John T. Halloran

CONTACT INFORMATION

Da Vinci Building
1632 Da Vinci Ct.
Davis, CA 95618

E-mail: jthalloran@ucdavis.edu
URL: https://jthalloran.bitbucket.io

RESEARCH INTERESTS

Machine learning algorithms for HPC architectures, parameter learning with theoretical guarantees, scalable large-scale learning algorithms, interpretable machine learning models, massive-scale problems in computational biology

EDUCATION AND EMPLOYMENT

- **2016-Present** Postdoctoral Scholar, University of California, Davis
- **2010-2016** PhD, University of Washington
  Department of Electrical Engineering
  Advisors: Jeff A. Bilmes and William S. Noble
- **2008-2010** MS, University of Hawaii at Manoa
  Department of Electrical Engineering
  Advisors: Narayana P. Santhanam and Yingbin Liang
- **2008** Double B.S., Seattle University
  Department of Mathematics, Department of Electrical Engineering
  Research advisor: Agnieska Miguel

AWARDS AND GRANTS

- **Award for Excellence in Postdoctoral Research, 2019**, from the UC Davis Postdoctoral Scholar Association.
- Top reviewer for NeurIPS 2019
- Nvidia Hardware Grant, 2016 (Tesla K40 GPU awarded for deep learning research).
- **Genome Training Grant** from the University of Washington, 2011-2013 (tuition, stipend, and travel funds awarded).

PUBLICATIONS

[1] **JT Halloran** and DM Rocke. "GPU-Accelerated Primal Learning for Extremely Fast Large-Scale Classification." *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. **20% Acceptance rate, 1900 out of 9454 submissions**.

[2] **JT Halloran**, H Zhang, K Kara, C Renggli, M The, DM Rocke, L Käll, WS Noble, and C Zhang. "Speeding up Percolator." *Journal of Proteome Research (JPR)*, 2019.

[3] **JT Halloran** and DM Rocke. "Learning Concave Conditional Likelihood Models for Improved Analysis of Tandem Mass Spectra." *Advances in Neural Information Processing Systems (NeurIPS)*, 2018. **20.8% Acceptance rate, 1011 out of 4856 submissions**.

[4] **JT Halloran** and DM Rocke. "A Matter of Time: Faster Percolator Analysis via Efficient SVM Learning for Large-Scale Proteomics." *Journal of Proteome Research (JPR)*, 2018.

[5] **JT Halloran**. "Analyzing Tandem Mass Spectra using the DRIP Toolkit: Training, Searching, and Post-Processing." In *Data Mining for Systems Biology: Methods and Protocols*, Ed. by H Mamitsuka. Humana Press, 2018.

[6] **JT Halloran** and DM Rocke. "Gradients of Generative Models for Improved Discriminative Analysis of Tandem Mass Spectra." *Advances in Neural Information Processing Systems (NIPS)*, 2017. **Spotlight oral presentation; 3.5% Acceptance rate, 112 out of 3240 submissions**.

[7] J Liu, **J Halloran**, J Bilmes, R Daza, C Lee, E Mahen, D Prunkard, C Song, S Blau, M Dorschner, V Gadi, J Shendure, A Blau, and W Noble. "Comprehensive statistical inference of the clonal structure of cancer from multiple biopsies." *Scientific Reports*, 2017.

[8] **JT Halloran**, WS Noble, and JA Bilmes. "A dynamic Bayesian network for accurate detection of peptides from tandem mass spectra." *Journal of Proteome Research (JPR)*, 2016.

[9] S Wang, **JT Halloran**, JA Bilmes, and WS Noble. "Faster and more accurate graphical model identification of tandem mass spectra using trellises." *Bioinformatics (Proc. of the ISMB)*, 2016.

[10] **JT Halloran**, JA Bilmes, and WS Noble. "Learning Peptide-Spectrum Alignment Models for Tandem Mass Spectrometry." *Uncertainty in Artificial Intelligence (UAI)*, 2014.

[11] AP Singh, **J Halloran**, JA Bilmes, K Kirchhoff, and WS Noble. "Spectrum Identification using a Dynamic Bayesian Network Model of Tandem Mass Spectra." *Uncertainty in Artificial Intelligence (UAI)*, 2012.

[12] Y Liang, L Lai, and **J Halloran**. "Distributed cognitive radio network management via algorithms in probabilistic graphical models." *IEEE JSAC, Special Issue on Advances in Cognitive Radio Networking and Communications*, 2011.

[13] Y Liang, L Lai, and **J Halloran**. "Distributed algorithm for collaborative detection in cognitive radio networks." *The Allerton Conference on Communication, Control, and Computing*, 2009.

Contributed Abstracts and Workshops

[14] **JT Halloran** and DM Rocke. "GPU-Accelerated SVM Learning for Massive-Scale Proteomics Analysis." *ICML Workshop on Computational Biology (WCB@ICML)*, 2020.

[15] **JT Halloran** and DM Rocke. "GPU-Accelerated SVM Learning for Extremely Fast Large-Scale Proteomics Analysis." *Machine Learning in Computational Biology (MLCB) Meeting*, 2019.

[16] **JT Halloran**. "Analyzing Tandem Mass Spectra: A Graphical Models Perspective." *Proceedings of Machine Learning Research vol* 73 (2017): 6-6.

[17] **JT Halloran** and DM Rocke. "Fisher Kernels for Improved Analysis of Tandem Mass Spectra." *NIPS Workshop on Machine Learning in Computational Biology (MLCB)*, 2016.

[18] **JT Halloran**, AP Singh, JA Bilmes, and WS Noble. "Peptide Identification of Tandem Mass Spectra via Spectrum Alignment using a Dynamic Bayesian Network." *NIPS Workshop on Machine Learning in Computational Biology (MLCB)*, 2012.

[19] AP Singh, **J Halloran**, JA Bilmes, K Kirchhoff, and WS Noble. "Spectrum Identification with a Dynamic Bayesian Network model of Tandem Mass Spectra." *RECOMB Satellite Conference on Computational Proteomics*, 2012.

In Submission

[20] **JT Halloran**, G Urban, DM Rocke, and P Baldi. "Deep Semi-Supervised Learning Improves Universal Peptide Identification of Shotgun Proteomics Data."

Talks and Presentations

[1] NeurIPS, Vancouver, BC, Canada. December, 2020. "GPU-Accelerated Primal Learning for Extremely Fast Large-Scale Classification."

[2] ICML Workshop on Computational Biology, July 2020. "GPU-Accelerated SVM Learning for Massive-Scale Proteomics Analysis."

[3] Machine Learning in Computational Biology (MLCB) Meeting, Vancouver, BC, Canada. December, 2019. "GPU-Accelerated SVM Learning for Extremely Fast Large-Scale Proteomics Analysis."

[4] UC Riverside Data Science Seminar, Riverside, CA. November, 2019. "Accelerated Machine Learning for Computational Proteomics."

[5] UC Irvine AI/ML Seminar, Irvine, CA. November, 2019. "Accelerated Machine Learning for Computational Proteomics."

[6] UC Davis Postdoctoral Research Symposium, Davis, CA. **Oral presentation**. April, 2019. "Faster Percolator Analysis via Efficient SVM Learning for Faster Large-Scale Protein Analysis."

[7] NeurIPS, Montreal, Canada. December, 2018. "Learning Concave Conditional Likelihood Models for Improved Analysis of Tandem Mass Spectra."

[8] NIPS, Long Beach, CA. **Spotlight oral presentation**. December, 2017. "Gradients of Generative Models for Improved Discriminative Analysis of Tandem Mass Spectra."

[9] Advanced Methodologies for Bayesian Networks (AMBN), Kyoto, Japan. September, 2017. **Invited keynote speaker**. "Analyzing Tandem Mass Spectra: A Graphical Models Perspective."

[10] NIPS Workshop on Machine Learning in Computational Biology (MLCB), Barcelona, Spain. December 2016. "Fisher Kernels for Improved Analysis of Tandem Mass Spectra."

[11] Department of Biostatistics, UC Davis. May, 2016. "Faster and More Accurate Graphical Models for Peptide Identification of Tandem Mass Spectra."

[12] Bioinstrumentation and BioMEMS Laboratory, UC Davis. April, 2016. "Faster and More Accurate Graphical Models for Peptide Identification of Tandem Mass Spectra."

[13] MODE Laboratory, University of Washington. May, 2015. "Graphical Models for Peptide Identification of Tandem Mass Spectra."

[14] NIPS Workshop on Machine Learning in Computational Biology (MLCB), Lake Tahoe, CA. December, 2012. "Peptide Identification of Tandem Mass Spectra via Spectrum Alignment using a Dynamic Bayesian Network."

[15] "Peptide Identification in Tandem Mass Spectrometry using Dynamic Bayesian Networks." Department of Electrical and Computer Engineering, Seattle University. April, 2012.

Software

All of the software listed below are available with source code at the URLs specified.

[1] **JT Halloran**, G Urban, DM Rocke, and P Baldi. ProteoTorch is a light and flexible Python package for deep semi-supervised analysis of shotgun proteomics data, with an emphasis on speed. By default, deep neural network classifiers are used to accurately recalibrate PSMs and features collected from a database-search. A host of other classifiers are also available, including ultrafast support vector machine (SVM) implementations. https://github.com/proteoTorch/proteoTorch. 2020.

[2] **JT Halloran**. GPU-optimized LIBLINEAR solvers: The following contains highly optimized GPU-training primal algorithms for logistic regression (in LIBLINEAR for sparse features) and SVMs (in Percolator for dense features). The underlying CPU-centric algorithms considered natively resist GPU-optimizations, due to the heavy sequential dependence of variables in the underlying algorithm (i.e., the trust-region second-order algorithm, TRON) and reliance on random access. Thus, the code enables GPU-optimizations using the techniques described in our NeurIPS 2020 paper. https://jthalloran.bitbucket.io/code/tronGpu_neurips2020.html. 2020.

[3] **JT Halloran**, M The, WS Noble, and L Käll. PERCOLATOR is a popular post-processing for enriching the results of a shotgun proteomics database search. I've rewritten the SVM training engine in version 3.5 , incorporating several of our recent works to accerate SVM training and overall analysis times. http://percolator.ms/. 2020.

[4] RK Iyer, **JT Halloran**, and K Wei. JENSEN is an easily-customizable C++ toolkit for massive-scale machine learning (ML) and convex optimization. Jensen natively supports a large number of popular ML loss functions, state-of-the-art optimization algorithms, and ML applications. https://github.com/rishabhk108/jensen. 2018-Present.

[5] **JT Halloran**, JA Bilmes, David M. Rocke, and WS Noble. The DRIP TOOLKIT (DTK) is a tandem mass spectrometry search engine that uses a dynamic Bayesian network. DTK supports parameter estimation, multithreading, utilities for cluster use, instantiating/decoding/plotting DRIP PSMs in the python shell, and in-browser analysis of identified spectra. https://jthalloran.bitbucket.io/dripToolkit. 2016-Present.

[6] J Liu, **JT Halloran**, JA Bilmes, and WS Noble. THEMIS infers the genotypes and relative frequencies of cancer clones from deep DNA sequence of multiple biopsies from a single cancer using a dynamic Bayesian network. https://github.com/jieliu6/THEMIS. 2017.

[7] R Rogers, JA Bilmes, S Wang, and **JT Halloran**. The GRAPHICAL MODELS TOOLKIT (GMTK) is an open source, publicly available toolkit for rapidly prototyping statistical models using dynamic graphical models (DGMs) and dynamic Bayesian networks (DBNs). The discriminative training engine for general DBNs, based on the Maximum Mutual Information (MMI) criterion, was written by Shengjie Wang and I. https://melodi.ee.washington.edu/gmtk. 2016.

PROFESSIONAL ACTIVITIES

Reviewer for: ICLR, NeurIPS, ICML, BMC Bioinformatics, RECOMB, WABI, Journal of Bioinformatics and Computational Biology, BMC Genomics, IEEE Journal of Biomedical and Health Informatics
Program Committee for: AAAI 2019-2021, MLCB 2019-2020

SOFTWARE SKILLS

C/C++, Python, PyTorch, Keras, CUDA, OpenMP, Matlab, Git, Bash, LaTeX, HTML, Linux, Windows, Cygwin

REFERENCES

Available on request.