Michael C. Hughes

+1-617-371-76 one one • mike@michaelchughes.com
www.michaelchughes.com • michaelchughes

Research Expertise

Machine Learning *2012-*
- Probabilistic models, Bayesian nonparametrics, approximate inference, semi-supervised learning

Clinical Informatics *2017-*
- Personalized risk and treatment prediction, phenotype discovery

Data Analysis *2012-*
- Multivariate time series, images and videos, electronic health records

Education

Brown University
Ph.D., Computer Science. *2016*

Brown University
M.S., Computer Science. *2012*

Olin College of Engineering
B.S. Electrical & Computer Engineering *2010*

Research Experience

Assistant Professor of Computer Science
Tufts University, Medford, MA *2018 - present*
- Research statistical machine learning methods and applications to health informatics.
- Advise Ph.D., M.S., and B.S. students in machine learning research projects.
- Teach machine learning courses targeted at advanced undergraduates (COMP 135 Intro to ML) and graduate students (COMP 150 Bayesian Deep Learning).
- Appointed as the Ann W. Lambertus and Peter Lambertus Assistant Professor in 2019

Postdoctoral fellow: Machine learning to improve clinical decisions in healthcare
Adviser: Prof. Finale Doshi-Velez (Harvard) *2016 - 2018*
- Developed semi-supervised models for characterizing and treating depression (with Dr. Perlis and Dr. McCoy).
- Applied time-series models to predict ventilator interventions in the ICU for public dataset of >36,000 patients.
- Created methods for training deep models so they are more interpretable to clinicians or other users.

Postdoc project: Estimating carbon biomass from LiDAR waveforms

Adviser: Prof. Erik Sudderth & Prof. Jim Kellner (Brown U., Ecology & Evolutionary Biology) 2016

- Predicted forest biomass from LiDAR waveforms to better understand land use and climate change.
- Modeled waveforms and biomass predictions jointly via nonparametric regression using our BNPy toolbox.
- Intended for use in NASA's upcoming Global Ecosystem Dynamics Investigation (GEDI).

Ph.D. thesis: Reliable and scalable variational inference for Bayesian nonparametrics

Adviser: Prof. Erik Sudderth *2016*

- Thesis Title: Reliable and scalable variational inference for nonparametric mixtures, topics, and sequences
- Developed optimization algorithms for Bayesian nonparametric models that scale to millions of examples.
- Optimized lower bound on marginal likelihood, thus penalizing too simple and too complex explanations.
- Escaped local optima via data-driven proposals that add useful new clusters and remove redundant ones.
- Applied to topic models of 2 million NY Times articles and sequential models of the whole human genome.
- Implemented algorithms in open-source package: Bayesian Nonparametrics for Python (BNPy).

Master's project: Sequential Models for Video and Motion Capture

Adviser: Prof. Erik Sudderth *2012*

- Developed methods to discover common actions from many videos of humans performing household exercises.
- Improved existing inference algorithms with data-driven Metropolis-Hastings proposals.

Highlighted Publications

1. "Regional Tree Regularization for Interpretability in Black Box Models." Mike Wu[d], Sonali Parbhoo, Michael C. Hughes, Ryan Kindle, Leo Celi, Maurizio Zazzi, Volker Roth, and Finale Doshi-Velez. To Appear at AAAI 2020, 2020.
2. "Rapid Model Comparison by Amortizing Across Models." Lily H. Zhang[b], and Michael C. Hughes. To Appear at Second Symposium on Advances in Approximate Bayesian Inference (AABI 2019), 2019.
3. "Feature Robustness in Non-stationary Health Records: Caveats to Deployable Model Performance in Common Clinical Machine Learning Tasks." Bret Nestor[d], Matthew B. A. McDermott, Willie Boag, Gabriela Berner, Tristan Naumann, Michael C. Hughes, Anna Goldenberg, and Marzyeh Ghassemi. Machine Learning for Healthcare, 2019.
4. "Supervised Machine Learning Algorithms Using Patient Related Factors to Predict in-Hospital Mortality Following Acute Myeloid Leukemia Therapy." Nauman Saleem Siddiqui[c], Andreas Klein, Amandeep Godara, Cindy Varga, Rachel J. Buchsbaum, and Michael C. Hughes. Proceedings of 61st Annual Meeting of the American Hematology Society, 2019.
5. "Prediction-Constrained POMDPs." Joseph Futoma, Michael C. Hughes, and Finale Doshi-Velez. Reinforcement Learning under Partial Observability (RLPO) workshop at NeurIPS 2018, 2018.
6. "Semi-Supervised Prediction-Constrained Topic Models." Michael C. Hughes, Gabriel Hope[d], Leah Weiner[d], Thomas H. McCoy Jr, Roy H. Perlis, Erik B. Sudderth, and Finale Doshi-Velez. Artificial Intelligence and Statistics (AISTATS), 2018.
7. "Beyond Sparsity: Tree Regularization of Deep Models for Interpretability." Mike Wu[u], Michael C. Hughes, Sonali Parbhoo, Maurizio Zazzi, Volker Roth, and Finale Doshi-Velez. Association for Advancement of Artificial Intelligence (AAAI), 2018.
8. "From Patches to Images: A Nonparametric Generative Model." Geng Ji[d], Michael C. Hughes, and

Erik B. Sudderth. International Conference on Machine Learning (ICML), 2017.

Superscripts indicate mentored student's status: u = undergraduate, m = masters, d = Ph.D. student, b = post-bacc, c = medical student. Complete publication list at end of this document.

Highlighted Preprints

1. "MIMIC-Extract: A Data Extraction, Preprocessing, and Representation Pipeline for MIMIC-III." Shirly Wang[m], Matthew B. A. McDermott, Geeticka Chauhan, Michael C. Hughes, Tristan Naumann, and Marzyeh Ghassemi. In Preparation for Journal Submission, 2019.

Honors and Awards

Top 400 Reviewer Award, NeurIPS 2019 *2019*

- Recognized as one of top 400 of more than 3500 expert reviewers at the top international machine learning conference.

Top 200 Reviewer Award, NeurIPS 2018 *2018*

- Recognized as one of top 200 of more than 3500 expert reviewers at the top international machine learning conference.

Best Paper Award, SoCal NLP Symposium 2018 *2018*

- Awarded for 2 page summary of our AISTATS 2018 paper.

Nominee for AMIA Clinical Informatics Research Award *2017*

- 1 of 7 papers nominated at AMIA's 2017 Joint Summits on Translational Science, out of >50 papers.

NSF Graduate Research Fellowship Award *2011*

- Three year award to fund Ph.D. studies. Covers tuition and provides research stipend.

NDSEG Graduate Research Fellowship Award *2011*

- Three year funding award. Declined to accept NSF fellowship.

Current Funding Support

Amortized Inference for Large-Scale Graphical Models

NSF CISE: Robust Intelligence: Small *9/1/19 - 8/31/22*

- Co-Investigators: Liping Liu (PI, Tufts CS) and Thomas Stopka (Tufts Public Health)
- Total Amount: $399,923

Estimating Individual Treatment Effects from Randomized Clinical Trials using Machine Learning

Tufts Collaborates Award (Internal) *7/1/19 - 6*

- Co-Investigators: David Kent (Tufts Medical Center)
- Total Amount: $50,251

MASTR-E: Machine Learning for Human Performance Prediction & Down-selection
U.S. Army NSRDEC, Natick *6/24/19 - 6/23/20*
- Co-Investigators: Eric Miller (PI, Tufts ECE) and Shuchin Aeron (Tufts ECE)
- Total Amount: $252,348

Invited Talks

Invited Talk at BNP 2019 *2019*
- Title: Scalable and Reliable Variational Inference for Dirichlet Process Clustering with Sparse Assignments
- Venue: 12th International Conference on Bayesian Nonparametrics
- Summarizes the effective learning methods behind our BNPy toolbox

Invited Mentor at 2019 PLA General Hospital - MIT Critical Data Datathon *2019*
- 4th annual PLAGH-MIT Datathon
- Event held in Beijing, China with 25 teams of local clinicians and computational scientists
- Team goal: Answer clinical question on historical public dataset (MIMIC) over 1 weekend
- Event goal: Develop local teams' skills via intense practice with expert oversight
- My role: Advise teams toward principled and clinically-useful analysis

Invited Tutorial at MLHC 2018 *2018*
- Machine Learning for Clinicians: Advances for Multi-Modal Health Data at MLHC '18
- Designed to help clinicians understand enough modern machine learning to collaborate successfully with ML researchers.

Invited Panelist *2016*
- Software panel at Advances in Approximate Bayesian Inference workshop at NIPS '16.

Professional Service

Senior Program Committee / Meta-Reviewer
- 2020 - AAAI

Program Committee / Reviewer
- 2020 - AISTATS, ICLR, AMIA CRI
- 2019 - NIPS, AISTATS, ICLR
- 2018 - NIPS (reviewer award), AAAI, AISTATS, ICLR, AMIA CRI
- 2017 - NIPS, ICML, AAAI
- 2016 - NIPS
- 2015 - NIPS, ICML
- 2014 - NIPS, ICML
- 2013 - NIPS (reviewer award)

Workshop Organizer: ML4H at NeurIPS 2018 *2018*
- Machine Learning for Health workshop at NeurIPS '18 (NeurIPS ML4H 2018).
- Full-day workshop with invited keynotes, accepted papers/posters, and lively panel discussions.
- Provided a forum for interdisciplinary interaction between clinicians, statisticians, and computer scientists.
- Helped with website, PR, and continuity in peer-review process from previous years.

Workshop Organizer: BNP at NeurIPS 2018 *2018*

- All of Bayesian Nonparametrics workshop at NeurIPS '18 (NeurIPS BNP 2018).
- Full-day workshop with invited keynotes, accepted papers/posters, and lively panel discussions.
- Helped with peer-review process for accepted posters, https://sites.google.com/view/nipsbnp2018/schedule.

Workshop Organizer: ML4H at NeurIPS 2017 *2017*

- Machine Learning for Health workshop at NIPS '17 (NIPS ML4H 2017)
- Full-day workshop with invited keynotes and panels involving clinicians, statisticians, and computer scientists.
- Organized peer-review process for 118 submitted papers.

Workshop Organizer: BNP at NeurIPS 2016 *2016*

- Practical Bayesian Nonparametrics workshop at NIPS '16.
- Full-day workshop with invited speakers, contributed talks, two panel discussions, and lively poster session.
- Led decisions on >25 submitted papers based on peer review.

Teaching and Mentorship

Tufts CS Dept.

Course: COMP 150 Bayesian Deep Learning *Fall 2019*

- Taught advanced topics seminar to 23 students
- Course format: weekly homeworks for first month, then 2-month open-ended team project

Tufts CS Dept.

Course: COMP 135 Introduction to Machine Learning *Spring 2019*

- Taught core principles of machine learning to about 50 students
- Course format: 3 open-ended projects, weekly homeworks, and 2 exams

Tufts CS Dept.

Course: COMP 150 Bayesian Deep Learning *Fall 2018*

- Taught advanced topics seminar to about 18 students
- Course format: weekly homeworks for first month, then 2-month open-ended team project
- One project resulted in publication at IEEE conference (ICDL-EpiRob 2019)

Tufts University

Research Mentor *2018-*

- Mentored undergraduate projects:
- – Manh Duc Nguyen. 2019 CS honors thesis: Particle-based algorithms for Bayesian Neural Networks - Hamiltonian Monte Carlo and Stein Variational Gradient Descent

Harvard University SEAS

Research Mentor *2016-2017*

- Mentored undergraduate senior thesis projects on Bayesian nonparametric inference.
- Frederick Widjaja. 2017 honors thesis: Streaming Variational Inference for the Indian Buffet Process.
- Madhu Vijay. 2017 honors thesis: Characterizing Posterior Uncertainty for the Indian Buffet Process.

Brown University

Research Mentor *2014-2016*

- Mentored students on projects related to Bayesian nonparametric clustering and the BNPy Python package.
- William Stephenson. 2015 undergraduate honors thesis: Variational Inference for Hierarchical Dirichlet Process based Nonparametric Models.
- Sonia Phene. 2015 undergraduate honors thesis: Multiprocessor Parallelization of Variational Inference for Bayesian Nonparametric Topic Models.
- Mengrui Ni. 2015 masters project: Variational Inference for Beta-Bernoulli Dirichlet Process Mixture Models.
- Mert Terzihan. 2015 masters project.

Lead Graduate TA for CS 142: Intro to Machine Learning *Fall 2013*

- Led weekly 1 hour recitation session to review key concepts for 50+ students.
- Designed homework assignments and exam questions.

Outreach Experience

TEALS and Boston Latin Academy **Roxbury, MA**

Volunteer AP Computer Science Instructor *2014-2016*

- Taught 1-2 classes / week for 2 years via TEALS "CS in every high school" initiative sponsored by Microsoft.
- Developed hands-on lessons to excite students from diverse backgrounds about computational thinking.
- Mentored full-time teacher Ingrid Roche as she transitioned from media arts to AP computer science (Java).

Harvard Humanitarian Initiative **Cambridge, MA**

Signal Program Fellow *2014*

- Developed prototype detector for common housing structures in sub-Saharan Africa from satellite images.
- Intended for humanitarian oversight of conflict areas where burning structures is common attack pattern.
- Featured in TEDx talk: `https://youtu.be/u7l9rBwOnwU`

Olin College Engineering Discovery **Needham, MA**

Co-Founder and Curriculum Director *2007-2010*

- Managed 15 undergrads in developing hands-on lessons for 4th-8th graders.
- Hosted workshops for 30 children to design, build, and launch bottle rockets.
- Pioneered green energy workshop which earned over $750 in outside funding.

Industry Experience

Google **Mountain View, CA**

Software Engineering Intern *Summer 2013*

- Improved walking/biking/running classifier using smartphone accelerometer data.
- Led collection of dataset from dozens of individuals for classifier evaluation via custom Android app.

All Conference Publications (in reverse chronological order)

1. "Regional Tree Regularization for Interpretability in Black Box Models." Mike Wu[d], Sonali Parbhoo, Michael C. Hughes, Ryan Kindle, Leo Celi, Maurizio Zazzi, Volker Roth, and Finale Doshi-Velez. To Appear at AAAI 2020, 2020.

2. "Feature Robustness in Non-stationary Health Records: Caveats to Deployable Model Performance in Common Clinical Machine Learning Tasks." Bret Nestor[d], Matthew B. A. McDermott, Willie Boag, Gabriela Berner, Tristan Naumann, Michael C. Hughes, Anna Goldenberg, and Marzyeh Ghassemi. Machine Learning for Healthcare, 2019.
3. "Supervised Machine Learning Algorithms Using Patient Related Factors to Predict in-Hospital Mortality Following Acute Myeloid Leukemia Therapy." Nauman Saleem Siddiqui[c], Andreas Klein, Amandeep Godara, Cindy Varga, Rachel J. Buchsbaum, and Michael C. Hughes. Proceedings of 61st Annual Meeting of the American Hematology Society, 2019.
4. "Sensorimotor Cross-Behavior Knowledge Transfer for Grounded Category Recognition." Gyan Tatiya[d], Ramtin Hosseini[d], Michael C. Hughes, and Jivko Sinapov. Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob), 2019.
5. "Semi-Supervised Prediction-Constrained Topic Models." Michael C. Hughes, Gabriel Hope[d], Leah Weiner[d], Thomas H. McCoy Jr, Roy H. Perlis, Erik B. Sudderth, and Finale Doshi-Velez. Artificial Intelligence and Statistics (AISTATS), 2018.
6. "Beyond Sparsity: Tree Regularization of Deep Models for Interpretability." Mike Wu[u], Michael C. Hughes, Sonali Parbhoo, Maurizio Zazzi, Volker Roth, and Finale Doshi-Velez. Association for Advancement of Artificial Intelligence (AAAI), 2018.
7. "From Patches to Images: A Nonparametric Generative Model." Geng Ji[d], Michael C. Hughes, and Erik B. Sudderth. International Conference on Machine Learning (ICML), 2017.
8. "Right for the Right Reasons: Training Differentiable Models by Constraining their Explanations." Andrew Slavin Ross[m], Michael C. Hughes, and Finale Doshi-Velez. International Joint Conference on Artificial Intelligence (ICJAI), 2017.
9. "Predicting Intervention Onset in the ICU with Switching State Space Models." Marzyeh Ghassemi, Mike Wu[u], Michael C. Hughes, Peter Szolovits, and Finale Doshi-Velez. AMIA Summit on Clinical Research Informatics, 2017.
10. "Scalable Adaptation of State Complexity for Nonparametric Hidden Markov Models." Michael C. Hughes, William Stephenson[u], and Erik B. Sudderth. Neural Information Processing Systems (NIPS), 2015.
11. "Reliable and Scalable Variational Inference for the Hierarchical Dirichlet Process." Michael C. Hughes, Dae Il Kim, and Erik B. Sudderth. Artificial Intelligence & Statistics (AISTATS), 2015.
12. "Memoized Online Variational Inference for Dirichlet Process Mixture Models." Michael C. Hughes and Erik B. Sudderth. Neural Information Processing Systems (NIPS), 2013.
13. "Effective Split-Merge Monte Carlo Methods for Nonparametric Models of Sequential Data." Michael C. Hughes, Emily Fox, and Erik B. Sudderth. Neural Information Processing Systems (NIPS), 2012.
14. "The Nonparametric Metadata Dependent Relational Model." Dae Il Kim, Michael C. Hughes, and Erik B. Sudderth. International Conference on Machine Learning (ICML), 2012.

All Journal Publications (in reverse chronological order)

1. "Predicting Treatment Discontinuation after Antidepressant Initiation.." Melanie F. Pradier, Thomas H. McCoy, Michael C. Hughes, Roy H. Perlis, and Finale Doshi-Velez. To Appear in Translational Psychiatry, 2019.
2. "Refinery: An Open Source Topic Modeling Web Platform." Daeil Kim, Benjamin F. Swanson, Michael C. Hughes, and Erik B. Sudderth. JMLR Machine Learning Open Source Software (MLOSS),

2017.

3. "Joint Modeling of Multiple Time Series via the Beta Process with Application to Motion Capture Segmentation." Emily Fox, Michael C. Hughes, Erik B. Sudderth, and Michael I. Jordan. Annals of Applied Statistics, Vol. 8(3), 2014.

All Workshop Papers (in reverse chronological order)

1. "Rapid Model Comparison by Amortizing Across Models." Lily H. Zhang[b], and Michael C. Hughes. To Appear at Second Symposium on Advances in Approximate Bayesian Inference (AABI 2019), 2019.
2. "Classification of Enzyme Promiscuity Using Positive, Unlabeled, and Hard Negative Examples." Gian Marco Visani, Michael C. Hughes and Soha Hassoun. To Appear at Machine Learning in Computational Biology Workshop (MLCB), 2019.
3. "Prediction-Constrained POMDPs." Joseph Futoma, Michael C. Hughes, and Finale Doshi-Velez. Reinforcement Learning under Partial Observability (RLPO) workshop at NeurIPS 2018, 2018.
4. "Rethinking clinical prediction: Why machine learning must consider year of care and feature aggregation." Bret Nestor[d], Matthew B. A. McDermott, Geeticka Chauhan, Tristan Naumann, Michael C. Hughes, Anna Goldenberg, Marzyeh Ghassemi. Machine Learning for Healthcare (ML4H) workshop at NeurIPS 2018, 2018.
5. "Prediction-Constrained Topic Models for Antidepressant Prediction." Michael C. Hughes, Gabriel Hope[d], Leah Weiner[d], Thomas H. McCoy, Roy H. Perlis, Erik B. Sudderth, and Finale Doshi-Velez. NIPS Workshop on Machine Learning for Health (NIPS ML4H), 2017.
6. "Associations between aboveground forest biomass and waveform LiDAR metrics: implications for modeling footprint-level biomass using Global Ecosystem Dynamics Investigation data." J. Kellner, J. B. Blair, L. Duncanson, L., S. Hancock, M. A. Hofton, M. C. Hughes, S. Marselis, S., J. Armston, E. B. Sudderth, H. Tang, L. Weiner[d], and R. Dubayah. American Geophysical Union, Fall General Assembly, 2016.
7. "Supervised topic models for clinical interpretability." Michael C. Hughes, Huseyin Melih Elibol, Thomas McCoy, Roy Perlis, and Finale Doshi-Velez. NIPS Workshop on Machine Learning for Health (NIPS ML4H), 2016.
8. "Nonparametric Discovery of Activity Patterns from Video Collections." Michael C. Hughes and Erik B. Sudderth. CVPR Workshop on Perceptual Organization in Computer Vision (POCV), 2012.