



THE UNIVERSITY OF
CHICAGO

Department of Statistics

MASTER'S THESIS PRESENTATION

PAUL HIVELY

Department of Statistics
The University of Chicago

Crowdsourced Statistics Advice: Topic Modeling
stats.stackexchange.com

FRIDAY, May 5, 2017, at 10:00 AM
Jones 226, 5747 S. Ellis Avenue

ABSTRACT

Topic models are hierarchical Bayesian models used to discover latent semantic structure within collections of documents, allowing them to be reduced from millions of words to a few dozen interpretable topics. This paper presents three closely related methods: latent Dirichlet allocation, correlated topic models, and structural topic models. I discuss the estimation challenges associated with topic modeling and compare the three methods by analyzing a collection of 182,308 posts contributed by the general public to the statistics and machine learning community website stats.stackexchange.com.

For information about building access for persons with disabilities, please contact Laura Rigazzi at 773.702-0541 or send an email to lrigazzi@galton.uchicago.edu. If you wish to subscribe to our email list, please visit the following web site: <https://lists.uchicago.edu/web/arc/statseminars>.