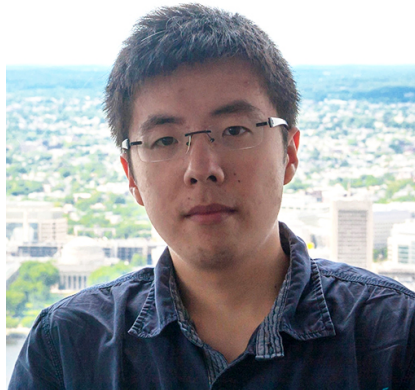


University of Chicago Computer Science Department

PRESENTS:



Hongzi Mao
PhD Student
MIT

Title: Learning to Schedule from Experience

Abstract: Efficiently scheduling data processing jobs on distributed compute clusters currently requires hand-tuned complex algorithms. In this work, we show that modern machine learning techniques can generate highly-efficient scheduling policies automatically. Decima uses reinforcement learning (RL) and neural networks to learn workload-specific scheduling algorithms without any human instruction beyond a high-level objective. However, off-the-shelf RL techniques cannot handle the complexity and scale of the scheduling problem. To build Decima, we had to develop new representations for jobs' dependency graphs, design scalable RL models, and invent new RL training methods to handle continuous stochastic job arrivals. Our prototype integration with Spark on a 25-node cluster shows that Decima improves average job completion time by at least 21% over human-engineered scheduling heuristics, achieving up to 2× improvement during periods of high cluster load.

Bio: Hongzi Mao is a PhD student at the Electrical Engineering and Computer Science Department of MIT advised by Mohammad Alizadeh. His research interests focus on applying reinforcement learning to build and optimize data-driven networked resource management systems. His work also contributes to the development of new machine learning algorithms for broader applications beyond computer systems. He is a recipient of ICML'19 best workshop paper award, Qualcomm Innovation Fellowship and Andrew (1956) and Erna Viterbi Fellowship.

Host: Heather Zheng and Junchen Jiang

Monday, September 23, 1 – 2 pm
JCL Room 223