

# Interactive topic modeling

Yuening Hu · Jordan Boyd-Graber · Brianna Satinoff ·  
Alison Smith

Received: 17 August 2012 / Accepted: 20 August 2013  
© The Author(s) 2013

**Abstract** Topic models are a useful and ubiquitous tool for understanding large corpora. However, topic models are not perfect, and for many users in computational social science, digital humanities, and information studies—who are not machine learning experts—existing models and frameworks are often a “take it or leave it” proposition. This paper presents a mechanism for giving users a voice by encoding users’ feedback to topic models as correlations between words into a topic model. This framework, interactive topic modeling (ITM), allows untrained users to encode their feedback easily and iteratively into the topic models. Because latency in interactive systems is crucial, we develop more efficient inference algorithms for tree-based topic models. We validate the framework both with simulated and real users.

**Keywords** Topic models · Latent Dirichlet Allocation · Feedback · Interactive topic modeling · Online learning · Gibbs sampling

---

Editors: Winter Mason, Jennifer Wortman Vaughan, and Hanna Wallach.

Y. Hu (✉) · B. Satinoff · A. Smith  
Computer Science, University of Maryland, College Park, Maryland, USA  
e-mail: [ynhu@cs.umd.edu](mailto:ynhu@cs.umd.edu)

B. Satinoff  
e-mail: [bsonrisa@cs.umd.edu](mailto:bsonrisa@cs.umd.edu)

A. Smith  
e-mail: [amsmit@cs.umd.edu](mailto:amsmit@cs.umd.edu)

J. Boyd-Graber  
iSchool and UMIACS, University of Maryland, College Park, Maryland, USA  
e-mail: [jbg@umiacs.umd.edu](mailto:jbg@umiacs.umd.edu)