

# Stephen H. Bach

## Curriculum Vitae

sbach@cs.brown.edu  
cs.brown.edu/people/sbach

**Research** I investigate **machine learning** methods for incorporating high-level knowledge into statistical models. My approach is to develop new methods that blend statistical methods with rule-based systems to achieve benefits from each. Topics I have worked on include weak supervision, statistical relational learning, and information extraction.

**Positions** **Assistant Professor, 2018–Present**

**Brown University**  
Computer Science Department

**Visiting Scholar, 2018**

**Google**  
Mountain View, CA

**Postdoctoral Scholar, 2015–2018**

**Stanford University**  
Computer Science Department  
Advisor: Christopher Ré

**Education** **Ph.D., Computer Science, 2015**

**University of Maryland, College Park**

Advisor: Lise Getoor

Dissertation: Hinge-Loss Markov Random Fields and Probabilistic Soft Logic:  
A Scalable Approach to Structured Prediction

Committee: Rama Chellapa, Hal Daumé III, Larry Davis, Kevin Murphy

**Larry S. Davis Doctoral Dissertation Award**

**B.S., Computer Science and Mathematics (double major), 2010**

**Georgetown University**

Advisor: Mark Maloof

Magna Cum Laude

## Publications

### Pre-Prints

N. V. Nayak and S. H. Bach. Zero-shot learning with common sense knowledge graphs. arXiv:2006.10713 [cs.LG], 2020.

### Journal Papers

A. J. Ratner, S. H. Bach, H. E. Ehrenberg, J. Fries, S. Wu, and C. Ré. Snorkel: Rapid training data creation with weak supervision. *The VLDB Journal*, 29(2):709–730, 2020.

S. H. Bach, M. Broecheler, B. Huang, and L. Getoor. Hinge-loss Markov random fields and probabilistic soft logic. *Journal of Machine Learning Research (JMLR)*, 18(109):1–67, 2017.

G. Farnadi, S. H. Bach, M. Blondeel, M.-F. Moens, L. Getoor, and M. De Cock. Soft quantification in statistical relational learning. *Machine Learning*, 2017.

### Peer-Reviewed Conference Papers

E. Safranchik, S. Luo, and S. H. Bach. Weakly supervised sequence tagging from noisy rules. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2020.

S. H. Bach, D. Rodriguez, Y. Liu, C. Luo, H. Shao, C. Xia, S. Sen, A. Ratner, B. Hancock, H. Alborzi, R. Kuchhal, C. Ré, and R. Malkin. Snorkel DryBell: A case study in deploying weak supervision at industrial scale. In *International Conference on Management of Data (SIGMOD) Industry Track*, 2019.

- A. J. Ratner, S. H. Bach, H. E. Ehrenberg, J. Fries, S. Wu, and C. Ré. Snorkel: Rapid training data creation with weak supervision. *PVLDB*, 11(3):269–282, 2017. **Best of VLDB 2018.**
- S. H. Bach, B. He, A. J. Ratner, and C. Ré. Learning the structure of generative models without labeled data. In *International Conference on Machine Learning (ICML)*, 2017.
- H. Lakkaraju, S. H. Bach, and J. Leskovec. Interpretable decision sets: A joint framework for description and prediction. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2016.
- S. H. Bach\*, B. Huang\*, J. Boyd-Graber, and L. Getoor. Paired-dual learning for fast training of latent variable hinge-loss MRFs. In *International Conference on Machine Learning (ICML)*, 2015.
- S. H. Bach, B. Huang, and L. Getoor. Unifying local consistency and MAX SAT relaxations for scalable inference with rounding guarantees. In *Artificial Intelligence and Statistics (AISTATS)*, 2015. **Selected for oral presentation, 6% of submitted papers (27/442).**
- G. Farnadi, S. H. Bach, M. Blondeel, M.-F. Moens, L. Getoor, and M. De Cock. Statistical relational learning with soft quantifiers. In *International Conference on Inductive Logic Programming (ILP)*, 2015. **Best Student Paper Award.**
- S. H. Bach, B. Huang, B. London, and L. Getoor. Hinge-loss Markov random fields: Convex inference for structured prediction. In *Uncertainty in Artificial Intelligence (UAI)*, 2013.
- S. H. Bach, M. Broecheler, L. Getoor, and D. P. O’Leary. Scaling MPE inference for constrained continuous Markov random fields with consensus optimization. In *Advances in Neural Information Processing Systems (NIPS)*, 2012.
- S. H. Bach and M. A. Maloof. A Bayesian approach to concept drift. In *Advances in Neural Information Processing Systems (NIPS)*, 2010.
- S. H. Bach\* and M. A. Maloof\*. Paired learners for concept drift. In *IEEE International Conference on Data Mining (ICDM)*, 2008.

### Demonstrations

- A. J. Ratner, S. H. Bach, H. E. Ehrenberg, and C. Ré. Snorkel: Fast training set generation for information extraction. *ACM SIGMOD Conference on Management of Data (SIGMOD)*, 2017.

### Workshop Papers

- E. Raisi and S. H. Bach. Selecting auxiliary data using knowledge graphs for image classification with limited labels. In *CVPR Workshop on Visual Learning with Limited Labels*, 2020.
- R. Patel, S. H. Bach, and E. Pavlick. Learning visually grounded representations with sketches. In *ICML Workshop on New Tasks for Vision and Language*, 2019.
- S. H. Bach, B. Huang, and L. Getoor. Rounding guarantees for message-passing MAP inference with logical dependencies. In *NIPS Workshop on Discrete and Combinatorial Problems in Machine Learning (DISCML)*, 2014.
- S. H. Bach, B. Huang, and L. Getoor. Probabilistic soft logic for social good. In *KDD Workshop on Data Science for Social Good*, 2014.
- G. Farnadi, S. H. Bach, M. Moens, L. Getoor, and M. De Cock. Extending PSL with fuzzy quantifiers. In *International Workshop on Statistical Relational Artificial Intelligence (StaRAI)*, 2014.
- S. H. Bach, B. Huang, and L. Getoor. Large-margin structured learning for link ranking. In *NIPS Workshop on Frontiers of Network Analysis: Methods, Models, and Applications*, 2013. **Best Student Paper Award.**
- S. H. Bach, B. Huang, and L. Getoor. Learning latent groups with hinge-loss Markov random fields. In *ICML Workshop on Inferning: Interactions between Inference and Learning*, 2013.
- B. London, S. Khamis, S. H. Bach, B. Huang, L. Getoor, and L. Davis. Collective activity detection using hinge-loss Markov random fields. In *CVPR Workshop on Structured Prediction: Tractability, Learning and Inference*, 2013.

- A. Kimmig, S. H. Bach, M. Broecheler, B. Huang, and L. Getoor. A short introduction to probabilistic soft logic. In *NIPS Workshop on Probabilistic Programming: Foundations and Applications*, 2012.
- B. Huang, S. H. Bach, E. Norris, J. Pujara, and L. Getoor. Social group modeling with probabilistic soft logic. In *NIPS Workshop on Social Network and Social Media Analysis: Methods, Models, and Applications*, 2012.
- A. Memory, A. Kimmig, S. H. Bach, L. Raschid, and L. Getoor. Graph summarization in annotated data using probabilistic soft logic. In *Proceedings of the International Workshop on Uncertainty Reasoning for the Semantic Web (URSW)*, 2012.
- S. H. Bach, M. Broecheler, S. Kok, and L. Getoor. Decision-driven models with probabilistic soft logic. In *NIPS Workshop on Predictive Models in Personalized Medicine*, 2010.

\* Equal Contributors

## Invited Talks

- Weakly Supervised Machine Learning at Industrial Scale*  
 ACM/IEEE Computer Society Boston Chapter Joint Seminar Series Jun. 20 2019  
 University of Massachusetts, Lowell, Computer Science Department Oct. 16 2019
- Programming Statistical Machine Learning with High-Level Knowledge*  
 University of Chicago, Computer Science Department Apr. 11 2018  
 Harvard University, Computer Science Department Apr. 5 2018  
 Northeastern University, College of Computer and Info. Science Apr. 3 2018  
 University of Pennsylvania, Computer Science Department Mar. 28 2018  
 Rutgers University, Computer Science Department Mar. 26 2018  
 Brown University, Computer Science Department Mar. 19 2018  
 University of Massachusetts Amherst, College of Info. and Computer Sciences Feb. 26 2018  
 Duke University, Computer Science Department Feb. 21 2018  
 Dartmouth College, Computer Science Department Feb. 16 2018  
 University of California, Irvine, Computer Science Department Feb. 13 2018  
 Purdue University, Computer Science Department Feb. 8 2018
- Snorkel: Creating Noisy Training Data to Overcome Machine Learning's Biggest Bottleneck*  
 AAAI Symposium: Towards AI for Collaborative Open Science Mar. 26 2019  
 University of California, Berkeley, Berkeley AI Research Lab Jul. 10 2017  
 SIGMOD Workshop on Data Management for End-to-End Machine Learning (DEEM) May 14 2017
- Probabilistic Soft Logic: Scaling Up Logical Reasoning in Statistical Machine Learning*  
 CSLI Workshop on Logic, Rationality, and Intelligent Interaction Jun. 2 2018  
 University of California, Santa Cruz, Computer Science Department Mar. 4 2016  
 Stanford University, InfoLab, Computer Science Department Feb. 11 2015  
 University of California, San Diego, San Diego Supercomputer Center Feb. 9 2015  
 Charles River Analytics, Cambridge, MA Aug. 20 2014

## Invited Panels

- Learning in the Presence of Label Scarcity*. NorthEast Computational Health Summit Apr. 26 2019

## Honors and Awards

- Larry S. Davis Doctoral Dissertation Award**  
 2015, Computer Science Department, University of Maryland, College Park
- John D. Gannon Award for Outstanding Graduate Student**  
 2013, Computer Science Department, University of Maryland, College Park
- Computer Science Award**  
 2010, Georgetown University

## Teaching

### Machine Learning

CSCI 1420, Brown University  
Spring 2019, 2020

### Learning with Limited Labeled Data

CSCI 2952-C, Brown University  
Fall 2018

## Professional Activities

### Workshop Organizer

2019 ICLR Workshop on Learning with Limited Labeled Data: Weak Supervision and Beyond  
2017 NIPS Workshop on Learning with Limited Labeled Data: Weak Supervision and Beyond

### National Science Foundation Peer Review Panelist

Division of Information and Intelligent Systems (IIS)

### Conference Area Chair / Senior Program Committee / Meta-Reviewer

Artificial Intelligence and Statistics (AISTATS)  
International Conference on Learning Representations (ICLR)  
International Conference on Machine Learning (ICML)  
International Joint Conference on Artificial Intelligence (IJCAI)  
Neural Information Processing Systems (NeurIPS)

### Conference Program Committee / Reviewer

International Conference on Learning Representations (ICLR)  
International Conference on Machine Learning (ICML)  
International Joint Conference on Artificial Intelligence (IJCAI)  
International World Wide Web Conference (WWW)  
Neural Information Processing Systems (NeurIPS)  
SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)  
Uncertainty in Artificial Intelligence (UAI)

### Journal Reviewer

ACM Transactions on Knowledge Discovery in Data (TKDD)  
Data Mining and Knowledge Discovery  
IEEE Transactions on Knowledge and Data Engineering (TKDE)  
IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)  
Journal of Machine Learning Research (JMLR)  
Science Advances  
Statistical Analysis and Data Mining

### Workshop Program Committee

Automatic Knowledge Base Construction (AKBC)  
Data Driven Discovery of Models (D3M)  
Declarative Learning Based Programming (DeLBP)  
Knowledge Base Construction, Reasoning and Mining (KBCOM)  
Mining and Learning with Graphs (MLG)

August 7, 2020