Statement of Purpose

Sahil Pattni
(sp72@hw.ac.uk)

Applying to MSc. Computer Science at McMaster University

Few topics have fostered such a sense of curiosity, wonder and interest in me as the broad field of computer science has. The process of seemingly arbitrary words on a screen translating to logic amazed me when I first discovered it, and still does as I conclude my undergraduate degree. It is for this reason that I wish to pursue a Master in Computer Science at McMaster University. My research interests include machine learning, data structures and natural language processing.

My motivation for pursuing research relating to machine learning originates from when I built my own neural network from scratch and optimized it using Particle Swarm Optimization for my Biologically Inspired Computation course. It was after I had built this model that I truly felt the power that machine learning can offer, and it was then that I decided I wanted to learn as much as I could about the work that had been done in this domain, and to contribute towards its advance.

I am working on a my undergraduate dissertation under the guidance of **Prof. Neamat El Gayar**, the purpose of which is to derive multi-element association rules by representing - as a graph - the association between product nodes as an edge that connects them, extracting the minimum spanning tree such that only the strongest associations remain, and determining clusters from this resulting minimum spanning tree using Markov Clustering. From each cluster, multi-element rules can be obtained from the edges within the cluster. This is an alternative to the established Apriori algorithm [1]. I am building on the work of **Marucio A. Valle et al** [2], who proposed the initial framework for single element association rule derivation from minimum spanning trees. This dissertation has allowed me to explore my interest in data structures and graph theory, but it has come with its share of technical challenges, all of which I enjoyed solving. Transforming the large dataset to my requirements computationally infeasible if performed linearly with a complexity of $O(n^2)$, due to the millions of entries in the dataset. I was able to use the knowledge I had gained in courses such as Operating Systems and Concurrency to develop a multi-core, parallelized solution to meet these challenges.

In the four years it took me to achieve a Bachelor in Science with Honors at Heriot-Watt University, I have acquired an understanding of this scientific discipline that has only fueled my curiosity and interest further. I have worked on several interesting projects - from using C and ARM Assembly to interact with hardware connected to a Raspberry Pi to writing our own parsers for compilers, and classifying the sentiments of tweets on a five point scale using LSTM RNNs. It was the latter - build for my Applied Text Analytics course - that sparked my interest in natural language processing (NLP). Diving deeper into the domain, I found generative language models such as GPT-2 to be utterly fascinating, coming one step closer towards bridging the gap that divides man and machine. Towards the end of the second semester of my third academic year, the world was hit with the Covid pandemic, and as universities, schools and businesses were closed under a government-mandated lockdown, my university did not have enough time to transition to an online examination system. Therefore, for courses whose grades were primarily determined through exams, a P (passing) grade was awarded since the exams could not be administered. For courses that were marked through courseworks,

the regular grading system was used as all courseworks had been submitted by the time the lockdown was put in effect.

Outside the academic environment, my passion for computer science has driven me to work on several projects on the side; most recently, I worked with two colleagues to develop a cross-platform mobile application for a hackathon that allows non-profits to crowd-source their funding. I am also currently developing an automated cryptocurrency trading system that uses machine learning[1] to determine when to execute trades. Laying the groundwork for this project has also fostered an interest in stochastic systems in general, and how machine learning could hold the potential to approximate them.

At McMaster University, due to the plethora of available opportunities, I believe that I will be able to gain higher exposure and with it, a deeper understanding of the various problems in the field. Being part of such an elite and diverse group would foster intellectual growth and enable me to work to my potential. It will be a privilege for me to spend a fruitful and rewarding time at the Hamilton campus.

References

[1] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," *Proceedings of the 20th International Conference on Very Large Databases*, pp. 487–489, 1994.

[2] M. A. Valle, G. A. Ruz, and R. Morrśs, "Market basket analysis: Complementing association rules with minimum spanning trees," *Expert Systems with Applications*, vol. 97, pp. 146–162, May 2018. DOI: 10.1016/j.eswa.2017.12.028.

[1] I am still comparing different classifiers and architectures to determine which one would best be suited for this task