Title: Artificial Intelligence for High Performance Job Scheduling

Q.1 I hope to learn about the inner workings of various artificial intelligence and machine learning algorithms. I hope to then apply that knowledge to create an artificial intelligence which will be able to predict certain information about HPC jobs run on the LC campus cluster (or elsewhere). After that, I would like to use that artificial intelligence to generate guesses about the performance of those jobs on the HPC cluster and evaluate its accuracy. If the AI is accurate enough, I would like to then expand it to generate an ordering of jobs such that all of the jobs finish as fast as possible. This is an extremely complex goal, computationally speaking, and it may not be achieved in full this semester. If possible, I would like to deploy an instance of the AI to the LC HPC system, in order to test it in a semi-production environment.

Q.2 My specific tasks will essentially be as follows: First, investigate existing technologies for AI model training with an informal literature review and find what existing platforms/tools/systems I should be looking at. The two that come to my mind already are PyTorch and TensorFlow. I also want to learn more about the factors that I should take into account with performance estimation. I would like to explore the classical methods of generating performance estimates of HPC jobs, including mathematical and algorithmic analysis and heuristic profiling. Second, create a tool that is able to use those tools/platforms/systems to generate an estimate of how long a job should take, how much memory it should use, etc. Third, test this “baseline” tool on a variety of types of HPC jobs (bioinformatics, text processing, physical simulation, etc.) and after I am able to verify that this tool works properly, then I will start making modifications that allow the generation of scheduling weights (the concept of deciding on the most intelligent order in which to run tasks, in order to minimize overall execution time). After that, I would like to design

a “production” version of my tool which can be made public for people inside and outside of the college to use if they wish.

Q.3 I will have meetings with my advisor (Jens Mache) approximately once a week during which we will discuss progress, challenges, and possible new directions if a change is necessary. My performance will be evaluated by how much of the things I set out to accomplish are designed and implemented (Note that some of the things in question 2 are very difficult and may not work out exactly as intended). Additionally, I will be evaluated on my new knowledge of AI and Machine Learning practices. At the end of this independent study, I hope to have some piece of software to show for my work, and a paper which I can submit to a conference.

Q.4 I will have a piece of software to show for my work, which means that it could be used to evaluate my work. Additionally, I would like to do a short write up (about 2 pages in ACM technical paper format) about my work if it is completed and works well. That paper, or an extended version of it, could be submitted to any of a number of CS conferences where I could go to get opinions from other CS researchers on my work.

Q.5 I will meet with my advisor in person approximately once a week or once every 2 weeks if not much progress has occurred in the last week.