David Mimno

Department of Information Science
Gates Hall
Cornell University
Ithaca, NY 14853
(607) 255-8919
mimno@cornell.edu

Education

University of Massachusetts, Amherst, 2005-2012

- Department of Computer Science. Advisor: Andrew McCallum
- 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

- Associate Professor, Cornell University, July 2019–present
- Assistant Professor, Cornell University, July 2013–July 2019
- 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, Univ. of Massachusetts, Amherst, 2005–2010
- Head Programmer, Perseus Project, Tufts University, 2002–2005

Publications

Journal Articles

- Computational Cut-Ups: The Influence of Dada. Laure Thompson and David Mimno. The Journal of Modern Periodical Studies vol. 8, no. 2, 2018, pp. 179?195.

- Evaluating the Stability of Embedding-based Word Similarities. Maria Antoniak and David Mimno. TACL (6) 2018, 107–119.

- Comparing Grounded Theory and Topic Modeling: Extreme Divergence or Unlikely Convergence? Eric P. S. Baumer, David Mimno, Shion Guha, Emily Quan, Geri K. Gay. JASIST 68(6) 2017.

- Applications of Topic Models. Jordan Boyd-Graber, Yuening Hu, and David Mimno. Foundations and Trends in Information Retrieval, Now Publishers. 2017.

- 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. Communications of the ACM. 2017.

- Posterior predictive checks to quantify lack-of-fit in admixture models of latent population structure. David Mimno, David M Blei, Barbara E Engelhardt. Proceedings of the National Academy of Sciences. 112(26):E3341 – 50, 2015.

- How Social Media Non-use Influences the Likelihood of Reversion: Self Control, Being Surveilled, Feeling Freedom, and Socially Connecting. Eric P. S. Baumer, Shion Guha, Emily Quan, David Mimno, and Geri Gay. Social Media and Society, 2015.

- Care and Feeding of Topic Models: Problems, Diagnostics, and Improvements. Jordan Boyd-Graber, David Mimno, and David Newman. In Handbook of Mixed Membership Models and Their Applications, CRC/Chapman Hall, 2014.

- 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 Conference Papers

- Authorless Topic Models: Biasing Models Away from Known Structure. Laure Thompson and David Mimno. COLING, 2018. [Winner: Best NLP Engineering Experiment]

- Quantifying the Visual Concreteness of Words and Topics in Multimodal Datasets. Jack Hessel, David Mimno, and Lillian Lee. NAACL 2018.

- Quantifying the Effects of Text Duplication on Semantic Models. Alexandra Schofield, Laure Thompson, and David Mimno. EMNLP 2017

- The Strange Geometry of Skip-Gram with Negative Sampling. David Mimno and Laure Thompson. EMNLP 2017

- Beyond Exchangeability: The Chinese Voting Process. Lee, Moontae, Seok Hyun Jin, and David Mimno. NIPS 2016, Barcelona, Spain.

- Robust Spectral Inference for Joint Stochastic Matrix Factorization. Moontae Lee, David Bindel, and David Mimno. NIPS, 2015, Montreal, QC, Canada.

- Evaluation methods for unsupervised word embeddings. Tobias Schnabel, Igor Labutov, David Mimno, and Thorsten Joachims. EMNLP, 2015, Lisbon, Portugal.

- Low-dimensional Embeddings for Interpretable Anchor-based Topic Inference. Moontae Lee and David Mimno. EMNLP 2014, Doha, Qatar.

- The Telltale Hat: LDA and Classification Problems in a Large Folklore Corpus. Peter Broadwell, David Mimno, and Timothy Tangherlini. DH 2014, Lausanne, Switzerland.

- 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.

Peer-reviewed Workshop Papers

- ○ The Sheepdog and the Telescope. David Mimno. NeurIPS workshop on Critiquing and Correcting Trends in Machine Learning, 2018.

- ○ Bayesian inference for Poisson community models. David Mimno, Prem Gopalan, David Blei. NIPS workshop on Variational Inference.

- ○ 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

Area chair, ICML 2015–2019, NeurIPS 2016–2018, ACL 2017, NAACL 2015.

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.

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

Awards and Grants

Zillow Research Grant. 2019.
NSF CAREER. 2017.
Sloan Research Fellow. 2016.
NSF CHS Small: Advancing the human work of data analytics. Sept 2015.
CRA Computing Innovation Fellow. Oct 2011–June 2013.
Google Digital Humanities Research grant. The Open Encyclopedia of Classical Sites. 2010.

Invited Talks

- Topic model selection in metadata-rich corpora. Princeton University, May 2015.

- Data science institute. University of Virginia, April 2015.

- Topic modeling seminar. Swarthmore College, March 2015.

- Micromegas Workshop. Stanford University, February 2015.

- Digital Philology Workshop. University of Leipzig, Germany, July 2014.

- Teaching Machine Learning with d3. OpenVisConf, Boston, MA, April 2014.

- QuanTM Workshop. Emory University, April 2014.

- Topic Modeling and Scholarly Methodologies. Stanford University, Jan 2014.

- Digital Humanities Hackathon. University of Wisconsin, Madison, Nov 2013.

- Digital Praxis Seminar. Graduate Center, CUNY, Sept 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.