HQP Training Plan

Training Philosophy

Supervision. I have worked closely with both graduate students and many undergraduate RAs (almost 30 from 2009-2017). My approach to mentoring supports their success through close supervision, matching the student's interests, group collaboration, and funding that frees up their time to focus on research. I hold at least weekly meetings with my students. Almost all my trainees have been coauthors, including undergraduates. For maximum motivation, I select projects according to the student interests. For example for undergraduate students who are interested in pursuing graduate school, I suggest a project that is close to publication to support their future application.

Lab environment. I strive to create a collaborative lab environment for research issues, software development, and general success in science. Trainees contribute to an extensive code base my lab has built up. Collaborative software development practices in my lab include repository use (github) and pair programming. For example, my students Yan Sun and Qing Li worked as a pair with a larger group of graduate students. One of my ideas for fostering collaboration and cohesion is a weekly coffee social, where I take my students out for an informal gathering with a looser atmosphere than the formal lab meetings. Over time I have supervised generations of trainees, where senior undergraduate students will encourage junior students to apply. For example, one of my especially talented RAs was Yuke Zhu (now a Ph.D. student at Stanford). Yuke introduced Zeyu Zhao to Oliver's lab saying "he is like my little brother". Zeyu has done brilliant work in our lab and is now an M.Sc. student at Harvard.

Recruitment. An important goal of my recruitment is gender diversity. I place priority on inviting female students to undertake RAships. My women graduate students are effective recruiters and mentors for all undergraduates but especially female ones. I also participate in international recruitment to bring highly skilled computer scientists to Canada. Several of my best undergraduate RAs have been from China. I also participate in the federal government's Globalink program, supervising students from India, China, and Germany. My current M.Sc. student Yejia Liu was formerly a Globalink student.

In addition to our joint research, extramural experiences will add value for my trainees. These include presentations at conferences, visits to industry collaborators (e.g. Sportloqig in Montreal), co-op (e.g. SAP, Google, Amazon), and research lab internships. For example this year, my Ph.D. student Mahmoud Khademi was an intern at the Microsoft research lab in Cambridge.

Research Training Plan

The main research skills my trainees will develop are in applying advanced data science, machine learning, and artificial intelligence. These are national priority areas for technology, as the federal government has recognized by allocating major funding to AI clusters. Data science skills are in demand in many parts of industry. The trainees will acquire skills in both machine learning and AI that support extracting large-scale information from the worldwide web, an increasingly important resource for many companies. My students also learn collaborative software development, which increases their job prospects in industry even for positions not related to machine learning.

I plan for PhD theses to take 4-5 years and MSc theses 2 years. My research program spans 3 PhD theses and 2 MSc theses. The PhD topics require substantive research into new statistical methods, whereas the MSc topics build heavily on previous research. My research plan aligns with projects for individual theses as follows.

PhD Thesis 1: Develop scalable relational learning for Bayes nets with class hierarchies. Apply the techniques to knowledge graph completion.

MSc Thesis 1: Improve the scalability of our current relational learning system (FactorBase) to 30K or more attribute and link types. This project will be carried out in close collaboration with PhD Thesis 1. The trainee will learn how to support large-scale machine learning, including data structures (e.g. ADtrees), systems resources (e.g. clusters), and software (Spark, Hadoop).

PhD Thesis 2: Leverage relational Bayes net learning for deep neural network learning for relational data with latent features (types, clusters, embeddings) and class hierarchies. Evaluate on knowledge graph completion. This is a PhD level topic because it requires using deep learning with nontraditional (relational) data (relational data) and combining it with other machine learning techniques.

MSc Thesis 2: Leverage relational Bayes net learning for deep learning for relational data with latent features (types, clusters, embeddings). This project will be carried out in close collaboration with PhD Thesis 2. The trainee will learn state-of-the-art relational embedding techniques and knowledge graph completion benchmarks.

MSc Theses 3 + 4: Apply relational Bayesian network learning to relation extraction from text or image data. These students will start in year 3 when our methods will be mature for challenging applications. They will be mentored by the PhD students in the statistical methods. The trainees will acquire knowledge of domains such as text and image data that are important in industry. These are challenging application domains each of which provides enough challenges for two MSc theses. The exact domain will depend on the interest and expertise of the available students.

Colleagues with domain expertise who are interested in collaboration include Sarkar for natural language processing, and Mori for vision.

Undergraduate students will be recruited to assist with different aspects, such as handling datasets, programming (over a longer summer term), and preparing publications.