

David Mimno

Department of Information Science
Cornell University
301 College Ave.
Ithaca, NY
mimno@cornell.edu

Education

University of Massachusetts, Amherst, 2005-2012

- Department of Computer Science.
- Advisor: Andrew McCallum
- Research Interests: Machine Learning, NLP, Text Mining, Topic Models
- MSc Jan 2009
- PhD Feb 2012, Thesis title: *Topic Regression* (defended Sept 13, 2011)

Swarthmore College, 1995-1999

- Bachelor of Arts with High Honors (Classics and Computer Science)

Employment

- Assistant Professor, Cornell University, July 2013–present
- CRA Computing Innovation Fellow, Princeton University, Oct 2011–June 2013
- Senior Research Assistant, Department of Computer Science, Princeton University, 2010–2011
- Research Assistant, Department of Computer Science, University of Massachusetts, Amherst, 2005–2010
- Head Programmer, Perseus Project, Tufts University, 2002–2005

Publications

Peer-reviewed Conference Papers

- A Practical Algorithm for Topic Modeling with Provable Guarantees. Sanjeev Arora, Rong Ge, Yonatan Halpern, David Mimno, Ankur Moitra, David Sontag, Yichen Wu, Michael Zhu. ICML, 2013, Atlanta, GA.
- Scalable Inference of Overlapping Communities. Prem Gopalan, David Mimno, Sean Gerrish, Michael J. Freedman, David Blei. NIPS, 2012, Lake Tahoe, NV.
- Sparse Stochastic Inference for Latent Dirichlet Allocation. David Mimno, Matthew Hoffman and David Blei. ICML, 2012, Edinburgh, Scotland.
- Topic Models for Taxonomies. Anton Bakalov, Andrew McCallum, Hanna Wallach, and David Mimno. Joint Conference on Digital Libraries (JCDL) 2012, Washington, DC.
- Reconstructing Pompeian Households. David Mimno. UAI, 2011, Barcelona, Spain.
- Bayesian Checking for Topic Models. David Mimno, David Blei. EMNLP, 2011, Edinburgh, Scotland.

- Optimizing Semantic Coherence in Topic Models. David Mimno, Hanna Wallach, Edmund Talley, Miriam Leenders, Andrew McCallum. EMNLP, 2011, Edinburgh, Scotland.
- Rethinking LDA: Why Priors Matter. Hanna Wallach, David Mimno and Andrew McCallum. NIPS, 2009, Vancouver, BC.
- Polylingual Topic Models. David Mimno, Hanna Wallach, Jason Naradowsky, David Smith and Andrew McCallum. EMNLP, 2009, Singapore.
- Evaluation Methods for Topic Models. Hanna Wallach, Iain Murray, Ruslan Salakhutdinov and David Mimno. ICML, 2009, Montreal, Quebec.
- Efficient Methods for Topic Model Inference on Streaming Document Collections. Limin Yao, David Mimno and Andrew McCallum. KDD, 2009, Paris, France.
- Topic Models Conditioned on Arbitrary Features with Dirichlet-multinomial Regression. David Mimno and Andrew McCallum. UAI, 2008, Helsinki, Finland.
- Expertise Modeling for Matching Papers with Reviewers. David Mimno and Andrew McCallum. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD) 2007, San Jose, CA.
- Mixtures of Hierarchical Topics with Pachinko Allocation. David Mimno, Wei Li and Andrew McCallum. International Conference on Machine Learning (ICML) 2007, Corvallis, OR.
- Mining a digital library for influential authors. David Mimno and Andrew McCallum. Joint Conference on Digital Libraries (JCDL) 2007, Vancouver, BC, Canada.
- Organizing the OCA: Learning faceted subjects from a library of digital books. David Mimno and Andrew McCallum. Joint Conference on Digital Libraries (JCDL) 2007, Vancouver, BC, Canada.
- Beyond Digital Incunabula: Modeling the Next Generation of Digital Libraries. Gregory Crane, David Bamman, Lisa Cerrato, Alison Jones, David Mimno, Adrian Packel, D. Sculley, and Gabriel Weaver. European Conference on Digital Libraries (ECDL) 2006, Alicante, Spain.
- Bibliometric Impact Measures Leveraging Topic Analysis. Gideon Mann, David Mimno and Andrew McCallum. Joint Conference on Digital Libraries (JCDL) 2006, Chapel Hill, NC.
- Finding a Catalog: Generating Analytical Catalog Records from Well-structured Digital Texts. David Mimno, Alison Jones and Gregory Crane. Joint Conference on Digital Libraries (JCDL) 2005, Denver, CO.
- Services for a Customizable Authority Linking Environment. Mark Patton and David Mimno. demonstration at Joint Conference on Digital Libraries (JCDL) 2004, Tucson, AZ.
- Towards a Cultural Heritage Digital Library. Gregory Crane, Clifford E. Wulfman, Lisa M. Cerrato, Anne Mahoney, Thomas L. Milbank, David Mimno, Jeffrey A. Rydberg-Cox, David A. Smith, and Christopher York. Joint Conference on Digital Libraries (JCDL) 2003, Houston, TX.

Journal Articles

- Significant themes in 19th-century literature. Matthew Jockers and David Mimno. *Poetics*. December, 2013.
- Database of NIH grants using machine-learned categories and graphical clustering. Edmund M Talley, David Newman, David Mimno, Bruce W Herr II, Hanna M Wallach, Gully A P C Burns, A G Miriam Leenders and Andrew McCallum, *Nature Methods*, Volume 8(7), June 2011, pp. 443-444.

- Computational Historiography. David Mimno. *ACM J. of Computing in Cultural Heritage*. 5, 1, Article 3 (April 2012), 19 pages.
- Classics in the Million Book Library. Gregory Crane, Alison Babeu, David Bamman, Thomas Breuel, Lisa Cerrato, Daniel Deckers, Anke Luedeling, David Mimno, Rashmi Singhal, David A. Smith, Amir Zeldes. *Digital Humanities Quarterly* 3(1), Winter 2009.
- Hierarchical Catalog Records: Implementing a FRBR Catalog. David Mimno, Alison Jones and Gregory Crane. *DLib*. October 2005.

Peer-reviewed Workshop Papers

- Measuring Confidence in Temporal Topic Models with Posterior Predictive Checks. David Mimno and David Blei. *Computational Social Science Workshop, NIPS 2010*, Whistler, BC.
- Reconstructing Pompeian Households. David Mimno. *Applications of Topic Models Workshop, NIPS 2009*, Whistler, BC.
- Gibbs Sampling for Logistic Normal Topic Models with Graph-Based Priors. David Mimno, Hanna Wallach and Andrew McCallum. *Workshop on Analyzing Graphs, NIPS 2008*, Whistler, BC.
- Community-based Link Prediction with Text. David Mimno, Hanna Wallach, and Andrew McCallum. *Statistical Network Modeling Workshop, NIPS 2007*, Whistler, BC.
- Probabilistic Representations for Integrating Unreliable Data Sources. David Mimno, Andrew McCallum and Gerome Miklau. *IIWeb workshop at AAAI 2007*, Vancouver, BC, Canada.

Service

Executive committee, Princeton University Digital Humanities Initiative.

Organizer: Topic Models: Computation, Application, and Evaluation. NIPS workshop, December 2013.

Organizer and Co-chair: Fourth Northeast Student Colloquium on Artificial Intelligence (NESCAI) 2010, Amherst, MA

- Organized a three-day, 100+ participant, peer-reviewed conference.
- Coordinated industry funding from Facebook and Yahoo.

Program committee: KDD 2012; JCDL 2008–12; NESCAI 2008

Reviewer: JMLR, IEEE Transactions on the Web, IEEE Transactions on Data and Knowledge Engineering, KDD, NIPS, ICML, UAI, ACL, HLT-NAACL

Chief maintainer: MACHINE Learning for Language Toolkit (MALLET).

- Determined development priorities
- Produced documentation and sample code
- Coordinated releases
- Answered user questions

Coordinator, Princeton Machine Learning Lunch. 2010–present.

Coordinator, UMass Machine Learning and Friends Lunch. 2008–2010.

Awards and Grants

CRA Computing Innovation Fellow. Oct 2011–June 2013.
Google Digital Humanities Research grant. The Open Encyclopedia of Classical Sites. 2010.
Mellon Foundation grant. Cybereditions. 2008.
NSF Fellowship, honorable mention, 2005.

Invited Talks

- Digital Humanities Hackathon. University of Wisconsin, Madison, Nov 2013.
- Topic Modeling with MALLET. Rice University, Apr 2013.
- Implementations and Infrastructure for Topic Modeling in the Humanities. Maryland Institute for Technology in the Humanities, University of Maryland, Nov. 2012.
- Statistical Inference. Mid-Atlantic Student Colloquium, University of Maryland, Oct. 2012.
- Finding Thousands of Topics in Millions of Books. Yale University, Dept of Computer Science, Sept. 2012
- The Library as Dataset: Machine Learning at Million-book Scale. Yale University, Sterling Library, Sept. 2012
- Topic modeling with MALLET. Tufts University, Aug 2012.
- The Princeton Digital Humanities Initiative. Princeton University, The Index of Christian Art, July 2012.
- An Introduction to MALLET. Washington University, St. Louis, Dept. of Germanic Studies. Mar 2012.
- Topic Modeling. Mid-Atlantic Student Colloquium, Johns Hopkins University, Fall 2011.
- Computational historiography in a century of classics journals. Maryland Institute for Technology in the Humanities Digital Dialog, University of Maryland, Sept 2011.
- Bayesian checking for topic models. New Directions in Text Analysis Workshop. Harvard University, May 2011.
- Verification of topic models. Columbia, April 2011.
- Topic models we can believe in. Rutgers, January 2011.
- NEH Institute: Networks and Network Analysis for the Humanities. UCLA, August 2010.
- Fast Gibbs sampling for topic models. ICML workshop on topic modeling, Haifa, Israel, June 2010.
- Tutorials: An Introduction to MALLET. University of Maryland, College Park / Johns Hopkins University, Feb 2010.
- Regression-based Topic Models. Google Pittsburgh, October 2009.
- Tutorial: An Introduction to Logistic Regression. Research Experience for Undergraduates. UMass, Amherst, July 2009.

- Text Mining for Humanistic Scholarship. New Directions in Text Analysis Workshop. Harvard University, May 2009.
- Classics in the Million Book Digital Library. Imperial College London/Humboldt University, Berlin, Jan 2008.