Liangyu Li

Website: https://lawrencelly.github.io

✉ llympcb@gmail.com   in linkedin.com/in/liangyu-li-239550256   ⓞ github.com/LawrenceLLY

Education

Washington University in St. Louis   **Aug 2024 – May 2029 (Expected)**
Doctor of Philosophy in Computational and Data Sciences   *St. Louis, MO*

Georgetown University   **Aug 2022 – May 2024**
Master of Science in Computer Science (GPA: 3.867/4.000)   *Washington, DC*

The Chinese University of Hong Kong   **Aug 2021 – Jul 2022**
Master of Science in Physics   *Shatin, Hong Kong*

BNU–HKBU United International College   **Sep 2017 – Jun 2021**
Bachelor of Science (2:1 Honors) in Computer Science and Technology (Major GPA: 3.47/4.00)   *Zhuhai, China*

Relevant Coursework

- Artificial Intelligence
- Machine Learning
- Deep Learning
- Reinforcement Learning
- Computer Vision
- Data Mining
- Natural Language Processing

Research Experience

Physics-Informed Neural Networks for Solving Nonlinear Partial Differential Equations   **Sep 2023 – Dec 2023**

- Using **Physics-Informed Neural Networks (PINNs)** to solve different physics nonlinear partial differential equations, including nonlinear Schrödinger equation, Burger's equation, diffusion equation, and Poisson's equation.
- PINNs are trained by incorporating physical laws as soft constraints in the loss function, considering the neural networks as a function.
- Define different initial conditions and boundary conditions in order to **create different loss functions**.

Inquisitive Conversational Agents by Offline Deep Reinforcement Learning   **May 2023 – May 2024**

- Dataset is the US Supreme Court, created different **measurement methods of Justice's utterance** and considered as the **reward function** of reinforcement learning.
- Using Reward-on-the-Line (ROL) by measuring the **ID and OOD agreement** of action selections to get the **linear relation** to fix rewards.
- Using **Double Conservative Q-Learning (DCQL)** with regularization terms to score the multi-responses to each state, evaluated by P@1.
- The multi-responses generated by **LLM** and using the DCQL model to **score** them, the response with the highest score is the final response.

Pun Detection and Location with BERT and Graph Convolutional Neural Networks   **Mar 2023 – May 2023**

- **SemEval-2017 Task 7**: each sentence contains a **pun**; need to identify whether a sentence contains a pun or not and locate the punning word.
- Using **Multi-Task Learning**. Using **BERT** to get the embedding of each sentence, building **dependency trees** for each sentence to be the **graphs**, combine them to become the input of **Graph Convolutional Neural Networks (GCNs)**, another input of the model is the **index** of pun in the sentence.
- Built a **control module** that lets the neural network focus on **learning the features** of a token through the input token index by extracting the corresponding token's features slice from the output tensor of BiLSTM, using Sigmoid to identify whether the word is a pun or not.

An Approach Based on Deep Learning for DDoS Detection and Classification   **Oct 2022 – Dec 2022**

- Using MLP, CNN, and RNN (GRU, LSTM) to train on the DDoS dataset (CIC-DDoS2019, Kaggle's DDoS) to do **classification**, try to find the performance by using different features, and implementations to identify DDoS attack traffic, including doing **real-time learning**.
- Using **Autoencoders** to do clustering for DDoS data, we found that the features of benign data are more diverse than DDoS.
- Using **transfer learning**, implement DDoS to collect the data every 1 hour, and use new data to train the model to improve the models.

Machine Learning and Deep Learning for Fluids and Crystal Structures   **Jan 2022 – May 2022**

- Using CNN, MLP, and RNN to do **classification** for crystal structures, including Liquid, BCC, FCC, HCP, BCCpre, FCCpre, and HCPpre.

- Using **unsupervised learning** such as Autoencoder, t-SNE, and PCA to do **clustering** for crystal structures and fluids.
- Visualizing energy landscapes of fluids by **manifold learning**, based on g(r) and LJ clusters of water molecules (use all pairwise distances).

Machine Learning and Deep Learning for Gas Sensors and Spectrums   Jan 2022 – May 2022
- Using MLP, LSTM, GRUs, Ridge/Lasso/Elastic Net/Linear regression to **predict** concentrations of different gases of the mixture gas.
- Data are the values of several gas sensors which are simulated by the absorption spectrums of Fourier-Transform Infrared Spectroscopy (FTIR).
- Need to find the best number of filters and their corresponding region of wavenumber, using R^2 Score to evaluate.
- Filters are doing several Gaussian filters of the bandwidth of 50 cm^{-1}, integral these intervals to become several numbers.

Classifying Patients' Medical Records Using Deep Learning and NLP   May 2020 – Dec 2020
- **Multi-label classification** of various diseases of patients, the dataset is text medical records of **different diagnoses**.
- Implement MLP, CNN, LSTM, BiLSTM, GRU, CRNN, BERT, ALBERT, TextCNN, FastText, Attention, Heterogeneous Graph Attention Network (HAN), etc. with different features of data, to compare the performance.
- Using back-translation, randomly disorder the words based on stop words to do **data augmentation**.

Artificial Intelligence and Computer Vision Four-Wheel Drive Smart Robot   Sep 2020 – Oct 2020
- Programming: Python, C, OpenCV, TensorFlow, VMware. Hardware: Raspberry Pi 4B mainboard, Driver Board, electronic parts
- **AI** and other functions: voice interactive, voice-activated movement, infrared obstacle avoidance; WIFI video, voice broadcast, mobile control
- **Computer Vision** functions: face/sphere tracking, face/color/QR-code/gesture recognition, tracks black lines, multi-class object recognition

Question-Answer Chatbot by using Natural Language Processing and Machine Learning Oct 2019 – Nov 2019
- The dataset is each question corresponds to an answer. Implement data augmentation by writing several similar questions to each question.
- Using Jieba for Chinese word segmentation and using pre-trained word2vec word vectors for word embedding with genism.
- Using sentence similarity (cosine similarity), match the input with the most relevant questions and get the corresponding answers.

Professional Experience

Georgetown University   Aug 2023 – Dec 2023
Teaching Assistant   *Washington, DC*
- Help to teach COSC-3440 **Deep Reinforcement Learning**.

Geosys Hong Kong Limited   Jun 2022 – Aug 2022
Artificial Intelligence Developer (Intern)   *Hong Kong*
- Using the technology of **artificial intelligence**, **deep learning**, **computer vision**, and **image processing** with PyTorch & OpenCV, **object detection** with YOLOv5, **object tracking** with StrongSORT and OSNet, and **image inpainting (GAN)** with LaMa.
- Corresponding tasks included human face detection, Hong Kong license plate detection, wall defect detection, multi-class object detection in the VR3D platform, and surveillance video real-time object detection and tracking. Developed a webpage for real-time surveillance video.

Tisson Regaltec Communications Tech Co., Ltd.   Jun 2020 – Jul 2020
Artificial Intelligence Developer (Intern)   *Guangzhou, China*
- The internship is developing software by using **artificial intelligence**, **machine learning**, **deep learning**, and **natural language processing**. Wrote a chat assistant software to assist communication with customers by displaying **prompts** based on conversation content.
- Using PageRank and TextRank to do **information extraction**, and match relevant content in the knowledge base as prompts. Completed various corresponding tasks including **information extraction**, **multi-label text classification**, word embedding, and so on.

Technical Skills

Languages: Python, Java, C, HTML/CSS, JavaScript, SQL, C++
Developer Tools: PyTorch, TensorFlow, NumPy, scikit-learn, SciPy