Selective Data Acquisition in Learning and Decision Making Problems

Abstract:
Classical statistics and machine learning posit that data are passively collected, usually assumed to be independently and identically distributed. In modern data science applications, however, many times a data analyst has control over how data are acquired or selected. For example, in simulation/hyper-parameter optimization the input parameter configurations can be adaptively chosen to obtain data resulting from the carefully chosen input parameters. In sequential decision making problems, data such as feedback or utility depend on the particular decisions which can be adaptively and selectively made.

The main topic of this thesis is to study how selective data acquisition techniques can be applied in estimation, optimization and/or decision making problems. Three representative problems are studied, as we explain in more details below:

1. Computationally tractable experimental design, which studies the classical question of (optimal) experimental design in linear and generalized linear models from a computational perspective. We design polynomial-time algorithms with rigorous approximation guarantees in terms of optimality criteria, and show an application to a 3D lightweight structure optimization problem.

2. Sample-efficient query regimes for nonparametric optimization, which tries to understand the most sample efficient regimes to make adaptive queries to a nonparametric function for optimization purposes. We consider three different settings of nonparametric optimization: smooth non-convex functions in low dimensions, high-dimensional convex functions with sparsity structures, and convex function sequences that evolve slowly over time.

3. Dynamic assortment optimization, which studies the classical assortment optimization problem in revenue management from a dynamic perspective, by combining statistical estimation of customers’ utility models and optimization of assortments based on estimated utilities into a unified theoretical framework. We characterize through statistical minimaxity the fundamental information-theoretic limits of these problems as well as notions of optimality of our proposed methodologies. On the practical side, we demonstrate industrial engineering and/or operations management applications such as lightweight structural design, dynamic pricing and assortment planning.

Speaker:
Yining Wang

Thesis Committee:
Aarti Singh (Chair)
Sivaraman Balakrishnan
Larry Wasserman
Robert Nowak (University of Wisconsin Madison)

March 22, 2019
12:00pm
GHC 4405

Link to draft document:
https://www.dropbox.com/s/rbsrh9ywwhi1pzq/defense.pdf?dl=0