Personal Statement

Dating back to the beautiful summer three year ago, I shared great time with a friend majoring in Computer Science. It was the first time that I stepped into the appealing AI world. Deeply impressed by his colorful codes in editor to build an automatic question answering system, I could not wait to run those codes with his help and interact with this system. Since then, I have developed an interest in the nature of Artificial Intelligence.

In the first summer vacation after I began to study at SJTU, I accepted my friend's invitation to participate in his research on the optimization for biterm topic model(BTM). It used Gibbs sampling method, which is very time-consuming while Alias method proposed by optimizer of LDA(Latent Dirichlet Allocation) performs better in time complexity. What our model does is to apply Alias to BTM. This was my first time to deal with NLP. Despite the fact that I had little knowledge about the topic model, and even had not taken any data structure class before, I tried my best to master all relevant knowledge as quickly as possible. I read papers on LDA and BTM and then traced back to learn the corresponding math knowledge. And then I carried out experiments on three datasets including long and short texts and our model presents a great improvement. Working with seniors is challenging but rewarding. I payed quite a lot efforts to keep pace with senior. As a result, I acquired not only LDA and BTM model but also some important math knowledge such as Dirichlet distribution as well as Markov chain Monte Carlo algorithm. Moreover, I progressively mastered the basic research methods and derived much satisfaction from this research. It was pretty exciting time when I saw the running time decreasing reflecting the substantial improvement in efficiency.

This research experience stimulated my great interests in NLP. Therefore, after completing the paper as a co-author, I kept paying close attention to the NLP field and took some online courses for machine learning and deep learning. However, as far as I am concerned, being exposed to different areas is vital for a researcher. Meanwhile, security issue my major focuses on is occurred everywhere and NLP is not an exception. There is a paper published in ACL 2018 proposed a model to solve a relevant problem. Texts convey information about their author and many individuals may wish to keep their information privacy. Genuinely impressed by its broad prospects in applying security knowledge to NLP, I became ambitious to promote the interdisciplinary research in this promising field. Thus, I was supposed to be more proficient in math and security problems. Participating a relevant research was the best way for me.

My great performance in the *Mathematic of Information Security* class earned me an opportunity to join a program called Optimization of Blockchain Confidential Transaction Algorithm. Setting up a Blockchain network using Hyper Fabric platform was the best way to get started in Blockchain, and I implemented the functions of transaction and query. Then I participated in a specific project, where I figured out the process and algorithm of confidential transaction including Pedersen commitments,

Borromean Ring Signatures and Homomorphic Encryption and Range Proof algorithm. Our target was to solve the time and space consuming problem, so I analyzed every step and process during a confidential transaction and explored the possibility to make any improvement. I cooperated with three peers under the guidance of a senior where we talked about the math details, proposed suggestions and programmed the java code to verify the improvement. My security knowledge played the role in practice and I believe it is also very significant in the NLP area such as privacy in NLP tasks.

Then it is time for me to go back to my way on NLP. I find an advisor called Prof. Hai Zhao in the Department of computer science, whose research interests are NLP, especially Chinese segmentation. After several discussion with him, I finally decided to build a universal Chinese word segmentation to adapt to different segmentation rules. I trained and tested the Fast and Accurate Neural Word Segmentation for Chinese(based on the work presented in ACL 2017) with different corpus and then analyzed the causes and types of errors using linguistic and statistical knowledge. And then I excluded the error when training and testing in the same domain. When I came across the problem of recognizing Chinese name and name entity of different lengths, I read a lot of relevant papers and tried to find a simple way to establish rules and found embedding is an appropriate way. Then I reproduced the work that Modeling Named Entity Embedding Distribution into Hypersphere and applied it to Named Entity Recognition (NER) independent of open codes. In the next stage of work I will use NER and semantic analysis to improve the accuracy of the neural word segmentation by rebuilding the decode layer.

I think NLP is a mixture of computer science, linguistics, and statistics to create systems that can cope with the richness and subtlety of human language. It requires both a strong computational background and a deep interest in human language. I have accumulated substantial programming ability by a large amount of coursework and research experience in NLP. My background in security can contribute to your program. I want to receive professional guidance for my further research and your program hits me, especially research on NLP such as Detecting and Avoiding Discrimination in NLP systems, Learning Joint Prediction Models with Auxiliary Supervision, conducted by Professor Kai-Wei Chang, largely overlaps with my long-lasting interests and previous research experiences. Thus, I sincerely hope that I will be offered the opportunity to pursue my graduate study in the Department of Computer Science at UCLA.