

Michael J. Paul

mpaul@colorado.edu
http://michaelpaul.com

Last Updated: August 15, 2018

Research interests: natural language processing, machine learning, text analysis, topic modeling, health and medical informatics, computational epidemiology, social media analysis

EMPLOYMENT

2015 – Present **University of Colorado Boulder**
Assistant Professor, Information Science

EDUCATION

2010 – 2015 **Johns Hopkins University**
PhD, Computer Science, August 2015
Committee: Mark Dredze (advisor), Jason Eisner (co-advisor),
Eric Horvitz, Hanna Wallach
MSE, Computer Science, May 2012

2005 – 2009 **University of Illinois at Urbana-Champaign**
BS, Computer Science, May 2009

AWARDS AND HONORS

2013 – 2015	Microsoft Research Fellowship	Microsoft Corporation
2010 – 2012	Graduate Research Fellowship	National Science Foundation
2010 – 2015	Dean's Fellowship	Johns Hopkins University
2005 – 2008	Edmund J. James Scholar	University of Illinois

PEER-REVIEWED BOOKS

2017 Michael J. Paul, Mark Dredze (2017) **Social monitoring for public health. *Synthesis Lectures on Information Concepts, Retrieval, and Services* 9(5): 1-185. ISBN: 9781681730950.**
http://www.morganclaypoolpublishers.com/catalog_Orig/product_info.php?products_id=1102

PEER-REVIEWED JOURNAL PUBLICATIONS (* = student)

2018 Meredith C. Meacham, Michael J. Paul, Danielle E. Ramo (2018) **Understanding emerging forms of cannabis use through an online community: an analysis of relative post volume and subjective highness ratings. *Drug and Alcohol Dependence* 188: 364-369.**
<https://www.sciencedirect.com/science/article/pii/S0376871618302424>

Sachin Muralidhara*, Michael J. Paul (2018) **#Healthy selfies: exploration of health topics on Instagram. *JMIR Public Health and Surveillance* 4(2): e10150 (~11 pages).** <http://publichealth.jmir.org/2018/2/e10150/>

Alexis S. Hammond, Michael J. Paul, Joseph Hobelmann, Animesh R.

Koratana*, Mark Dredze, Margaret S. Chisolm (2018) **Perceived attitudes about substance use in anonymous social media posts near college campuses: observational study.** *JMIR Mental Health* 5(3): e52 (~6 pages). <http://mental.jmir.org/2018/3/e52/>

2016

Michael J. Paul, Margaret S. Chisolm, Matthew W. Johnson, Ryan G. Vandrey, Mark Dredze (2016) **Assessing the validity of online drug forums as a source for estimating demographic and temporal trends in drug use.** *Journal of Addiction Medicine* 10(5): 324-330. <https://www.ncbi.nlm.nih.gov/pubmed/27466069>

Ye Zhang*, Erin Willis, Michael J. Paul, Noemie Elhadad, Byron C. Wallace (2016) **Characterizing the (perceived) newsworthiness of health science articles: a data-driven approach.** *JMIR Medical Informatics* 4(3): e27 (~19 pages). <https://medinform.jmir.org/2016/3/e27/>

Michael J. Paul, Ryen W. White, Eric Horvitz (2016) **Search and breast cancer: on episodic shifts of attention over life histories of an illness.** *ACM Transactions on the Web (TWEB)* 10(2): 1-27. <https://dl.acm.org/citation.cfm?id=2932204.2893481>

Atul Nakhasi, Ralph J. Passarella, Sarah G. Bell, Michael J. Paul, Mark Dredze, Peter J. Pronovost (2016) **The potential of Twitter as a data source for patient safety.** *Journal of Patient Safety.* (~14 pages) <https://www.ncbi.nlm.nih.gov/pubmed/26756726>

2015

Mauricio Santillana, Andre T. Nguyen, Mark Dredze, Michael J. Paul, Elaine O. Nsoesie, John S. Brownstein (2015) **Combining search, social media, and traditional data sources to improve influenza surveillance.** *PLOS Computational Biology* 11(10): e1004513 (~15 pages). <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004513>

Benjamin M. Althouse, Samuel V. Scarpino, Lauren Ancel Meyers, John W. Ayers, Marisa Bargsten, Joan Baumbach, John S. Brownstein, Lauren Castro, Hannah Clapham, Derek A.T. Cummings, Sara Del Valle, Stephen Eubank, Geoffrey Fairchild, Lyn Finelli, Nicholas Generous, Dylan George, David R. Harper, Laurent Hebert-Dufresne, Michael A. Johansson, Kevin Konty, Marc Lipsitch, Gabriel Milinovich, Joseph D. Miller, Elaine O. Nsoesie, Donald R. Olson, Michael J. Paul, Philip M. Polgreen, Reid Priedhorsky, Jonathan M. Read, Isabel Rodriguez-Barraquer, Derek J. Smith, Christian Stefansen, David L. Swerdlow, Deborah Thompson, Alessandro Vespignani, Amy Wesolowski (2015) **Enhancing disease surveillance with novel data streams: challenges and opportunities.** *EPJ Data Science* 4:47 (~8 pages). <https://www.ncbi.nlm.nih.gov/pubmed/27990325>

David A. Broniatowski, Mark Dredze, Michael J. Paul*, Andrea Dugas (2015) **Using social media to perform local influenza surveillance in an inner-city hospital: a retrospective observational study.** *JMIR Public Health and Surveillance* 1(1): e5 (~7 pages). <https://www.ncbi.nlm.nih.gov/pubmed/27014744>

Michael J. Paul*, Mark Dredze (2015) **SPRITE: Generalizing topic models with structured priors**. *Transactions of the Association for Computational Linguistics (TACL)* 3: 43-57.

<https://transacl.org/ojs/index.php/tacl/article/view/403>

Shiliang Wang*, Michael J. Paul*, Mark Dredze (2015) **Social media as a sensor of air quality and public response in China**. *Journal of Medical Internet Research (JMIR)* 17(3): e22 (~10 pages).

<https://www.ncbi.nlm.nih.gov/pubmed/25831020>

2014

Michael J. Paul*, Mark Dredze, David Broniatowski (2014) **Twitter improves influenza forecasting**. *PLOS Currents Outbreaks*. October 2014 (~13 pages). <http://currents.plos.org/outbreaks/article/twitter-improves-influenza-forecasting/>

Michael J. Paul*, Mark Dredze (2014) **Discovering health topics in social media using topic models**. *PLOS ONE* 9(8): e103408 (~11 pages).

<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0103408>

Ahmed Abbasi, Donald Adjeroh, Mark Dredze, Michael J. Paul*, Fatemeh Mariam Zahedi, Huimin Zhao, Nitin Walia, Hemant Jain, Patrick Sanvanson, Reza Shaker, Marco D. Huesch, Richard Beal, Wanhong Zheng, Marie Abate, Arun Ross (2014) **Social media analytics for smart health**. *IEEE Intelligent Systems* 29(2): 60-80. (My contribution was a ~4 page article within this collection of essays.)

<https://www.computer.org/csdl/mags/ex/2014/02/mex2014020060-abs.html>

Byron C. Wallace, Michael J. Paul*, Urmimala Sarkar, Thomas A. Trikalinos, Mark Dredze (2014) **A large-scale quantitative analysis of latent factors and sentiment in online doctor reviews**. *Journal of the American Medical Informatics Association (JAMIA)* 21(6), 1098-1103.

<https://www.ncbi.nlm.nih.gov/pubmed/24918109>

2013

David A. Broniatowski, Michael J. Paul*, Mark Dredze (2013) **National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic**. *PLOS ONE* 8(12): e83672 (~7 pages).

<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0083672>

2011

Roxana Girju, Michael J. Paul* (2011) **Modeling reciprocity in social interactions with probabilistic latent space models**. *Natural Language Engineering* 17(1): 1-36. <https://dl.acm.org/citation.cfm?id=1972848>

PEER-REVIEWED CONFERENCE PUBLICATIONS (* = student)

2018

Linzi Xing*, Michael J. Paul. **Diagnosing and improving topic models by analyzing posterior variability**. *AAAI Conference on Artificial Intelligence (AAAI)*, New Orleans, Louisiana. February 2018. (8 pages)

http://cmci.colorado.edu/~mpaul/files/aaai18_stability.pdf

- Shudong Hao*, Michael J. Paul. **Learning multilingual topics from incomparable corpora.** *International Conference on Computational Linguistics (COLING)*, Santa Fe, New Mexico. August 2018. (14 pages)
http://cmci.colorado.edu/~mpaul/files/coling2018_softlinks.pdf
- Xiaolei Huang*, Michael J. Paul. **Examining temporality in document classification.** *Association for Computational Linguistics (ACL)*, Melbourne, Australia, July 2018. Short paper, pages 694–699.
<http://aclweb.org/anthology/P18-2110>
- Shudong Hao*, Jordan Boyd-Graber, Michael J. Paul. **Lessons from the Bible on modern topics: low-resource multilingual topic model evaluation.** *North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, New Orleans, Louisiana, June 2018. Pages 1090–1100. <http://aclweb.org/anthology/N18-1099>
- 2017 Michael J. Paul. **Feature selection as causal inference: experiments with text classification.** *Conference on Computational Natural Language Learning (CoNLL)*, Vancouver, Canada, August 2017. Pages 163-172.
<http://www.aclweb.org/anthology/K17-1018>
- Xiaolei Huang*, Linzi Xing*, Jed R. Brubaker, Michael J. Paul. **Exploring timelines of confirmed suicide incidents through social media.** *IEEE International Conference on Healthcare Informatics (ICHI)*, Park City, Utah, August 2017. Pages 470-477.
<https://ieeexplore.ieee.org/document/8031196/>
- 2016 Adrian Benton*, Michael J. Paul, Braden Hancock*, Mark Dredze. **Collective supervision of topic models for predicting surveys with social media.** *AAAI Conference on Artificial Intelligence (AAAI)*, Phoenix, Arizona, February 2016. (7 pages)
<https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/11978>
- Michael J. Paul, Abeer Sarker, John S. Brownstein, Azadeh Nikfarjam, Matthew Scotch, Karen L. Smith, Graciela Gonzalez. **Social media mining for public health monitoring and surveillance.** *Pacific Symposium on Biocomputing (PSB)*, Big Island of Hawaii, January 2016. Pages 468-479.
<https://asu.pure.elsevier.com/en/publications/social-media-mining-for-public-health-monitoring-and-surveillance>
- 2015 Michael J. Paul*, Ryen W. White, Eric Horvitz. **Diagnoses, decisions, and outcomes: Web search as medical decision support for cancer.** *24th International World Wide Web Conference (WWW)*, Florence, Italy, May 2015. (11 pages)
http://cmci.colorado.edu/~mpaul/files/www2015_treatments.pdf
- 2013 Alex Lamb*, Michael J. Paul*, Mark Dredze. **Separating fact from fear: Tracking flu infections on Twitter.** *2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language*

Technologies (NAACL-HLT), Atlanta, June 2013. Short paper, pages 789–795. <http://www.aclweb.org/anthology/N13-1097>

Michael J. Paul*, Mark Dredze. **Drug extraction from the Web: Summarizing drug experiences with multi-dimensional topic models.** *2013 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)*, Atlanta, June 2013. Pages 168–178. <http://aclweb.org/anthology/N13-1017>

2012 Michael J. Paul*, Mark Dredze. **Factorial LDA: Sparse multi-dimensional models of text.** In *Advances in Neural Information Processing Systems (NIPS)*. Lake Tahoe, Nevada, December 2012. (9 pages) <https://papers.nips.cc/paper/4784-factorial-lda-sparse-multi-dimensional-text-models>

Michael J. Paul*. **Mixed membership Markov models for unsupervised conversation modeling.** In the *2012 Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, Jeju, Korea, July 2012. Pages 94–104. <https://dl.acm.org/citation.cfm?id=2390960>

Michael J. Paul*, Jason Eisner. **Implicitly intersecting weighted automata using dual decomposition.** In the *2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, Montreal, Canada, June 2012. Pages 232–242. <https://dl.acm.org/citation.cfm?id=2382060>

2011 Michael J. Paul*, Mark Dredze. **You are what you Tweet: Analyzing Twitter for public health.** In the *5th International AAAI Conference on Weblogs and Social Media (ICWSM)*, July 2011. (8 pages) <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.224.9974>

Delip Rao*, Michael Paul*, Clayton Fink, David Yarowsky, Timothy Oates, Glen Coppersmith. **Hierarchical Bayesian models for latent attribute detection in social networks.** In the *5th International AAAI Conference on Weblogs and Social Media (ICWSM)*, July 2011. (Short paper, 4 pages) <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2881>

2010 Michael J. Paul*, ChengXiang Zhai, Roxana Girju. **Summarizing contrastive viewpoints in opinionated text.** In the proceedings of the *2010 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Cambridge, Massachusetts, October 2010. Pages 66–76. <https://dl.acm.org/citation.cfm?id=1870665>

Michael Paul*, Roxana Girju. **A two-dimensional topic-aspect model for discovering multi-faceted topics.** In the proceedings of the *Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI)*, Atlanta, Georgia, July 2010. Pages 545–550. <https://dl.acm.org/citation.cfm?id=2898695>

2009 Michael Paul*, Roxana Girju. **Topic modeling of research fields: an interdisciplinary perspective.** In the proceedings of *Recent Advances in*

Natural Language Processing (RANLP), September 2009. (6 pages)
<http://cmci.colorado.edu/~mpaul/files/ranlp09.pdf>

Michael Paul*, Roxana Girju. **Cross-cultural analysis of blogs and forums with mixed-collection topic models.** *Empirical Methods in Natural Language Processing (EMNLP)*, Singapore. August 2009. (10 pages) <http://www.aclweb.org/anthology/D09-1146>

Michael Paul*, Roxana Girju, Chen Li. **Mining the web for reciprocal relationships.** In the proceedings of the *13th Conference on Computational Natural Language Learning (CoNLL)*, Boulder, Colorado, June 2009. Pages 75-83. <http://www.aclweb.org/anthology/W09-1111>

PEER-REVIEWED WORKSHOP PUBLICATIONS (* = student)

- 2017
- Ashlynn R. Daughton*, Dasha Pruss*, Brad Arnot*, Danielle Albers Szafir, Michael J. Paul. **Characteristics of Zika behavior discourse on Twitter.** *AMIA Workshop on Social Media Mining for Health Applications*, Washington, DC, November 2017. (5 pages)
http://cmci.colorado.edu/~mpaul/files/amia17_zika.pdf
- Linzi Xing*, Michael J. Paul. **Incorporating metadata into content-based user embeddings.** *EMNLP Workshop on Noisy User-generated Text (W-NUT)*, Copenhagen, Denmark, September 2017. (5 pages)
http://cmci.colorado.edu/~mpaul/files/wnut17_embedding.pdf
- Brian C. Keegan, Jofish Kaye, Patricia Cavazos-Rehg, Munmun de Choudhury, Anh Ngoc Nguyen, Michael J. Paul, Saiph Savage. **CHI-nnabis: Implications of marijuana legalization for and from human-computer interaction.** *CHI Conference Extended Abstracts on Human Factors in Computing Systems*, Denver, CO, May 2017. Panel presentation.
<http://cmci.colorado.edu/~mpaul/files/chinnabis17.pdf>
- Xiaolei Huang*, Michael C. Smith*, Michael J. Paul, Dmytro Ryzhkov*, Sandra C. Quinn, David A. Broniatowski, Mark Dredze. **Examining patterns of influenza vaccination in social media.** *AAAI Joint Workshop on Health Intelligence (W3PHIAI)*, San Francisco, February 2017. (5 pages)
http://cmci.colorado.edu/~mpaul/files/w3phi17_vaccines.pdf
- 2016
- Kevin Stowe*, Michael Paul, Martha Palmer, Leysia Palen, Ken Anderson. **Identifying and categorizing disaster-related tweets.** *EMNLP Workshop on Natural Language Processing for Social Media (SocialNLP)*, Austin, Texas, November 2016. (6 pages)
http://cmci.colorado.edu/~mpaul/files/socialnlp16_disasters.pdf
- Michael J. Paul. **Interpretable machine learning: lessons from topic modeling.** *CHI Workshop on Human-Centered Machine Learning*. San Jose, California, May 2016. (6 pages)
http://cmci.colorado.edu/~mpaul/files/chi16hcmli_interpretable.pdf
- Michael C. Smith*, David A. Broniatowski, Michael J. Paul, Mark Dredze. **Towards real-time measurement of public epidemic awareness:**

monitoring influenza awareness through Twitter. *AAAI Spring Symposium on Observational Studies through Social Media and Other Human-Generated Content*, Stanford, California, March 2016. (4 pages)
http://cmci.colorado.edu/~mpaul/files/aaai16ossm_awareness.pdf

Animesh Koratana*, Mark Dredze, Margaret S. Chisolm, Matthew W. Johnson, Michael J. Paul. **Studying anonymous health issues and substance use on college campuses with Yik Yak.** *AAAI Workshop on the World Wide Web and Public Health Intelligence*, Phoenix, Arizona, February 2016. (4 pages)
http://cmci.colorado.edu/~mpaul/files/w3phi16_yikyak.pdf

2015 Michael J. Paul*, Mark Dredze, David A. Broniatowski, Nicholas Generous. **Worldwide influenza surveillance through Twitter.** *AAAI Workshop on the World Wide Web and Public Health Intelligence*, Austin, Texas, January 2015. (7 pages)
http://cmci.colorado.edu/~mpaul/files/w3phi2015_worldwide.pdf

Byron Wallace, Michael J. Paul*, Noémie Elhadad. **What predicts media coverage of health science articles?** *AAAI Workshop on the World Wide Web and Public Health Intelligence*, Austin, Texas, January 2015. (4 pages)
http://cmci.colorado.edu/~mpaul/files/w3phi2015_media.pdf

2014 Shiliang Wang*, Michael J. Paul*, Mark Dredze. **Exploring health topics in Chinese social media: an analysis of Sina Weibo.** *AAAI Workshop on the World Wide Web and Public Health Intelligence*, Quebec City, July 2014. (4 pages)
http://cmci.colorado.edu/~mpaul/files/w3phi2014_weibo.pdf

Mark Dredze, Renyuan Cheng*, Michael Paul*, David Broniatowski. **HealthTweets.org: a platform for public health surveillance using Twitter.** *AAAI Workshop on the World Wide Web and Public Health Intelligence*, Quebec City, July 2014. (2 pages)
http://cmci.colorado.edu/~mpaul/files/w3phi2014_healthtweets.pdf

2013 Michael J. Paul*, Byron C. Wallace, Mark Dredze. **What affects patient (dis)satisfaction? Analyzing online doctor ratings with a joint topic-sentiment model.** *AAAI Workshop on Expanding the Boundaries of Health Informatics Using AI (HIAI)*, Bellevue, WA, July 2013. (6 pages)
http://cmci.colorado.edu/~mpaul/files/aaai13_flda_sentiment.pdf

Mark Dredze, Michael J. Paul*, Shane Bergsma, Hieu Tran*. **Carmen: a Twitter geolocation system with applications to public health.** *AAAI Workshop on Expanding the Boundaries of Health Informatics Using AI (HIAI)*, Bellevue, WA, July 2013. (5 pages)
http://cmci.colorado.edu/~mpaul/files/aaai13_geo.pdf

2012 Michael J. Paul*, Mark Dredze. **Experimenting with drugs (and topic models): Multi-dimensional exploration of recreational drug discussions.** *AAAI 2012 Fall Symposium on Information Retrieval and Knowledge Discovery in Biomedical Text*, Arlington, VA, November 2012. (6 pages)
<http://cmci.colorado.edu/~mpaul/files/aaaibio12-drugs.pdf>

Alex Lamb*, Michael J. Paul*, Mark Dredze. **Investigating Twitter as a source for studying behavioral responses to epidemics.** *AAAI 2012 Fall Symposium on Information Retrieval and Knowledge Discovery in Biomedical Text*, Arlington, VA, November 2012. (2 pages)
http://cmci.colorado.edu/~mpaul/files/aaaibio12-flu_concern.pdf

Atul Nakhasi*, Ralph J. Passarella*, Sarah G. Bell*, Michael J. Paul*, Mark Dredze, Peter J. Pronovost. **Malpractice and malcontent: Analyzing medical complaints in Twitter.** *AAAI 2012 Fall Symposium on Information Retrieval and Knowledge Discovery in Biomedical Text*, Arlington, VA, November 2012. (2 pages) http://cmci.colorado.edu/~mpaul/files/aaaibio12-patient_safety.pdf

William M. Darling*, Michael J. Paul*, Fei Song. **Unsupervised part-of-speech tagging in noisy and esoteric domains with a syntactic-semantic Bayesian HMM.** In the *EACL 2012 Workshop on Semantic Analysis in Social Media*, Avignon, France, April 2012. (9 pages)
<http://cmci.colorado.edu/~mpaul/files/poslda12.pdf>

2008 Michael Paul*, Roxana Girju. **AIRTA: an automatic interdisciplinary research topic advisor.** *NSF-sponsored Symposium on Semantic Knowledge Discovery, Organization and Use - Demo session*, New York University, November 2008. (2 pages)
<http://cmci.colorado.edu/~mpaul/files/nyu08.pdf>

NON-REFEREED PUBLICATIONS

2015 Michael J. Paul. **Topic modeling with structured priors for text-driven science.** *PhD thesis*, Johns Hopkins University. (277 pages, double spaced)
http://cmci.colorado.edu/~mpaul/files/phd_thesis.pdf

2014 David A. Broniatowski, Michael J. Paul, Mark Dredze. **Twitter: Big data opportunities (letter).** *Science* 345(6193): 148. 2014.
http://cmci.colorado.edu/~mpaul/files/science_letter_twitter.pdf

GRANTS

Current \$1,081,143 (investigator portion: ~\$270,285). **CHS: Medium: Scaling Qualitative Inductive Analysis through Computational Methods.** *National Science Foundation*, IIS-1764089, 9/1/2018 – 8/31/2021. PI: Danielle Szafir. Co-PIs: Casey Fiesler, Michael Paul, Jed Brubaker.

\$174,117 (investigator portion: \$174,117). **CRII: III: Real-World Machine Learning: Adaptation Methods for Addressing Temporal, Geographic, and Demographic Confounds in User-Generated Content.** *National Science Foundation*, IIS-1657338, 9/1/2017 – 8/31/2019. PI: Michael Paul.

Pending \$476,537. **CAREER: Causality, Certainty, and Coherence: Increasing the Robustness of Learning and Evaluation of Text Models.** *National Science Foundation*. PI: Michael Paul.

\$1,114,886. **SCH: INT: Understanding and Modeling Processes to Co-Create Chronic Disease Self-Management in Online Communities.**

PI: Michael Paul. Co-PIs: Erin Willis, Lisa Davis, Sheana Bull.

\$118,738 (CU portion). **Misuse of Opioids in Online Networks.** *National Institutes of Health.* PI: Urmimala Sarkar (UC San Francisco). Co-Is:

Danielle Ramo, Gem Le, Byron Wallace, Michael Paul (site PI).

Completed \$49,038. **Workshop on Ontology Learning for Behavioral Knowledge Embeddedness.** *National Science Foundation,* IIS-1612580, 7/1/2016 – 6/30/2018. PI: Kai Larsen. Co-PI: Michael Paul.

TEACHING

Instructor **INFO 1301: Quantitative Reasoning 1** University of Colorado

- Fall 2016, Spring 2017

INFO 2301: Quantitative Reasoning 2

- Spring 2017, Spring 2018

INFO 4604/5604: Applied Machine Learning

- Fall 2017

Guest Lecture **CMCI 1020: Concepts and Creativity 2** University of Colorado

- Spring 2016, two lectures

CSCI 4802/5802: Colorado Data Science Team

- 10/04/2016, talk in “Meet a data scientist” series

CBHS 6628: Technology-Based Health Promotion

- Recorded video lecture on machine learning for online course

CS 465: Natural Language Processing Johns Hopkins University

- 12/04/2013, “An introduction to topic models”

NETS 213: Crowdsourcing & Human Comp. University of Pennsylvania

- 10/30/2013, “Mining health information from Twitter”

Teaching Assistant **CS 476/676: Machine Learning in Complex Domains** Johns Hopkins University

- Spring 2013

CS 475: Machine Learning

- Fall 2012

INVITED TALKS

1/24/2018 “Social monitoring of substance abuse.” University of Colorado Boulder, Information Science Seminar. Boulder, CO.

11/04/2017 “Social monitoring of substance abuse.” Social Media Mining for Health Applications Workshop at AMIA 2017. Washington, DC.

2/04/2016 “Discovering patterns in text collections with structured topic models.” University of Colorado Boulder, Library Research Symposium. Boulder, CO.

10/12/2015 “Learning about population health by mining web activity.” University of Colorado Denver, School of Medicine, Computational Bioscience Seminar. Aurora, CO.

- 3/02/2015 "Topic modeling with structured-priors for text-driven science." University of Maryland, Baltimore County. Baltimore, MD.
- 2/27/2015 "Topic modeling with structured-priors for text-driven science." University of Colorado Boulder. Boulder, CO.
- 2/23/2015 "Topic modeling with structured-priors for text-driven science." Rice University. Houston, TX.
- 2/18/2015 "Topic modeling with structured-priors for text-driven science." University of California, Santa Barbara. Santa Barbara, CA.
- 1/21/2015 "Making sense of the web for public health using NLP." NIH National Cancer Institute (NCI). Rockville, MD.
- 12/03/2014 "Making sense of the web for public health using NLP." University of Maryland, College Park, CLIP Colloquium. College Park, MD.
- 5/29/2014 "Using social media to study public health." Café Scientifique. Annapolis, MD.
- 9/16/2013 "Data mining social websites for public health." Association of Public Data Users. Washington, DC.
- 12/08/2010 "Summarizing contrastive viewpoints in opinionated text." University of Maryland, College Park, CLIP Colloquium. College Park, MD.

STUDENT ADVISING

- PhD Advisor
- Ashlynn Daughton (Information Science, Fall 2016 – Present)
 - Yoshinari Fujinuma (Computer Science, Fall 2016 – Present)
 - Xiaolei Huang (Information Science, Fall 2016 – Present)
 - Shudong Hao (Computer Science, Fall 2017 – Present)
 - Dianna Radpour (Information Science, upcoming, Fall 2018)
 - Dasha Pruss (Information Science, Fall 2016 – Spring 2017)
- PhD Committee
- University of Colorado
- Forough Poursabzi (Computer Science, defended Summer 2018)
 - Soheil Danesh (Computer Science, defended Spring 2016)
- External committee member
- Jason Cho (Computer Science, University of Illinois at Urbana-Champaign)
- MS Advisor or Mentor
- Erik Holbrook (Computer Science, Fall 2016 – Spring 2018)
 - Linzi Xing (Computer Science, Spring 2017 – Spring 2018)
 - Sachin Muralidhara (Computer Science, Fall 2016 – Spring 2017)
 - Shiliang Wang (Computer Science, Johns Hopkins, Fall 2013)
 - Michael C. Smith (Computer Science, Johns Hopkins, Fall 2013 – Spring 2014)

- BS Mentor
- Jasmine Bascom (Computer Science, Summer 2018)
 - Justin Cai (Computer Science, Spring 2017 – Summer 2017)
 - Bradley Arnot (Computer Science, Fall 2016 – Spring 2017)
 - Branden Romero (Computer Science, Summer 2016)
 - Dmytro Ryzhkov (Computer Science, Summer 2016)
 - Alex Lamb (Computer Science, Johns Hopkins, Spring 2012 – Fall 2012)
 - Hieu Tran (Computer Science, Johns Hopkins, Spring 2012 – Fall 2012)

PROFESSIONAL EXPERIENCE

October 2011 – Present	<i>Advisory Board</i> Sickweather LLC	Baltimore, MD
Summer 2013, Summer 2014	<i>Intern</i> Microsoft Research Mentors: Eric Horvitz, Ryen White	Redmond, WA
Summer 2011	<i>Intern</i> Twitter Mentor: Alek Kolcz	San Francisco, CA
Fall 2009 – Summer 2010	<i>Research Programmer</i> University of Illinois, Dept. of Computer Science Advisor: Dan Roth	Urbana, IL
Summer 2009	<i>Research Assistant</i> University of Illinois, Dept. of Linguistics Advisor: Roxana Girju	Urbana, IL
Summer 2008	<i>Research Fellow</i> Univ. of Illinois, Data Sciences Summer Institute	Urbana, IL

ACADEMIC SERVICE

Organizer / Co-organizer	<ul style="list-style-type: none"> • NSF Workshop on Behavioral Ontology Learning <ul style="list-style-type: none"> ○ 8/08/2017 – 8/09/2017, Estes Park, Colorado • AAAI Workshop on Health Intelligence (W3PHI/HIAI) <ul style="list-style-type: none"> ○ 2/03/2018, New Orleans, LA ○ 2/04/2017, San Francisco, CA • Social Media Mining for Health Applications Workshop <ul style="list-style-type: none"> ○ EMNLP 2018: 10/31/2018 (upcoming), Brussels, Belgium ○ AMIA 2017: 11/04/2017, Washington, DC ○ PSB 2016: 1/08/2016, Big Island of Hawaii • ICWSM Social Media and Health Workshop <ul style="list-style-type: none"> ○ 6/25/2018, San Francisco, CA • Third Annual Mid-Atlantic Student Colloquium on Speech, Language and Learning (MASC-SLL) <ul style="list-style-type: none"> ○ 10/12/2013, University of Maryland Baltimore County
-----------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Journal
Reviewer

Computer and information science journals:

- Social Network Analysis and Mining (SNAM) (2017, 2018)
- Journal of Information Science (2018)
- Journal of Artificial Intelligence Research (JAIR) (2016)
- Human-Computer Interaction (2015)
- Transactions on Information Systems (TOIS) (2014)
- Journal of Natural Language Engineering (JNLE) (2012, 2013)
- Knowledge and Information Systems (KAIS) (2013)

Health and general science journals:

- PLOS ONE (2014-2016)
- Journal of the American Medical Informatics Association (JAMIA) (2015, 2018)
- Journal of Medical Internet Research (JMIR) (2016-2018)
- NPJ Digital Medicine (2018)
- PeerJ (2018)
- Journal of Biomedical Informatics (2017)
- American Journal of Preventative Medicine (2017)
- International Journal of Geographic Information Science (2016)
- Health Informatics Journal (2014)
- American Journal of Preventative Medicine (2014, 2016)
- Nicotine & Tobacco Research (2013, 2014)

Conference
Roles

- EMNLP 2017: Co-chair of Demo Track

Program
Committee

Conferences:

- AAAI Conference on Artificial Intelligence (AAAI): 2016-2018
- Association for Computational Linguistics (ACL): 2014-2018
- North American Chapter of ACL (NAACL): 2012-2013, 2015-2016, 2018
- European Chapter of ACL (EACL): 2014, 2017
- Empirical Methods in Natural Language Processing (EMNLP): 2015-2016, 2018
- Computational Natural Language Learning (CoNLL): 2018
- International World Wide Web Conference (WWW): 2017-2018
- International Conference on Weblogs and Social Media (ICWSM): 2014, 2017-2018
- ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD): 2015
- Conference on Lexical and Computational Semantics (*SEM): 2013
- International Conference on Machine Learning (ICML): 2015

Workshops

- Workshop on Natural Language Processing for Social Media (SocialNLP): EMNLP 2016, WWW 2018
- Workshop on Noisy User-generated Text (W-NUT): EMNLP 2018
- Workshop on Natural Language Processing and Computational Social Science (NLP+CSS): EMNLP 2016
- International Workshop on the World Wide Web and Population Health Intelligence (W3PHI): AAAI 2016

- Workshop on Entity Attribute Prediction in Social Media: ACL 2014
 - Workshop on Topic Models: NIPS 2013
- Ad-hoc Reviewer
- CHI 2018 (2 reviews), CHI 2017 (1 review), PSB 2017 (1 review), AAAI 2014 (1 review)
- Secondary Reviewer
- AAAI 2014, WWW 2014, NIPS 2011, NAACL 2010
- Grant Reviewer
- Panels:
- NIH PAR Panel: Cancer Communication in the New Media Environment (April 3, 2017)
 - NSF IIS-III panel reviewer (April 14, 2016)
- Ad-hoc reviewing:
- Chilean National Science and Technology Commission (CONICYT) – FONDECYT Regular 2018 grant competition (1 ad-hoc review) (October 2017)
 - NSF Smart & Connected Health (1 ad-hoc review) (March 2016)
 - Sloan Foundation (1 ad-hoc review) (2015)
- University Service
- Cloud working group (CU) (2017)
- College Service
- IT committee (CU) (Spring 2016)
 - Website committee (CU) (Fall 2015, Spring 2016)
- Departmental Service
- Seminar organizer (Spring 2018 – Present)
 - Undergraduate committee (CU) (Fall 2016 – Present)
 - BS curriculum committee (CU) (Fall 2015, Spring 2016)
 - PhD curriculum committee (CU) (Fall 2015, Spring 2016)
 - PhD admissions committee (CU) (Fall 2015, Spring 2016)
 - Faculty search committee (CU) (Fall 2015, Spring 2016, Spring 2018)
 - Student admissions committee (Johns Hopkins) (Spring 2015)
 - Student recruiting committee (Johns Hopkins) (Spring 2011, Spring 2012)