Statement of Purpose

By Haoyu Qi

I am committed to continuing my graduate studies in the Master of Information Systems Management (MISM) at Carnegie Mellon University. Specifically, I am interested in the curriculum and research projects centered around combining business and technology.

My fascination with information technology took root in freshman year during the course *Introduction to Computer Science*. My teammates and I developed a simple Gobang automatic battle program which was primarily designed using a maximum-minimum algorithm and therefore was not capable of learning and updating its own functions. Wishing to evolve what we had built, I searched a multitude of papers and discovered that machine learning can achieve self-updates by simulating Gobang battles, a startling advancement which triggered my passion for machine learning. To subsequently expand my knowledge in this field, I self-learned *Machine Learning* and *Convolutional Neural Networks for Visual Recognition* on Coursera. A brief introduction to deep learning as part of this process made me realize the power of deep learning algorithms in processing massive amounts of data. My abiding curiosity for deep learning was born.

To explore the application of deep learning in medical imaging in a more comprehensive fashion, I commenced a project titled ***Multi-label Fixmatch with MoCo Pretrained Models on Medical Image Domain.*** There are currently multiple challenges facing computer scientists in the processing of medical images. First, medical images have domain bias when compared to natural images. Pixel-based representation in medical images bear huge differences versus what we see in natural images. There also exists an image classification problem – medical images are usually multi-task and multi-label, as opposed to the common single label in the natural imaging field. Additionally, fine labeled medical data sets require expert domain knowledge and are also open to being challenged according to the rights of patients. Public medical data sets are even more rare, a situation which leads to poor performance via supervised learning. As a result, the direct use of natural imaging field methods greatly reduces the interpretability of the model, and also renders the classification results of the model unsatisfactory. After searching through a massive amount of information on the intersection of medicine and deep learning, I managed to improve the Fixmatch algorithm to make it suitable for both Multi-task and Multi-label models. Regarding the problem with fine labeled medical data, I adapted specially-designed data augmentation methods for medical images, such as Non-linear Transform, Local Shuffle, Out Painting, and In Painting combined with conventional data enhancement methods to generate negative samples with strong interpretability. According to the open-source data sets CheXpert and Chest8, the verified results showed that the outcomes of self-supervised learning reached 81.24 and 78.97 AUC, which exceeded or were consistent with the results of fully supervised learning (78.67 AUC) while possessing better interpretability. This unique research experience emphasized that aside from enhancing the algorithm itself, it is imperative to consider how to deploy a huge 3D model to disperse a large amount of data to different machines in the interest of balanced computing power. Only then can deep learning truly be applied to the recognition of medical images and achieve outcomes similar to professional doctors.

To better implement the deployment of the model, I taught myself ***Distributed Systems*** developed by the Massachusetts Institute of Technology (MIT), proactively read classic papers on subjects such as GFS, MapReduce, RAFT, VMFT, ZooKeeper, Paxos, and more, and mastered the K/V structure distributed storage system of the RAFT protocol simulation based on Go language. Still longing to gain systematic expertise related to computer systems, I took part in Professor Ding Zhijun's laboratory, where we implemented ***The Design and Development of Proactive Microservice Dynamic Deployment Tools***. Presently, the application of machine learning methods in active microservice deployment happens mostly

Statement of Purpose

By Haoyu Qi

through the known load sequence correlation training model, predicting the future load sequence correlation and allocating the resource deployment of the remaining services according to the request volume of certain services. However, traditional methods emphasize the extraction of time relationship features and ignore spatial factors, which are mostly comprised of dependency relationships between microservices, so that the prediction models cannot transfer well to more general architectures. To tackle this dilemma, I obtained the data correlation, time correlation, trend correlation, and volatility correlation characteristics between microservices by designing the network structure, and then further extracted the size, time, positive and negative relationship characteristics, and external event characteristics to mine the spatial (dependency) relationship between microservices. Consequently, the model now fits well within different deployments and the interpretability of the model is dramatically improved. Apart from having a crucial hand in the load prediction model, I also designed the active scaling part and modules to execute the scaling plan, and finally wrapped them together to form an active microservice dynamic deployment tool. Thanks to this project, I learned how cloud computing can better serve the deployment optimization of machine learning models, as well as how to combine continuous training components with responsive executing components.

With theory in hand, I desired to know more about how machine learning and computer systems are combined and applied in the industry. An internship at Intel was thus a natural choice, where I focused on OpenCL compile optimization using machine learning methods. Industrial entities can deploy machine learning in scenarios with high response speed and solve problems via multi-dimensional data analysis. This internship inspired me to set my career sights on becoming an excellent system architect and to adopt machine learning in service of optimizing systems. In the short term, I aspire to build a solid foundation of basic knowledge by participating in all kinds of projects. In the long term, I hope to combine machine learning and computer systems to create products that contribute to humanity's progress.

In conclusion, my studies and research experiences have led to my advanced maturation as a worthy postgraduate candidate. The Master of Information Systems Management (MISM) program at Carnegie Mellon University appeals to me because of its world-class faculty and outstanding academic atmosphere. What makes it irresistible is the inspiring and comprehensive curriculum the program offers. Courses, such as *Distributed Systems for ISM* and *Object-Oriented Programming in Java*, will help me build a sound knowledge base and prepare me for both my short-term and long-term career goals. And the finance and management-related courses, for instance, *Economic Analysis and Professional Speaking*, will give me a thorough understanding of starting up. The large volume of real-world projects, for instance, the Information Systems Capstone Project, will equip me with practical skills and state-of-the-art methodology. The required internship will prepare me well for my future career in advance. Under your constructive guidance and compelled by my ceaseless enthusiasm, I am confident that your program will shape me into a better person with limitless potential.