trained models on the full dataset showed that all performance metrics remained above 86%.

Implementation of a Multi-Class Multi-Layer Perceptron Classifier from Scratch | *Numpy Git* **Feb – Apr 2023**

- Built an object-oriented framework for a multi-layer perceptron, including Activation, HiddenLayer, and MLP class.
- **Activation Class**: Implemented multiple activation functions and their derivatives, including Tanh, Sigmoid, ReLU,Softmax, etc.
- **HiddenLayer Class**: Implemented forward and backward propagation, Xavier and Kaiming initialization, Batch Normalization, and Dropout techniques.
- **MLP Class**: Implemented cross-entropy loss function, mini-batch training process, and various optimizers such as Adam, etc.
- **Experiments**: The model achieved comparable accuracy but higher computational efficiency compared to that using PyTorch.