

Brendan O'Connor

PhD Student, Machine Learning Department
School of Computer Science, Carnegie Mellon University
5000 Forbes Ave., Pittsburgh, PA 15213
brenocon@cmu.edu <http://brenocon.com>

Education

Carnegie Mellon University, Pittsburgh, PA 2010 – present
PhD student, Machine Learning Department 2009 – 2010
Previously enrolled: Masters in Language Technologies 2001 – 2006
Stanford University, Stanford, CA: co-terminal master student 2001 – 2006
Master of Science in Symbolic Systems
Bachelor of Science with honors in Symbolic Systems
Edina High School, Edina, MN 1998 – 2001
University of Minnesota Talented Youth Mathematics Program, Minneapolis, MN 1997 – 2000

Work Experience

Carnegie Mellon University (Graduate Research Assistant) Pittsburgh, PA Sept 2009 – present
Research assistant in statistical NLP group (Prof. Noah Smith): discriminative and generative models for text analysis in social media and other domains, with applications to economics, linguistics, and political science. Press coverage in NPR, Wall Street Journal, BBC, Washington Post, etc. See publication list.

Facebook, Inc. (Research Intern) Pittsburgh, PA Summer 2010
Statistical modeling of interest preferences and homophily in the Data Science group.

Crowdfunder / Dolores Labs, Inc. (Scientific Consultant) San Francisco, CA 2008 – 2009
Designed quality control system to use online workers from Amazon's Mechanical Turk service (an online labor marketplace) to solve customer annotation problems in information extraction, information retrieval, market research, and content moderation. Co-authored highly cited paper on applications to NLP. Wrote blog posts in a popular science style, attracting more than 200,000 views.

Powerset, Inc. (Engineer) San Francisco, CA 2006 – 2008
Designed and implemented many systems for a semantic search engine, especially relevance ranking and natural language engineering. Thorough end-to-end knowledge of a highly complex question answering system. Early user of Hadoop and EC2. Built practical data visualization and development tools that were still in use more than a year after I left.

Free University (Research Intern) Berlin, Germany 2004
Worked on e-Chalk educational instruction software, incorporating handwriting recognition.

Stanford University (Research Assistant) Stanford, CA 2003 – 2006
Designed and developed an open source web application for online deliberation (discussion, document collaboration, voting).

Stanford University (Advising Fellow) Stanford, CA 2003 – 2005
Undergraduate student advisor for Stanford's interdisciplinary Symbolic Systems Program. Academic peer advising, organizing and inviting speakers, running student and campus events, giving input to curriculum decisions.

Stanford University (Teaching) Stanford, CA 2003 and 2005
1) Section leader (i.e., TA) for introductory computer science course.
2) Developed and taught student-initiated reading course on *Gödel, Escher, Bach*, on the philosophy of artificial intelligence. Invited faculty speakers, led discussions.

Paul Wellstone for Senate (Volunteer) Minneapolis, MN July 2002
Assisted in information management for U.S. Senate campaign.

Refereed Conference Papers (Full Length)

“Predicting a Scientific Community's Response to an Article.”

Dani Yogatama, Michael Heilman, Brendan O'Connor, Chris Dyer, Bryan R. Routledge, and Noah A. Smith.
EMNLP-2011: Proceedings of the 2011 Conference on Empirical Methods on Natural Language Processing.

“A Latent Variable Model for Geographic Lexical Variation.”

Jacob Eisenstein, Brendan O'Connor, Noah A. Smith, Eric P. Xing.
EMNLP-2010: Proceedings of the 2010 Conference on Empirical Methods on Natural Language Processing.

“From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series.”

Brendan O'Connor, Ramnath Balasubramanyan, Bryan Routledge, Noah A. Smith.
ICWSM-2010: Proceedings of 4th International AAAI Conference on Weblogs and Social Media.

“Cheap and Fast — But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks.”

Rion Snow, Brendan O'Connor, Daniel Jurafsky, Andrew Y. Ng.
EMNLP-2008: Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing.

Refereed Conference Papers (Short Length)

“Part-of-Speech Tagging for Twitter: Annotations, Features, Experiments.”

Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeff Flanigan, Noah A. Smith, 2011.
ACL-2011: Proceedings of the Annual Meeting of the Association for Computational Linguistics.

“Discovering Demographic Language Variation.”

Brendan O'Connor, Jacob Eisenstein, Eric P. Xing, Noah A. Smith.
NIPS-2010 [workshop]: Forthcoming, *Journal of Machine Learning Research*, Workshop and Conference Proceedings, for the Workshop on Machine Learning for Social Computing, forthcoming 2011

“TweetMotif: Exploratory Search and Topic Summarization for Twitter.”

Brendan O'Connor, Michel Krieger, David Ahn.
ICWSM-2010 [demo]: Proceedings of the 4th International AAAI Conference on Weblogs and Social Media.

“Groupware for Groups: Problem-Driven Design in Deme.”

Todd Davies, Brendan O'Connor, Alex Cochran, Andrew Parker.
CHI-2005 [workshop]: Position paper from the Beyond Threaded Conversation Workshop at Computer-Human Interaction.

Book Chapters

“Superficial Data Analysis: Exploring Millions of Social Stereotypes.”

Brendan O'Connor and Lukas Biewald. In *Beautiful Data: The Stories Behind Elegant Data Solutions*, ed. Toby Segaran and Jeff Hammerbacher. O'Reilly Media. 2009.

“An Online Environment for Democratic Deliberation: Motivations, Principles, and Design.”

Todd Davies, Brendan O'Connor, Alex Angiolillo Cochran, Andrew Parker, Jonathan Effrat, Benjamin Newman, Aaron Tam. In Todd Davies and Seeta Pena Gangadharan (editors), *Online Deliberation: Design, Research, and Practice*, CSLI Publications/University of Chicago Press, 2009.

Book Review

Review of Donald Knuth, “Things a Computer Scientist Rarely Talks About.”

Brendan O'Connor. In Raul Rojas (editor), “Reviews.” *IEEE Annals of the History of Computing*, Oct-Dec 2004.

Invited Talks

“Dolores Labs and Amazon Mechanical Turk.”

- Invited talk, Stanford Human-Computer Interaction group (hci.stanford.edu). Sept 2008.
- Invited talk, the Stanford Natural Language Processing group (nlp.stanford.edu). May 2008.
- Invited talk, Freebase / Metaweb Technologies, Inc., San Francisco, CA. (freebase.com). May 2008.
- Invited talk, Dataviz Salon SF, San Francisco, CA. February 2009.

“Modeling Social Phenomena in Microblog Text.”

- Invited talk, Social Computing research group, CMU HCII
- Invited talk, Machine Learning seminar, CMU MLD

Tutorial

“An Interactive Introduction to R.”

Michael Driscoll and Brendan O'Connor.

Tutorial by Dataspora LLC, for Amyris, Inc., Emeryville, CA. June 2009. (dataspora.com, amyrisbiotech.com)

Academic Service

Reviewer for

- IJCNLP. 2011.
- EMNLP. 2011.
- ICWSM: International Conference on Weblogs and Social Media. 2011.
- Communications of the ACM. 2010.
- CHI: Computer-Human Interaction. 2009.

Program Committee for Workshop on Creating Speech and Language Data With Amazon's Mechanical Turk. The 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics / Human Language Technologies, June 2010, Los Angeles.

Awards

Finalist, Facebook Fellowship Program. 2010 and 2011. (<http://www.facebook.com/careers/fellowship.php>)

K. Jon Barwise Award for Distinguished Contribution to the Symbolic Systems Program. 2005.

National AP Scholar (2001), AP State Scholar (2001), National Merit Scholar Finalist (2001), Qualifier for American Invitational Mathematics Examination (2000, 2002).

Theses

“The social evolution of human cooperation via group competition and conflict.”

B.S. honors thesis, 2006. Readers: Noah Mark and James Fearon.

"Biased evidence assimilation under bounded Bayesian rationality."

M.S. thesis, 2006. Readers: Jonathan Bendor and James McClelland.

Other Projects

Netflix Prize competition. 2008.

Spent one month researching and implementing baseline recommendation system.

Final rank (2009): 645 out of 41305 teams.

KDD Cup competition. 2009.
Task: Prediction system for marketing analysis.
Final rank: 287 of 4921 (valid) entries.

Fleshmap.com. 2008.
With Fernanda Viegas and Martin Wattenberg.
Mechanical Turk data collection supporting the spatial visualization of human sexual preferences.

PalinSpeak.com. 2008.
Brendan O'Connor and Douglas Wilson.
Markov model chatbot as political satire.

Skills

Programming languages: R, Matlab, Prolog, C/C++, Java, Scala, Python, Ruby, Awk/Sed/Perl, PHP, Javascript, Flash/Flex.

Other technologies: MySQL, PostgreSQL, EC2, S3, Hadoop, Hive, Linux, LaTeX, Excel, Processing.

Patents Pending

“Emphasizing Search Results According to Conceptual Meaning.”
Barney Pell, Scott Prevost, Giovanni Lorenzo Thione, Brendan O'Connor, Lukas Biewald.
Filed August 29, 2008. Application No. 12/201,504. Powerset, Inc. and Microsoft Corporation.

“Efficient Storage and Retrieval of Posting Lists.”
Chad Walters, Giovanni Lorenzo Thione, Barney Pell, Lukas Biewald, Brendan O'Connor.
Filed August 29, 2008. Application No. 12/201,079. Powerset, Inc. and Microsoft Corporation.

“Browsing Knowledge on the Basis of Semantic Relations.”
Franco Salvetti, Giovanni Lorenzo Thione, Richard Crouch, Lukas Biewald, Brendan O'Connor, Barney Pell.
Filed August 29, 2008. Application No. 12/201,978. Powerset, Inc. and Microsoft Corporation.