Statement of Purpose

Lili Wang

During high school, I participated in the National Informatics Contest (examining all kinds of algorithms and data structures) and eventually won the silver medal. This accomplishment allowed me to be admitted to Shanghai Jiao Tong University (SJTU) without taking the National Entrance Exam. I entered the ACM class (a pilot computer science program in SJTU) to which only the top 5% of nearly 500 students in the entire School of Electronic Information and Electrical Engineering are admitted. I studied incessantly and performed well in all courses related to computer science and mathematics. In addition, I have witnessed the great breakthrough in Computer Science and particularly in Machine Learning over the past several years. I also had the honor of participating in several related scientific researches regarding topics of Computational Biology, Nature Language Processing and Computer Vision. These projects expanded my horizons and guided me to my interest in computer science. So I've decided to apply for your MMath program to further explore the field of computer science after my undergraduate study.

In 2015, I was involved in a bus accident after the driver fell asleep at the wheel. After I recovered, I decided to devote myself to research on Computational Biology. I joined Professor Lu's lab to develop a way to detect people's quality of sleep attained the night before by measuring their brain waves. First, I recruited eight volunteers (seven as my training set and the other individual as a testing set) and collected their sleeping data. After de-noising the original data, I extracted the Power Spectral Density (PSD) and Differential Entropy (DE) features to conveniently classify the pattern recognition. However, I countered a difficulty in that the extracted electroencephalograph (EEG) features were usually unstable and changed acutely. Conversely, the quality of human's sleep should be a stable and slow transition, rather than a sudden and dramatic one, and, as a result, the quality of sleep associated with the EEG feature should be smooth with the changing time intervals. To solve this problem, I used the LDS method to smooth the EEG features, but the recognition accuracy rate was still not ideal. After reviewing much literature, I tried the minimal-redundancy-maximal-relevance (MRMR) algorithm feature selection algorithm, after which the accuracy increased. I took the relatively new Discriminative Graph Regularized Extreme Learning Machine as the classifier and finally improved my detection accuracy to 83.57%. This was a great result, especially because this figure demonstrated that although my testing method to determine someone's quality of sleep was completely new, it was also completely feasible. Based on these experiment results, I succeeded in publishing a first-author paper in IJCNN 2016. Furthermore, our lab is now furthering my results and applying projects to the Railway Academy of Sciences. I look forward to seeing how this method can be applied to prevent the drivers and pilots from nodding off and causing disastrous accidents. From this research project, I gained the ability to design experiment flow and solve problems independently. I also gained a real understanding of how scientific research can improve people's daily lives. I felt a sense of mission to continue my MMath study in scientific research after realizing the results of such research would be adopted to practical application to save human lives.

After finishing the aforementioned research, I began another project at CoNLL 2016 on natural language processing

(NLP), in which my task involved using a Connective classifier, Explicit Sense Classifier and Non-Explicit Sense Classifier. The function of the connective classifier was to determine whether each connective functions as a discourse connective. The function of the latter two classifiers was used to distinguish the discourse relation that the sentence expresses (e.g. cause & effect, transition). One of the first problems presented itself while selecting a proper model. Because this was my first attempt in this field, the results of a series of tests were not satisfying. At this point, a graduate student in our laboratory recommended me the constituent parse tree model. After deducing each sentence into a constituent parse tree and extracting simple features to conduct the experiment, I found the results were very good. I continued to explore which features could lead to better results. Based on this model, I spent just one week experimenting and selecting several features which performed well for each part of my task. After an intense training and testing session, the F1 scores of my charged connective classifier and explicit sense classifier finally exceeded the best results of CoNLL-2015, both on dev set and blind set. I successfully published my second paper in CoNLL 2016 as the third author. Through this experience, I practiced NLP and realized that adding new content to improve an experiment based on previous studies is also an efficient way to conduct scientific research. Through this project, my NLP abilities as well as machine learning were largely improved and I began to be more skilled when cooperating with others in scientific research.

After three years of studying at SJTU, I got the chance to work as a research intern at Microsoft Research Asia in Beijing on an online people counting project in which I have to implement a people tracking algorithm to track multiple human bodies in the video, which of course presented many challenges. To begin my work, I had to catch up with the whole team's pace as soon as possible. I accomplished this by reading through the core framework of the code and becoming familiar with the convolutional neural networks and Caffe (an open-source deep learning framework). I then started in on the important details. Another challenge involved the actual algorithm. To improve real-time efficiency, it should test interval frames instead of each one. However, for the current algorithm was based on Lucas-Kanade Optical Flow, the error will be very large if the detected frames are far apart. Currently, I am trying to overcome these difficulties and contribute my best work to the project, and my ability to code and learn independently have largely improved during this process.

These past and ongoing experiences have equipped me a deep understanding and skills in different areas in computer science. Through these varying experiences, I gained a clear understanding of what I really want to do, which is Computational Biology and Bioinformatics. After careful deliberation, I have determined to pursue my MMath degree to further my research in computer science combined with health, pharmacy, or safe medical treatment in order to make this world better. My background of mastering of various algorithms as well as different ideas toward computer science will help me in future studies. University of Waterloo is one of the best schools for Bioinformatics; Prof. Bin Ma's research is incredibly interesting, since the analysis of DNA sequences can help us understand and conquer different diseases, and I hope to become involved in his lab. I think with my great passion and skills, I am a good candidate for your program, and I look forward to furthering my studies at your school.