COLLOQUIUM

Eitan Levin

Applied and Computational Mathematics
California Institute of Technology

Any-Dimensional Data Science

THURSDAY, January 22nd at 4:00 PM

Jones 303, 5747 S. Ellis Ave. Chicago, IL 60637

ABSTRACT

Many applications throughout data science require methods that are well-defined and performant for problems or data of any size. In machine learning, we are given training data from which we wish to learn algorithms capable of solving problems of any size. In particular, the learned algorithm must generalize to inputs of sizes that are not present in the training set. For example, algorithms for processing graphs or point clouds must generalize to inputs with any number of nodes or points. A second challenge pertaining to any-dimensionality arises in applications such as game theory or network statistics in which we wish to characterize solutions to problems of growing size. Examples include computing values of games with any number of players, or proving moment inequalities for random vectors and graphs of any size. From an optimization perspective, this amounts to deriving bounds that hold for entire sequences of problems of growing dimensionality. Finally, in applications involving graph-valued data, we wish to produce constant-sized summaries of arbitrarily-large networks that preserve their essential structural properties. These summaries can then be used for efficiently testing properties of the underlying large network, e.g., testing for the presence of hubs is of interest in massive biological and traffic networks. We develop a unified framework to tackle such any-dimensional problems by using random sampling maps to compare and summarize objects of different sizes. Our methodology leverages new de Finetti-type theorems and the recently-identified phenomenon of representation stability. We illustrate the resulting framework for any-dimensional problems in several applications.