

Workshop in Methods

Probabilistic Topic Models and User Behavior

Dr. David Blei

Professor of Statistics and Computer Science, Columbia University

Topic modeling algorithms analyze a document collection to estimate its latent thematic structure. However, many collections contain an additional type of data: how people use the documents. For example, readers click on articles in a newspaper website, scientists place articles in their personal libraries, and lawmakers vote on a collection of bills. Behavior data is essential both for making predictions about users (such as for a recommendation system) and for understanding how a collection and its users are organized.



I will review the basics of topic modeling and describe our recent research on collaborative topic models, models that simultaneously analyze a collection of texts and its corresponding user behavior. We studied collaborative topic models on 80,000 scientists' libraries from Mendeley and 100,000 users' click data from the arXiv. Collaborative topic models enable interpretable recommendation systems, capturing scientists' preferences and pointing them to articles of interest. Further, these models can organize the articles according to the discovered patterns of readership. For example, we can identify articles that are important within a field and articles that transcend disciplinary boundaries.

Friday, October 16, 2015, 2-4pm

Social Science Research Commons Grand Hall
Woodburn Hall 200