

Sameer Singh

Research Interests

- Large-Scale Machine Learning
- Information Extraction/NLP
- Probabilistic Programming
- Interactive Machine Learning
- Distributed and Parallel Inference
- Semi-Supervised Learning

Education

- 2008–2014 **PhD (Computer Science)**, *University of Massachusetts*, Amherst, MA.
title Scaling MCMC Inference and Belief Propagation to Large, Dense Graphical Models
committee Andrew McCallum (chair), Carlos Guestrin, Ben Marlin, David Jensen, Michael Zink
- 2004–2007 **MS (Computer Science)**, *Vanderbilt University*, Nashville, TN.
- 2000–2004 **BE (Electronics & Comm Engg)**, *NSIT (Delhi University)*, New Delhi, India.

Experience

Academic

- Sept 2013–present **Postdoc (Research Associate)**, *Computer Science*, University of Washington, Seattle.
Working with Carlos Guestrin, Luke Zettlemoyer, and Dan Weld on large-scale, interactive machine learning for extracting structured data from text corpora.
- 2008–2013 **Research Assistant**, *Information Extraction and Synthesis Lab (IESL)*, UMass, Amherst.
Worked on various large-scale applications of machine learning in natural language processing
- May 05–May 07 **Research Assistant**, *Inst. for Software & Integrated Systems (ISIS)*, Vanderbilt University.
- Summer 2006 **Research Assistant**, *Department of Biomedical Informatics*, Vanderbilt University.
- 2003–2004 **Student Collaborator**, *Defense Research & Development Organisation (DRDO)*, India.
- 2002–2004 **Student Collaborator**, *School of Physical Sciences*, Jawaharlal Nehru Univ. (JNU), India.

Industry

- June–Sept 2012 **Research Intern**, *Microsoft Research*, Cambridge, UK.
Generative probabilistic modeling of relational databases
- June–Sept 2010 **Research Intern**, *Google Research*, Mountain View, CA.
Large-scale coreference using Map-Reduce based distributed inference & a novel hierarchical model
- July–Sept 2009 **Applied Research Intern**, *Yahoo! Labs*, Santa Clara, CA.
Unsupervised extraction of entities from Search Ads using constraint-based CRF learning
- May–Dec 2007 **Software Engineer Intern**, *Google Inc*, Pittsburgh, PA.
Machine Learning and Feature Selection algorithms for Map-Reduce for the predicting Ad clicks
- May–Aug 2006 **Systems Engineer Intern**, *Adv Computing Center for Research and Education (ACCRES)*, Vanderbilt University, Nashville, TN.
Scheduling algorithms for queries to a Distributed File System
- Dec 03–Jan 2004 **Programming Intern**, *PRAVAK Cybernetics Ltd*, New Delhi, India.
- May–Aug 2003 **Robotics Intern**, *Central Electronic Engineering Research Inst (CEERI)*, New Delhi, India.
Prosthetic hand that enables gripping of irregular, slippery and delicate objects

Teaching

- Fall 2015 **Guest Lecturer**, CSEP 517: *Natural Language Processing*, instr: Yejin Choi, UW.
- Fall 2015 **Guest Lecturer**, CSE 546: *Machine Learning*, instr: Sham Kakade, UW.
- Fall 2014 **Guest Lecturer**, CSE 546: *Machine Learning*, instr: Carlos Guestrin, UW.
- Spring 2012 **Guest Lecturer**, CS691: *Probabilistic Graphical Models*, instr: Ben Marlin, UMass.
- Spring 2011 **Teaching Assistant**, CS691: *Probabilistic Graphical Models*, instr: Andrew McCallum, UMass.
- Spring 2008 **Teaching Assistant**, CS121: *Introduction to Programming*, instr: Erik Learned-Miller, UMass.
- Fall 04, Spring 05 **Teaching Assistant**, CS201: *Data Structures*, instr: Doug Fisher, Vanderbilt University.

Committees

- 2015 **PhD Committee**, Xiao Ling, Advised by Daniel S. Weld, UW.

Outreach

- mentor High School and Middle School teams, Paul Allen Computing Challenge, 2014
- judge Science Fair, Hampden Charter School of Science, 2012
- session chair Massachusetts Statewide Undergraduate Research Conference, 2012

Other Service

- member UW CSE Postdoc Best Practices Committee, 2015
- member UW Graduate Admissions Committee, 2013
- member UMass Faculty Hiring Committee (CS + Computational Biology), 2012
- steward Computer Science, Graduate Employee Organization (GEO), UMass, 2010, 2011
- member New Students Committee, UMass Computer Science Department, 2009, 2010, 2012
- general secretary Vanderbilt India Association (VIA), 2005-2006

Awards and Honors

- 2015 **DARPA Riser**, DARPA Wait, What? Event, St. Louis, MO.
- 2015 **Grand Prize Winner**, Yelp Dataset Challenge, Round 4.
- 2015 **Support Grant for Junior Researchers**, Schloss Dagstuhl-NSF, Dagstuhl, Germany.
- 2014, 2015 **UW CSE Postdoc Research Award**, University of Washington, Seattle, WA.
- 2014 **Amazon Machine Learning Data Grant**, Amazon, Seattle, WA.
- 2014 **Exceptional Submission Award**, ACL Workshop on Semantic Parsing, Baltimore, MD.
- 2012 **Finalist**, Facebook PhD Fellowship, Menlo Park, CA.
- 2011 **Travel Award**, Neural Information Processing Systems (NIPS), Granda, Spain.
- 2011 **Best Talk Award**, Machine Reading Project Phase 3 Kickoff, Seattle, WA.
- 2010-2011 **Yahoo! Key Scientific Challenges (KSC) Award**, in Machine Learning & Statistics.
- 2010-2011 **Accomplishments in Search & Mining Award**, UMass CS Dept and Yahoo!.
- 2009-2010 **Graduate School Fellowship Award**, *for continuing students*, Univ of Massachusetts.
- 2009 **Travel Award**, Neural Information Processing Systems (NIPS), Vancouver, Canada.
- 2004 **Best Undergraduate Project**, Department of Instrumentation & Control (NSIT).

Professional Activities

Conference and Workshop Organization

- organizer NAACL Workshop on "Automated Knowledge Base Construction", 2016
- organizer AAAI Workshop on "Declarative Learning Based Programming", 2016
- organizer NIPS Workshop on "Machine Learning Systems (LearningSys)", 2015

- organizer** NIPS Workshop on “Automated Knowledge Base Construction”, 2014
organizer CIKM Workshop on “Automated Knowledge Base Construction”, 2013
organizer NIPS Workshop on “Big Learning”, 2011, 2012, 2013
organizer ICML Workshop on “Inferning”: Interactions between Inference and Learning, 2012, 2013
co-chair North-East Student Colloquium on Artificial Intelligence (NESCAI), 2010

Professional Service

- reviewer** AKBC 2010, NESCAI 2010, EMNLP 2010, KDD 2011, IJCAI 2011, EMNLP 2011, NIPS 2011, EACL 2012, OPTMAS 2012, AKBC 2012, NIPS 2012, ICML 2013, NAACL 2013, UAI 2013, NIPS 2013, ACL 2014, EACL 2014, ICML 2014, COLING 2014, ACL 2014, EMNLP 2014, CIKM 2014, NAACL 2015, ACL 2015, ICML 2015, EMNLP 2015, NIPS 2015
- pc member** NESCAI 2010, IJCAI 2011, EMNLP 2011, EACL 2012, OPTMAS 2012, AKBC 2012, EMNLP 2012, OPTMAS 2013, CONLL 2013, LML 2013, AKBC 2013, StarAI 2014, CIKM 2014, LTPM 2014, NIPS 2014, AAAI 2015, IJCAI 2015, KDD 2015, WSDM 2016, ACL 2016, NAACL 2016, IJCAI 2016
- standing reviewer** Transactions of the ACL (2015-2016), Computational Linguistics (2016), ACM Transactions on Intelligent Systems and Technology (2016)

Invited Talks

- Jan 27 2016 **Interactive Training of Relation Embeddings**, *University of Maryland*, College Park, MD.
Jan 26 2016 **Interactive Training of Relation Embeddings**, *Johns Hopkins University*, Baltimore, MD.
Aug 18 2015 **Wolfe: A Declarative Machine Learning Stack**, *Big Data Scala By the Bay*, San Francisco, CA.
April 30 2015 **Declarative ML With Wolfe**, *Dagstuhl Seminar on Probabilistic Programming*, Dagstuhl, Germany.
Jan 9 2015 **Interactive Matrix Factorization**, *UW CSE Department Summit*, Seattle, WA.
Dec 11 2014 **Wolfe: Declarative Machine Learning**, *Montreal Scala Meetup*, Montreal, Canada.
Aug 23 2014 **Large-Scale Entity Resolution**, *COLING AHA! Workshop*, Dublin, Ireland.
Nov 18 2013 **Universal Schema for TACKBP**, *NIST*, Gaithersburg, MD.
Feb 8 2013 **Large-Scale Entity Resolution**, *Univ of Pennsylvania*, Philadelphia, PA.
Nov 2 2012 **Machines That Read**, *Computer Science Dept Colloquium*, Williams College, MA.
Apr 6 2011 **Large-scale Cross-doc Coreference**, *Machine Reading Project Ph 3 Kickoff*, Seattle, WA.
Nov 5 2008 **Deep Belief Nets**, *Machine Learning and Friends Lunch (MLFL)*, UMass, Amherst.

Posters

- Sept 9 2015 **Interactive Machine Learning for NLP**, *DARPA Wait, What Risers*, St Louis, MO.
Jul 9 2012 **Parallel & Distributed Graphical Models**, *Graphlab Workshop*, San Francisco, CA.
May 16 2012 **Scaling Up MCMC**, *New England Machine Learning Day*, Boston, MA.

Publications

Theses and Book Chapters

- PhD Thesis** · **Sameer Singh**. *Scaling MCMC Inference and Belief Propagation for Large, Dense Graphical Models*. PhD thesis, University of Massachusetts, 2014
- Book Chapter** · Jeremy Kubica, **Sameer Singh**, and Daria Sorokina. Parallel large-scale feature selection, 2011

Journal Publications

- Xiao Ling, **Sameer Singh**, and Dan Weld. Design challenges for entity linking. *Transactions of the Association for Computational Linguistics (TACL)*, 3, 2015
- D.G. Feitelson, T.O.S. Adeshiyan, D. Balasubramanian, Y. Etsion, G. Madl, E.P. Osses, **Sameer Singh**, K. Suwanmongkol, M. Xie, and S.R. Schach. Fine-grain analysis of common coupling and its application to a linux case study. *Journal of Systems and Software (JSS)*, 80, 2007

- S.R. Schach, T.O.S. Adeshiyan, D. Balasubramanian, G. Madl, E.P. Osses, **Sameer Singh**, K. Suwanmongkol, M. Xie, and D.G. Feitelson. Common coupling and pointer variables, with application to a linux case study. *Software Quality Journal (SQJ)*, 15, 2007

Peer-Reviewed Conference Papers

- Tim Rocktaschel, **Sameer Singh**, and Sebastian Riedel. Injecting logical background knowledge into embeddings for relation extraction. In *Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2015
- Tianqi Chen, **Sameer Singh**, Ben Taskar, and Carlos Guestrin. Efficient second-order gradient boosting for conditional random fields. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2015
- **Sameer Singh**, Tim Rocktaschel, Luke Hewitt, Jason Naradowsky, and Sebastian Riedel. WOLFE: An NLP-friendly declarative machine learning stack. In *Demo at the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2015a
- Ignacio Cano, **Sameer Singh**, and Carlos Guestrin. Distributed non-parametric representations for vital filtering: UW at TREC KBA 2014. In *Text REtrieval Conference (TREC): Knowledge-Base Acceleration (KBA) Track*, 2014
- Fabian M. Suchanek, **Sameer Singh**, Sebastian Riedel, and Partha P. Talukdar. Akbc 2013: Third workshop on automated knowledge base construction. In *ACM Conference of Information and Knowledge Management (CIKM)*, 2013
- **Sameer Singh** and Thore Graepel. Automated probabilistic modeling for relational data. In *ACM Conference of Information and Knowledge Management (CIKM)*, 2013
- Jiaping Zheng, Luke Vilnis, **Sameer Singh**, Jinho D. Choi, and Andrew McCallum. Dynamic knowledge-base alignment for coreference resolution. In *Conference on Computational Natural Language Learning (CoNLL)*, 2013
- **Sameer Singh**, Limin Yao, David Belanger, Ari Kobren, Sam Anzaroot, Michael Wick, Alexandre Passos, Harshal Pandya, Jinho D. Choi, Brian Martin, and Andrew McCallum. Universal schema for slot filling and cold start: Umass iesl at tackbp 2013. In *Text Analysis Conference on Knowledge Base Population (TAC KBP)*, 2013c
- Michael Wick, **Sameer Singh**, and Andrew McCallum. A discriminative hierarchical model for fast coreference at large scale. In *Association for Computational Linguistics (ACL)*, 2012
- **Sameer Singh**, Michael Wick, and Andrew McCallum. Monte carlo mcmc: Efficient inference by approximate sampling. In *Empirical Methods in Natural Language Processing (EMNLP)*, 2012d
- **Sameer Singh**, Amarnag Subramanya, Fernando Pereira, and Andrew McCallum. Large-scale cross-document coreference using distributed inference and hierarchical models. In *Association for Computational Linguistics (ACL)*, 2011b
- **Sameer Singh**, Dustin Hillard, and Chris Leggetter. Minimally-supervised extraction of entities from text advertisements. In *Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2010a
- **Sameer Singh**, Limin Yao, Sebastian Riedel, and Andrew McCallum. Constraint-driven rank-based learning for information extraction. In *Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2010d
- **Sameer Singh**, Karl Schultz, and Andrew McCallum. Bi-directional joint inference for entity resolution and segmentation using imperatively-defined factor graphs. In *Machine Learning and Knowledge Discovery in Databases (Lecture Notes in Computer Science) and European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*, 2009b
- **Sameer Singh**, Jeremy Kubica, Scott E. Larsen, and Daria Sorokina. Parallel large scale feature selection for logistic regression. In *SIAM International Conference on Data Mining (SDM)*, 2009a
- Andrew McCallum, Karl Schultz, and **Sameer Singh**. Factorie: Probabilistic programming via imperatively defined factor graphs. In *Neural Information Processing Systems (NIPS)*, 2009

- Michael Wick, Khashyar Rohanimanesh, **Sameer Singh**, and Andrew McCallum. Training factor graphs with reinforcement learning for efficient map inference. In *Neural Information Processing Systems (NIPS)*, 2009
- **Sameer Singh** and Julie A. Adams. Transfer of learning for complex domains: A demonstration using multiple robots. In *International Conference on Robotics and Automation (ICRA)*, 2006
- **Sameer Singh**. Finding the shortest path for a mobile robot in an unmapped maze from minimum runs. In *Int Conf on CAD, CAM, Robotics and Autonomous Factories (INCARF)*, 2003

Peer-reviewed Workshops and Symposia

- **Sameer Singh** and Sebastian Riedel. Creating interactive and visual educational resources for AI. In *AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI)*, 2016
- **Sameer Singh**, Tim Rocktaschel, and Sebastian Riedel. Towards combined matrix and tensor factorization for universal schema relation extraction. In *NAACL Workshop on Vector Space Modeling for NLP*, 2015b
- Ivan Sanchez, Tim Rocktaschel, Sebastian Riedel, and **Sameer Singh**. Towards extracting faithful and descriptive representations of latent variable models. In *AAAI Spring Symposium on Knowledge Representation and Reasoning (KRR): Integrating Symbolic and Neural Approaches*, 2015
- Guillaume Bouchard, **Sameer Singh**, and Theo Trouillon. On approximate reasoning capabilities of low-rank vector spaces. In *AAAI Spring Symposium on Knowledge Representation and Reasoning (KRR): Integrating Symbolic and Neural Approaches*, 2015
- Tim Rocktaschel, **Sameer Singh**, Matko Bosnjak, and Sebastian Riedel. Low-dimensional embeddings of logic. In *ACL 2014 Workshop on Semantic Parsing (SP14)*, 2014
- Sebastian Riedel, **Sameer Singh**, Vivek Srikumar, Tim Rocktaschel, Larysa Visengeriyeva, and Jan Noessner. Wolfe: Strength reduction and approximate programming for probabilistic programming. In *International Workshop on Statistical Relational AI (StarAI)*, 2014
- Xiao Ling, **Sameer Singh**, and Dan Weld. Context representation for named entity linking. In *Pacific Northwest Regional NLP Workshop (NW-NLP)*, 2014
- Mathias Niepert and **Sameer Singh**. Out of many, one: Unifying web-extracted knowledge bases. In *Workshop on Automated Knowledge Base Construction (AKBC)*, 2014
- **Sameer Singh**, Sebastian Riedel, Luke Hewitt, and Tim Rocktaschel. Designing an IDE for probabilistic programming: Challenges and a prototype. In *NIPS Workshop on Probabilistic Programming*, 2014b
- Victoria (Xi) Lin, **Sameer Singh**, Luheng He, Ben Taskar, and Luke Zettlemoyer. Multi-label learning with posterior regularization. In *NIPS Workshop on Modern Machine Learning and Natural Language Processing*, 2014
- **Sameer Singh**, Sebastian Riedel, Brian Martin, Jiaping Zheng, and Andrew McCallum. Joint inference of entities, relations, and coreference. In *CIKM Workshop on Automated Knowledge Base Construction (AKBC)*, 2013a
- Michael Wick, **Sameer Singh**, Ari Kobren, and Andrew McCallum. Assessing confidence of knowledge base content with an experimental study in entity resolution. In *CIKM Workshop on Automated Knowledge Base Construction (AKBC)*, 2013a
- Michael Wick, **Sameer Singh**, Harshal Pandya, and Andrew McCallum. A joint model for discovering and linking entities. In *CIKM Workshop on Automated Knowledge Base Construction (AKBC)*, 2013b
- **Sameer Singh**, Sebastian Riedel, and Andrew McCallum. Anytime belief propagation using sparse domains. In *Neural Information Processing Systems (NIPS) Workshop on Resource Efficient Machine Learning*, 2013b
- **Sameer Singh**, Michael Wick, and Andrew McCallum. Monte carlo mcmc: Efficient inference by sampling factors. In *NAACL/HLT Workshop on Automated Knowledge Base Construction (AKBC-WEKEX)*, 2012c
- **Sameer Singh** and Thore Graepel. Compiling relational database schemata into probabilistic graphical models. In *NIPS Workshop on Probabilistic Programming*, 2012

- **Sameer Singh** and Andrew McCallum. Towards asynchronous distributed mcmc inference for large graphical models. In *Neural Information Processing Systems (NIPS) Workshop on Big Learning*, 2011
- **Sameer Singh**, Brian Martin, and Andrew McCallum. Inducing value sparsity for parallel inference in tree-shaped models. In *Neural Information Processing Systems (NIPS) Workshop on Computational Trade-offs in Statistical Learning (COST)*, 2011a
- **Sameer Singh**, Amarnag Subramanya, Fernando Pereira, and Andrew McCallum. Distributed map inference for undirected graphical models. In *Neural Information Processing Systems (NIPS) Workshop on Learning on Cores, Clusters, and Clouds (LCCC)*, 2010b
- Andrew McCallum, Khashyar Rohanimanesh, Michael Wick, Karl Schultz, and **Sameer Singh**. Factorie: Efficient probabilistic programming via imperative declarations of structure, inference and learning. In *NIPS Workshop on Probabilistic Programming*, 2008
- T. Kichkaylo, C. van Buskirk, **Sameer Singh**, H. Neema, M. Orosz, and R. Neches. Mixed-initiative planning for space exploration missions. In *International Conference on Automated Planning and Scheduling Workshop (ICAPS)*, 2007

Unrefereed Publications

- Nitish Gupta and **Sameer Singh**. Collective factorization for relational data: An evaluation on the yelp datasets. Technical report, Yelp Dataset Challenge, Round 4, 2015
- **Sameer Singh**, Amarnag Subramanya, Fernando Pereira, and Andrew McCallum. Wikilinks: A large-scale cross-document coreference corpus labeled via links to wikipedia. Technical report, University of Massachusetts Amherst, CMPSCI UM-CS-2012-015, 2012b
- **Sameer Singh**, Gregory Druck, and Andrew McCallum. Constraint-driven training of complex models using mcmc. Technical report, University of Massachusetts Amherst, CMPSCI UM-CS-2012-032, 2012a
- **Sameer Singh**, Michael Wick, and Andrew McCallum. Distantly labeling data for large scale cross-document coreference. Technical report, Computing Research Repository (CoRR) eprint arXiv:1005.4298, 2010c
- **Sameer Singh**. Option discovery in hierarchical reinforcement learning for training large factor graphs for information extraction, 2009
- Khashyar Rohanimanesh, Michael Wick, **Sameer Singh**, and Andrew McCallum. Reinforcement learning for map inference in large factor graphs. Technical report, University of Massachusetts Amherst, CMPSCI UM-CS-2008-040, 2008

Patents

- **Sameer Singh**, Thore Graepel, Lucas J. Bordeaux, and Andrew D. Gordon. Relational database management. Technical report, US Patent Number 0188928, 2014a
- **Sameer Singh**, E. S. Larsen, Jeremy Kubica, and Andrew W. Moore. Feature selection for large scale models. Technical report, US Patent Number 8190537, 2008

Open Source Software

- previous maintainer** **Factorie**, *Design, training and inference for probabilistic graphical models.*
factorie.cs.umass.edu
- contributor** **Wolfe**, *Language for declarative machine learning.*
wolfe.ml
- owner** **Moro**, *Interactive documentation and educational resources.*
github.com/wolfe-pack/moro

References

Available upon request