

Computational mathematics for large-scale data analysis

Alexander Gray*

Starting with examples of challenges in astronomy, this talk will consider a selected survey of the computational mathematics, both current and future, which can be brought to bear to the problem of analyzing massive datasets. First, what are the data analysis (machine learning and statistics) methods we want to perform on large datasets in general, and how much do they cost? Then, what is a short-list of the basic mathematical operations underlying these methods, or bottleneck subroutines for which we can focus on developing fast algorithms? I will discuss aggregations, N-body-like problems, graphical model inference, linear algebra, and optimization; for each I will discuss the main mathematical challenge and highlight at least one current interesting solution. I will close with some general solution avenues for the future, including the deep interplay between statistics and computation, visual analytics, and algorithm/software synthesis.

*College of Computing, Georgia Institute of Technology