

Abstract:

Increasingly often, data mining has to learn predictive models from big data, which may have many examples or many input/output dimensions and may be streaming at very high rates. Contemporary predictive modelling problems may also be complex in a number of other ways: they may involve (a) structured data, both as input and output of the prediction process, (b) incompletely/partially labelled data, and (c) data placed in a spatio-temporal or network context.

The talk will first give an introduction to the different tasks encountered when learning from big and complex data. It will then present some methods for solving such tasks, focusing on structured-output prediction, semi-supervised learning (from incompletely annotated data), and learning from data streams. Finally, some illustrative applications of these methods will be described, ranging from genomics and medicine to image annotation and space exploration.