

Lecturer: Koby Crammer
Department of Computer and Information Science, University of Pennsylvania

Title: Learning From Related Sources

Abstract:

We often like to build a model for one scenario based on data from similar or nearby cases. For example, consider the problem of building a model which predicts a sentiment about books from short reviews, using reviews and sentiment of DVDs. Another example is of learning movies preference for one viewer from ratings provided by other similar users. There is a natural tradeoff between using data from more users and using data from only similar users.

In this talk, I will discuss the problem of learning good models using data from multiple related or similar sources. I will present a theoretical approach which extends the standard probably approximately correct (PAC) learning framework, and show how it can be applied in order to determine which sources of data should be used and how. The bounds explicitly model the inherit tradeoff between building a model from many but inaccurate data sources or building it from a few accurate data sources. The theory shows that optimal combinations of sources can improve performance bounds on some tasks.