Statement of Purpose

Gerome Yoo
geromeyoo@gmail.com

Motivation

I am motivated to pursue a Ph.D. in Computer Science with the goal of conducting research in the intersection of Machine Learning and data management. My particular areas of fascination encompass applied ML, data cleansing for ML, and efficient data management.

My motivation to engage in research is rooted in good experince of my master's degree and my practical experience in the field. These encounters have revealed my enthusiasm for tackling unsolved and challenging problems in research projects. Moreover, I have acquired essential skills in implementing and effectively communicating ideas to enhance outcomes through my real-world experiences.

I am deeply enthusiastic about Computer Science research due to its inherent ability to unify various disciplines. My initial exposure to Machine Learning through coursework led me to concentrate on the foundational principles that constitute models from a statistical perspective. As I immersed myself in the field of profession, my focus shifted towards the practical aspects of accuracy and cost-effectiveness of making predictions using models. This transition reflects my evolving interest in computer science to address real-world challenges.

Research in Word Sense Disambiguation

In Fall 2018, during my master's program, I undertook a graduation project focused on disambiguating polysemies of the homonym. The approach employed was as follows: I implemented word2vec to generate word embeddings from a thesis abstracts corpus. These embeddings were then utilized to recalculate context embeddings for each instance of the target word by averaging embeddings within specific windows. Subsequently, I clustered the context embeddings of these instances to assess the method's effectiveness, which yielded successful results. However, drawbacks of this approach were the need for an additional dimension to store context embeddings separately from the original word embeddings, also required secondary calculations for every word in the corpus.

Experience in Real World Data

During the spring of 2020, I opted for practical experience in the field to engage with real-world data. In my first year as a Marketing Intelligence Analyst at Hotel Shilla, I gained insights into the critical role of data management. During this period, I carried out fundamental data analysis tasks and tried to implement data pipelines for OLAP system. This experience taught me the significance of data governance, as I encountered challenges in fully leveraging my analytical capabilities. The absence of robust data governance policies and irregularities in data fusion impeded the optimal utilization of my analytical skills.

In the spring of 2021, I secured a position as an Analytics Engineer at Netmarble Company in South Korea. My primary responsibilities involved analyzing and defining key indices for games, as well as constructing data pipelines. This process encompassed collecting raw data from the source and transforming it to generate index dashboards. Additionally, I actively participated in a data migration project, gaining valuable experience in data management and conducting comparative analyses between heterogeneous databases for migration purposes from HIVE to BigQuery. Another significant undertaking was a data model integration project, wherein I developed an integrated data model to enhance the efficiency of processing and storing data from diverse genres of games.

Furthermore, I actively contributed by sharing statistical insights through delivering many game analysis reports and updating the indices based on my experience. My research focused on employing statistical methods to generalize game user segmentation. In an effort to enhance business intelligence, I applied methods such as A/B testing and clustering methods to various games. Additionally, our team conducted seminars on basic statistics to facilitate understanding and utilization of statistical analysis results. Currently, my ongoing project involves constructing time series models to analyze the impact of game updates.

Future Work

Prior to embarking on my PhD journey, I plan to enroll in theoretical foundational courses on platforms like Coursera. Recognizing that my undergraduate major did not align with computer science, I recognize the significance of establishing a robust theoretical foundation. While field experience is valuable for building proficiency as a developer, I acknowledge its limitations in cultivating the skills necessary for effective academic research. Then during my PhD, I want to enjoy working on challenging problems which both advance theoretical understanding and improve developer productivity. I am dedicated to immersing myself in research endeavors that not only push the boundaries of existing knowledge but also offer practical solutions with tangible benefits for the development community.