

## CCST - CBST Seminar

CENTER FOR CATALYTIC SCIENCE & TECHNOLOGY | CENTER FOR BIOMANUFACTURING SCIENCE & TECHNOLOGY



# Kristen Severson

Ph.D. candidate  
Massachusetts Institute of Technology

**Tuesday, February 13, 2018 | 2:00 PM | 366 Colburn Lab**

## Data-driven modeling in complex systems

Chemical and biological systems are increasingly implemented with advanced sensor systems that collect large amounts of data. For example, a single microarray can measure thousands of genes and a typical offshore oil platform generates 1 to 2 TB of data per day. New algorithms are needed to efficiently and effectively use these datasets to increase predictive capability and improve system understanding. In this talk, algorithmic advances to bridge the gap between data and system insights are addressed. Two case studies are presented. The first case study considers how to learn interpretable models from high-throughput, high-dimensional biological assays where some data may be missing. A new approach for building a classification model is presented and demonstrated using two different types of datasets: gene expression microarrays and reverse phase protein arrays. The resulting models use very few measurements and are at least as accurate as competing approaches. The second case study analyzes early prediction of lithium-ion battery life cycle. Analysis is performed using high-throughput cycling data. The resulting model outperforms all existing models.