Research Statement

Lin Ling

My research interests cover convex optimization and distributed algorithms, as well as their applications in statistical machine learning, graph analytics and personalized learning at scale. More specific topics include the design of distributed and iterative algorithms, automated theorem proving, and crowd-sourced data aggregation.

1 Ph.D. Research

1.1 Scalable Automated Theorem Proving in Information Theory by Convex Optimization

Proving and disproving information inequalities is a fundamental task in information theory. When the inequality involves more than a few random variables, it is often tedious or impossible to prove manually. In [1], we propose a linear programming framework to automate the process. By inspecting the Lagrange duality of the optimization problem, we can obtain hints to construct an analytic proof or a counterexample to disprove the inequality. Then in [2, 3], we develop a scalable iterative algorithm to efficiently solve the specific optimization problem we will encounter when proving information inequalities. Our experiments demonstrate that our algorithm gives superior performance (up to ten times faster) than any existing state of the art.

1.2 Human-Assisted Computation for Automated Grading at Scale

Marvin Minsky once advocated that computer scientists have the responsibility to help students "debug" their thinking process. In [4, 5], we develop a statistical machine learning method to automatically grade multiple-choice questions (MCQs) without having access to the correct answers beforehand. The unsupervised algorithm can learn each student's accuracy in answering the questions and adjust the parameters over time, and using the learned parameters it can accurately recover the correct MCQ answers given the answers from a sufficient number of students. We also demonstrate that the same technique can be applied to grade other question types that are seemingly unrelated to MCQs. This research has led to successful knowledge transfer and entrepreneurship opportunities, recognized by the Hong Kong Science Park's prestigious IncuTech incubation program.

1.3 Rank Aggregation from Crowd-Sourced Pairwise Comparisons

Rank aggregation is the process of uncovering the underlying "true" ranking of items from a set of pairwise comparison results. This is a common task in machine learning and has been covered extensively in the literature. However, in large-scale applications the comparisons are often crowd-sourced, where users might produce biased or adversarial comparisons. In November 2019, I was invited by Dr. Yuxin Chen at the Department of Electrical Engineering, Princeton University to work with him on this problem as a visiting research collaborator. By applying the matrix sampling

technique, we develop a spatial algorithm that can accurately recover the true ranking from noisy crowd-sourced data. We are currently working on a conference submission.

2 Future Research Directions

I enjoy doing research that blends human intelligence and the computational power of computers. I plan to undertake in the following long term directions in my future research:

- **Automated reasoning by convex optimization:** By inspecting the dual problems in convex optimization, we can uncover hidden structures and gain additional insights. This allows algorithms to display complex problem-solving skills. Can we apply the same principle to make computers assist humans in proving new theorems and making scientific discoveries? Can we use it in an educational setting to automate grading and provide instant feedback?
- **Statistical machine learning from crowd-sourced data:** Machine learning requires a massive amount of training data, and the data is often crowd-sourced from users so it is inherently biased. How can we design efficient and accurate learning algorithms that are robust enough to deal with biased or even adversarial training data? Can we accurately identify the biased users and minimize their effect in the learning process?
- **Computational tools that facilitate large-scale personalized and remote learning:** One of the U.S. National Academy of Engineering Grand Challenges is to advance personalized learning, and remote learning has become the norm due to the global Covid-19 pandemic. How can we use our mature recommendation systems and the collaborative filtering technique to recommend personalized learning content to students at different levels? Can we develop an algorithm that can automatically grade open-ended questions and mathematical derivations?

My Ph.D. research represents a first step along these directions. In my future research career, I hope to explore the computational edge between human minds and computers, and develop solutions to practical problems that combine the best of the two worlds.

References

[1] S.-W. Ho, L. Ling, C. W. Tan, and R. W. Yeung, "Proving and disproving information inequalities: Theory and scalable algorithms," *IEEE Transactions on Information Theory*, 2020.

[2] L. Ling, C. W. Tan, S.-W. Ho, and R. W. Yeung, "Scalable automated proving of information theoretic inequalities with proximal algorithms," in *2019 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2019, pp. 1382–1386.

[3] C. W. Tan and L. Ling, "Automated reasoning by convex optimization: Proof simplicity, duality and sparsity," in *2020 54th Annual Conference on Information Sciences and Systems (CISS)*. IEEE, 2020, pp. 1–5.

[4] L. Ling and C. W. Tan, "Human-assisted computation for auto-grading," in *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2018, pp. 360–364.

[5] ——, "Social learning network and its applications in large scale online education through chatbot," in *Online Social Networks: Perspectives, Applications and Developments*. Nova Science Publishers, Inc., 2020.