GREGORY DRUCK

email:	gdruck@gmail.com
web:	http://www.cs.umass.edu/~gdruck/
cell:	(443) 604-8603
address:	55 Webster St., San Francisco, CA 94117
citizenship:	United States citizen

RESEARCH INTERESTS

Lightly Supervised, Semi-Supervised, and Active Machine Learning
Natural Language Processing
Information Extraction, Mining, and Retrieval

EDUCATION

University of Massachusetts Amherst Amherst, MA
Ph.D., Computer Science, September 2011
Dissertation: Generalized Expectation Criteria for Lightly Supervised Learning
Committee: Andrew McCallum (chair), Xiaojin Zhu, David A. Smith, James Allan, Andrew Cohen

University of Massachusetts Amherst Amherst, MA
M.S., Computer Science, 2008

Johns Hopkins University Baltimore, MD
B.S., Computer Science, 2005
Minor in Applied Mathematics and Statistics

EXPERIENCE

Postdoctoral Scientist September 2011 - Present
Yahoo! Research Santa Clara, CA
Manager: Dr. Evgeniy Gabrilovich

Developing semi-supervised machine learning approaches to natural language processing and information retrieval problems. Ongoing projects include mining user-generated refinements to online instructions, identifying entities and classifying intent in queries, and learning about users from observational data.

Research Assistant July 2005–August 2011
University of Massachusetts Amherst Amherst, MA
Adviser: Dr. Andrew McCallum

Research in semi-supervised and active machine learning for information extraction and natural language processing.

Intern (Research) June 2008–September 2008
Google New York, NY
Adviser: Dr. Gideon Mann

Developed and implemented a method for semi-supervised dependency parsing that uses linguistic prior knowledge as a training signal. This work was published at ACL 2009.

Intern (Live Labs) June 2006–August 2006
Microsoft Redmond, WA

Advisers: Dr. Mukund Narasimhan and Dr. Paul Viola

Developed and implemented a method for fast test time inference in graphical models using A* search with a learned heuristic function. This work was published at AISTATS 2007.

Undergraduate Research Assistant January 2004–May 2005
Center for Language and Speech Processing, Johns Hopkins University Baltimore, MD
Adviser: Dr. David Yarowsky

Developed methods for extracting and synthesizing information from scanned bilingual dictionaries for use in low-density language machine translation.

Undergraduate Research Assistant May 2003–September 2004
Far Ultraviolet Spectroscopic Explorer, Johns Hopkins University Baltimore, MD
Adviser: Dr. David Sahnow

Developed data mining tools to help predict events that interfere with satellite operations.

HONORS

Yahoo Key Scientific Challenges Award (Information Extraction), 2009
Honorable Mention, NSF Graduate Research Fellowship, 2005
Computer Science Department Honors, Johns Hopkins University, 2005

REFEREED CONFERENCE PUBLICATIONS

Gregory Druck and Bo Pang. Spice it up? Mining Refinements to Online Instructions from User Generated Content. To appear in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL 2012).*

Gregory Druck and Andrew McCallum. Toward Interactive Training and Evaluation. In *Proceedings of the ACM Conference on Information and Knowledge Management (CIKM 2011)*, pages 947–956, Glasgow, U.K., October 24-28, 2011 (15% accepted).

Gregory Druck and Andrew McCallum. High-Performance Semi-Supervised Learning using Discriminatively Constrained Generative Models. In *Proceedings of the 27th International Conference on Machine Learning (ICML 2010)*, pages 319–326, Haifa, Israel, June 21-24, 2010 (25.6% accepted).

Gregory Druck, Burr Settles, and Andrew McCallum. Active Learning by Labeling Features. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP 2009)*, pages 81–90, Singapore, August 6-7, 2009 (34% accepted).

Gregory Druck, Gideon Mann, and Andrew McCallum. Semi-supervised Learning of Dependency Parsers using Generalized Expectation Criteria. In *Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL-IJCNLP 2009)*, pages 360–368, Singapore, August 2-7, 2009 (21% accepted).

Kedar Bellare, Gregory Druck, and Andrew McCallum. Alternating Projections for Learning with Expectation Constraints. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI 2009)*, pages 35–42, Montreal, Canada (31% accepted).

Gregory Druck, Gideon Mann, and Andrew McCallum. Learning from Labeled Features using Generalized Expectation Criteria. In *Proceedings of the 31st Annual International ACM SIGIR*

Conference, pages 595-602, Singapore, July 2008 (17% accepted).

Gregory Druck, Chris Pal, Xiaojin Zhu, and Andrew McCallum. Semi-Supervised Classification with Hybrid Generative/Discriminative Methods. In *Proceedings of the Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 280–289, San Jose, CA, August 2007. (18% accepted).

Gregory Druck, Mukund Narasimhan, and Paul Viola. Learning A* Underestimates: Using Inference to Guide Inference. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics (AISTATS)*, San Juan, Puerto Rico, March 21-24, 2007.

Andrew McCallum, Chris Pal, Gregory Druck, and Xuerui Wang. Multi-Conditional Learning: Generative/Discriminative Training for Clustering and Classification. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI)*, pages 433-439, Boston, MA, June 2006. (30% accepted).

REFEREED WORKSHOP PUBLICATIONS

Gregory Druck, Gerome Miklau, and Andrew McCallum. Learning to Predict the Quality of Contributions to Wikipedia. In *Proceedings of the AAAI Workshop on Wikipedia and Artificial Intelligence*, pages 7-12, Chicago, IL, July 2008. (37% accepted).

Gregory Druck, Gideon Mann, and Andrew McCallum. Leveraging Existing Resources using Generalized Expectation Criteria. In *Advances in Neural Information Processing Systems (NIPS) Workshop on Learning Problem Design*, Vancouver, Canada, December 2007.

Chris Pal, Michael Kelm, Xuerui Wang, Gregory Druck and Andrew McCallum. On Discriminative and Semi-Supervised Dimensionality Reduction. In *Advances in Neural Information Processing Systems (NIPS) Workshop on Novel Applications of Dimensionality Reduction*, Vancouver, Canada, December 2006.

TECHNICAL REPORTS

Gregory Druck and David Smith. Computing Conditional Feature Covariance in Non-Projective Tree Conditional Random Fields. Technical Report UM-CS-2009-060, University of Massachusetts Amherst, December 2009.

PROFESSIONAL SERVICE

Tutorial Instructor: Rich Prior Knowledge in Learning for Natural Language Processing (with Kuzman Ganchev and João Graça). *The 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011)*. One of the most popular tutorials at the conference, with 95 attendees.

Tutorial Instructor: Rich Prior Knowledge in Learning (with Kuzman Ganchev and João Graça). *12th Annual Conference of the International Speech Communication Association (Interspeech 2011)*. One of the most popular tutorials at the conference, with ≈ 30 attendees.

Tutorial Instructor: Rich Prior Knowledge in Learning for Reducing Annotation Cost (with Kuzman Ganchev and João Graça). To be given at *The 8th International Conference on Language Resources and Evaluation (LREC 2012)*.

Program Committee: SIGIR 2012, ICML 2012, EACL 2012, ROBUS-UNSUP 2012, NIPS Big Learning Workshop 2011, EMNLP 2011, ACL 2011, IJCAI 2011, ACL 2010, COLING 2010, AAAI 2010: AI and the Web Track, and the 2009 NAACL HLT Student Research Workshop.

Reviewer: EMNLP 2012, AISTATS 2012, EMNLP 2011, ACM Transactions on Intelligent Systems and Technology, Journal of Machine Learning Research, IEEE Transactions on Pattern Analysis and Machine Intelligence, The Netherlands Organisation for Scientific Research, KDD 09, ICML 08, ACL 08, NIPS 07, UAI 07, ICML 07, ACL 07.

Guest Lecturer: *Introduction to Natural Language Processing* and *Graphical Models* (both taught by Prof. Andrew McCallum)
University of Massachusetts Amherst, Fall 2007 & Spring 2011

Teaching Assistant: *Computational Linguistics* (taught by Prof. Andrew McCallum)
University of Massachusetts Amherst, Spring 2006

SOFTWARE ENGINEERING

Software: Key contributor to the development, maintenance, and support of the *MAchine Learning for LanguagE Toolkit (MALLET)*: `http://mallet.cs.umass.edu/`.

Primary Programming Languages: Java, Perl

Secondary Programming Languages: C++, Scala, Matlab

Parallel Computing: Apache Hadoop, Apache Pig, Java multi-threading