

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Hughes, Michael C.

eRA COMMONS USER NAME (credential, e.g., agency login): N/A

POSITION TITLE: Assistant Professor of Computer Science

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE (if applicable)	Completion Date MM/YYYY	FIELD OF STUDY
Franklin W. Olin College of Engineering	B.S.	05/2010	Electrical & Computer Engineering
Brown University	M.S.	05/2012	Computer Science
Brown University	Ph.D.	05/2016	Computer Science
Harvard University	(post-doc)	08/2018	Computer Science

A. Personal Statement

My research goal is to develop machine learning methods that discover actionable knowledge from large, messy datasets. I take a results-oriented approach to model development, iteratively adding principled features to statistical models to capture key properties of the underlying data and task then carefully validating using application-specific performance metrics. My expertise in statistical machine learning, optimization algorithms, biomedical informatics, and open-source software engineering makes me uniquely positioned to achieve these goals. Armed with this background, I am excited to pursue an interdisciplinary collaboration in developing computational approaches to interpret untargeted metabolomics measurements.

My previous work in developing new methods is often centered around applications in the biomedical sciences and healthcare. I have worked on methods that are applied to chromatin state discovery in epigenomics (Hughes et al. NeurIPS 2015) as well as several projects applying machine learning to electronic health records (Ghassemi et al. AMIA CRI 2017; Hughes et al. AISTATS 2018). I also have a track record of building and evaluating models for networks (Kim et al. ICML 2012). My research provides better methods to the machine learning community, patient-specific treatment advice to doctors, and scientific insight into possible subtypes or trajectories of targeted diseases. I strive to deliver top-tier publications in both computational and applied venues as well as reproducible open-source software and deployed predictive systems.

B. Positions and HonorsEmployment

2016-2018 Postdoctoral Fellow, School of Engineering and Applied Sciences, Harvard University

2018- Assistant Professor, Department of Computer Science, Tufts University

Honors

2018 Best Paper Award, SoCal Natural Language Processing (NLP) Symposium
2017 Nominee for AMIA Clinical Informatics Research Award (1 of 7)
2011 NSF Graduate Research Fellowship Award

C. Contributions to Science

1. Semi-supervised models for personalized medicine from large observational EHR data

Observational medical datasets contain orders of magnitude more data than the largest randomized controlled trials and thus offer promise as a source of information about why drugs might be effective for some patients but not others or finding useful subgroups of patients with similar symptoms, responses, or trajectories. However, there exist many challenges in using such data to draw meaningful conclusions, including learning from many sources of data that are often noisy and high-dimensional and determining clinically-useful outcomes (which drugs were successful, not just attempted) when manual annotation is unavailable or too expensive. My postdoc work at the intersection of machine learning and clinical informatics tries to overcome these barriers.

For predicting needed interventions in the intensive care unit (ICU), our AMIA CRI 2017 paper applied switching-state models to time-series data from 36,000 ICU patients. Although trained in an unsupervised way (discovering common patient states without labels), our learned representations improved 4-hour look-ahead predictions of need for mechanical ventilation. Such models can improve personalized care and smooth logistics in the busy ICU, but to reach higher accuracies we must supervise training by including labels.

For our work on antidepressant recommendation for 50,000+ patients in Boston-area hospitals, we have developed a semi-supervised dimensionality reduction approach that learns from few labeled examples and many unlabeled ones. Supervised topic models which explain both the data and predict the labels can achieve two needed goals: good generative models of data (scientific insight on possible patient subgroups or important variable interactions) and good prediction of labels from data (effective personalized treatment recommendations). We have shown these methods offer competitive predictions that generalize well to new hospital systems.

M. C. Hughes, L. Weiner, G. Hope, T. H. McCoy Jr, R. H. Perlis, E. B. Sudderth, and F. Doshi-Velez
Semi-supervised prediction-constrained topic models.

In Artificial Intelligence and Statistics (AISTATS), 2018.

https://www.michaelchughes.com/papers/HughesEtAl_AISTATS_2018.pdf

M. Ghassemi*, M. Wu*, **M. C. Hughes***, P. Szolovits, and F. Doshi-Velez
Predicting intervention onset in the ICU with switching state space models.

In AMIA Summit on Clinical Research Informatics, 2017

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5543372/pdf/2611344.pdf>

2. Reliable inference for Bayesian nonparametric unsupervised learning of structured or networked datasets

During my Ph.D. thesis, I made core contributions to statistical machine learning by inventing new algorithms to train a broad class of Bayesian nonparametric (BNP) models for unsupervised structure discovery, including mixture models, models for social or ecological networks (Kim et al. ICML 2012), topic models for grouped data, and hidden Markov models for sequential data. While well-known parametric methods like k-means require an a priori specification of the exact number of clusters, BNP models, such as the Dirichlet process mixture model, can manage uncertainty in the number of clusters needed to explain a dataset and let the data determine model complexity without restrictive assumptions.

While this class of models is promising for many real datasets, before our work standard inference methods such as sampling or variational optimization did not live up to this promise. Such methods optimize non-convex objective functions via iterative updates to subsets of parameters. Even on small data, random initializations become stuck at poor local optima and fail to reach even the best single mode, let alone explore the real posterior. Scaling to millions of examples only exaggerates this problem. My Ph.D. work on reliable BNP suggests that these standard updates should be interleaved with data-driven proposals that create new clusters to better explain some examples or remove irrelevant clusters. Our proposal-driven algorithms reach qualitatively better solutions without expensive restarts or cross-validation. Experiments demonstrate the usefulness of our approach in discovering clusters from large collections of news articles (every New York Times article in last 20 years) and discovering recurring patterns of epigenetic modifications to the human genome. My collaborators and I released an open-source Python package called BNPy (Bayesian Nonparametrics for Python, <https://bnpy.readthedocs.io>) to make these techniques broadly available.

D. Kim and M. C. Hughes and E. B. Sudderth
The Nonparametric Metadata Dependent Relational Model
In International Conference on Machine Learning (ICML), 2012
https://www.michaelchughes.com/papers/KimHughesSudderth_ICML_2012.pdf

M. C. Hughes and E. B. Sudderth.
Memoized online variational inference for Dirichlet process mixture models.
In Neural Information Processing Systems (NeurIPS), 2013.
https://www.michaelchughes.com/papers/HughesSudderth_NIPS_2013.pdf

M. C. Hughes, D. I. Kim, and E. B. Sudderth.
Reliable and scalable variational inference for the hierarchical Dirichlet process.
In Artificial Intelligence and Statistics (AISTATS), 2015.
https://www.michaelchughes.com/papers/HughesKimSudderth_AISTATS_2015.pdf

M. C. Hughes, W. Stephenson and E. B. Sudderth.
Scalable adaptation of state complexity for nonparametric hidden Markov models.
In Neural Information Processing Systems (NeurIPS), 2015.
https://www.michaelchughes.com/papers/HughesStephensonSudderth_NIPS_2015.pdf

3. Optimizing deep neural network classifiers for interpretability

While deep learning has made impressive strides on image and language tasks, many clinical practitioners are reluctant to adopt deep models because their predictions are difficult to interpret, explain, and trust. During my postdoc, I've pursued new machine learning methods to close this trust gap.

First, our IJCAI 2017 paper shows how to train models to respect expert annotations of features relevant to specific examples. Furthermore, even without expert annotations our method can actively discover models with different per-example decision rationales via the same regularization process. This gives practitioners valuable tools to debug pretrained models or discover confounding features hiding in their datasets.

Second, our AAAI 2018 paper introduces "tree regularization", a method to directly train deep models to be human-simulatable. Small decision trees with only a few nodes make it easy for humans to step through the entire prediction process, and thus enjoy widespread use in manual medical diagnosis. In contrast, feed-forward networks with a dozen hidden units can have far too many parameters and connections for a human to simulate. Deep models for sequential data are even more challenging. Simulatability allows clinicians to audit predictions easily. They can manually inspect how outputs change under slightly-perturbed inputs or identify when predictions are made due to systemic bias or confounders rather than real causes. Our work shows that recurrent neural networks for predicting treatments for patients with sepsis or HIV can be trained to have simpler, tree-like decision boundaries with fewer than 25 nodes, while still predicting better than standalone trees.

A. S. Ross, **M. C. Hughes**, and F. Doshi-Velez.

Right for the right reasons: Training differentiable models by constraining their explanations.
In International Joint Conference on Artificial Intelligence, 2017.

https://www.michaelchughes.com/papers/RossHughesDoshiVelez_IJCAI_2017.pdf

M. Wu, **M. C. Hughes**, S. Parbhoo, M. Zazzi, V. Roth, and F. Doshi-Velez.

Beyond sparsity: Tree regularization of deep models for interpretability.

In AAAI Conference on Artificial Intelligence, 2018.

<https://arxiv.org/pdf/1711.06178.pdf>

D. Additional Information: Research Support and/or Scholastic Performance

As a new faculty member (started August 2018), I have no external funding to date.

My postdoctoral work was funded by Oracle Labs.