Essay

I want to pursue a Ph.D. degree in Computer Science, and my primary research interest is in the intersection area of data mining with natural language processing, machine learning or deep learning. My initial passion and interest in data mining was inspired by an undergraduate course.

My undergraduate study at Hainan University was full and tense. In my sophomore year, I transferred my major to computer science due to my interest in programming. Then I finished all the required courses within two years and kept my GPA in the top 3% ranking. Among all the courses, "data warehouse and data mining" attracted me the most. I was always wondering how people could get valuable guidance from a series of seemingly messy numbers. This curiosity persuaded me to participate in the American Mathematical Contest in Modeling. As the captain of a team of three girls, I mainly charged with analyzing data, building mathematical models, and setting paper's writing logic. We used prediction and evaluation models to determine whether an economical opportunity exists to remove the space debris and provided a recommended solution to remove the debris. After a month of training and four days of tough contests, we won the second prize. This experience greatly improved my confidence and inspired me to further explore interesting problems in data mining.

With the honor of waiving the entrance exam, I was admitted to the Department of Computer Science in Harbin Institute of Technology, Shenzhen, which is famous for its artificial intelligence research in China. Following my passion, I joined Professor Philippe Fournier-Viger's group to continue my study in data mining. My research projects focus on high utility pattern mining, which is about mining patterns that have great importance in the database. Observing the limitation that existing models often ignore the cost (e.g., effort, resources) required for obtaining the high utility, I added a cost model into the utility model to mine the patterns that can generate a high utility but consume a relatively small cost. First, I designed innovative pruning measures and a buffer structure to improve the algorithm's efficiency under the huge search space. The experiments showed that the execution time is improved by up to 10 times and the memory usage is reduced by up to 20%. Second, two reasonable statistical measures were proposed to rank the discovered patterns in different scenarios. The proposed models were applied to e-learning datasets to provide insights about time-efficient ways of utilizing learning materials to obtain good performance.

Last August, I presented my work at the Utility-Driven Mining and Learning Workshop, held in conjunction with the 24th KDD conference in London. It was an excellent experience. In the conference, I learned about many interesting research topics and exciting applications in the area of natural language processing, machine learning, and deep learning. It also gave me an opportunity to discuss my research problems with several pioneer researchers in data mining. Their inspiring responses encouraged me to further investigate one of the problems. In my current project, I try to

provide guidance for people with different characteristics by implementing mining methods to properly combine heterogeneous data sources. In summary, I need to integrate people's profile and their sequence of activities to learn and mine the best ways of utilizing resources for an individual or a group.

Conducting research at school cannot satisfy my curiosity on understanding data mining's power in solving real-world problems. I was eager to work with researchers in industry and apply my knowledge on data mining to solve challenging problems with real data. In the summer of 2019, I joined the 2012 Lab of Huawei Technology as an intern. During my internship, I designed and implemented a framework to discover erroneous patterns from the CAD drawing data. This system was deployed in production and greatly improved the CAD design team's efficiency by reducing many manual efforts. This successful internship experience brought me another opportunity with Huawei. Now I am participating in a project of alarm compression and causal root location in the telecommunication network with Noah' Ark Lab. The main focuses of the project are using dynamic attributes graph as an approach to mine important sequences of alarms that have high priorities and generating potential correlated sequence rules of alarms to identify the root-cause alarms.

Renowned for its prominent faculty and state-of-the-art research facilities, the graduate program at The University of Texas at Dallas is an ideal platform for me to advance my pursuit in data mining research. In particular, there are several professors whose projects are especially appealing to me: Prof. Jessica Ouyang (natural language processing on summarization, paraphrasing, and style transfer); Prof. Vincent Ng (unsupervised and semi-supervised machine learning for natural language processing); Prof. Feng Chen (data mining in complex and heterogeneous network data). Reading those professors' project website and seeing the researches' broad applications have convinced me that pursuing Ph.D. degree in your university is a great opportunity to fulfill my long-term professional objectives: keep learning advanced technics, make data mining technology a positive impact on human life and devote myself to build a more intelligent and convenient world. I hope that I have convinced you of my passion and ability, and I sincerely wish to be admitted to your program.