

Rajhans Samdani

Phone: (217) 377 7023

E-mail: rajhans@gmail.com

Work Experience

Google Inc. September 2013 - Present
Senior Research Scientist
Working on machine learning (including deep learning) techniques for natural language processing.

Education

PhD, Computer Science August 2009 - August 2013
University of Illinois at Urbana-Champaign (GPA 4.00) Advisor: Prof. Dan Roth
Thesis title: Algorithms for Structural Learning with Decompositions

Master of Science, Computer Science August 2007 - August 2009
University of Illinois at Urbana-Champaign (GPA 4.00) Advisor: Prof. Dan Roth
Thesis title: Learning Multi-linear Representations for Efficient Inference

Bachelor of Technology, Computer Science July 2003 - August 2007
Indian Institute of Technology Bombay (GPA 9.41/10.00) Advisor: Prof. Soumen Chakrabati
Thesis title: Entity Annotation and Reconciliation in Graphs Using
Local and Global Dependencies

Publications

- ◇ **R. Samdani** and K.-W. Chang and D. Roth, A Discriminative Latent Variable Model for Online Clustering. International Conference on Machine Learning (ICML), 2014
- ◇ **R. Samdani** and G. Mann, Restricted Transfer Learning for Text Categorization. Neural Information Processing Systems (NIPS) workshop on New Directions in Transfer and Multi-Task: Learning Across Domains and Tasks, 2013
- ◇ K.-W. Chang and **R. Samdani** and D. Roth, A Constrained Latent Variable Model for Coreference Resolution. Empirical Methods in Natural Language Processing (EMNLP), 2013
- ◇ **R. Samdani** and D. Roth, Efficient Decomposed Learning for Structured Prediction. International Conference on Machine Learning (ICML), 2012
- ◇ **R. Samdani** and M. Chang and D. Roth, A Framework for Tuning Posterior Entropy. International Conference in Machine Learning (ICML) workshop on Inferring: Interactions between Inference and Learning, 2012
- ◇ K.-W. Chang and **R. Samdani** and A. Rozovskaya and M. Sammons and D. Roth, Illinois-Coref: The UI System in the CoNLL-2012 Shared Task. Conference on Computational Natural Language Learning (CoNLL): Shared Task, 2012
- ◇ **R. Samdani** and M. Chang and D. Roth, Unified Expectation Maximization. North American Chapter of the Association for Computational Linguistics (NAACL), 2012
- ◇ **R. Samdani** and W. Yih. Domain Adaptation with Ensemble of Feature Groups. International Joint Conference on Artificial Intelligence (IJCAI), 2011
- ◇ G. Kundu and **R. Samdani** and D. Roth, Constrained Conditional Models for Information Fusion. International Conference on Information Fusion (Fusion), 2011
- ◇ K.-W. Chang and **R. Samdani** and A. Rozovskaya and N. Rizzolo and M. Sammons and D. Roth, Inference Protocols for Coreference Resolution. Conference on Computational Natural Language Learning (CoNLL): Shared Task, 2011
- ◇ G. Levine and G. DeJong and L. Wang and **R. Samdani** and S. Vembu and D. Roth, Automatic Model Adaptation for Complex Structured Domains. European Conference in Machine Learning (ECML), 2010

- ◇ D. Roth and **R. Samdani**. Learning Multi-linear Representations for Efficient Inference. Machine Learning Journal, 2009. **Amongst 14 out of 600 papers accepted for the Machine Learning Journal version of European Conference on Machine Learning (ECML), 2009**

Industry Research Internships

Google Research Internship

Restricted Task-transfer for Text Categorization

May - August 2012

Manager: Dr. Gideon Mann

- ◇ Restricted Transfer Learning: transferring information from one classification task to another classification task without storing the labeled data of the former due to concerns of privacy, memory, scalability, etc.
- ◇ Developed a restricted transfer learning algorithm for text categorization which achieves 2-6% higher accuracy.
- ◇ Implemented the algorithm, which is highly parallelizable, using Map/Reduce to efficiently create a word cluster feature space over 500k words.

Microsoft Research Internship

Domain Adaptation with Ensemble of Feature Groups

May - August 2010

Supervisor: Dr. Scott Wen-Tau Yih

- ◇ Domain Adaptation: adapting to a prediction domain with a different distribution than the training domain.
- ◇ Developed an algorithm for domain adaptation and supported it with theoretical justifications.
- ◇ Conducted experiments on sentiment and spam detection, obtaining 1-10% higher accuracy than baselines.

Yahoo! Labs Internship

Query-based Summarization of Search Results

May - August 2008

Supervisor: Dr. Tapas Kanungo

- ◇ Developed a system for query-based document summarization primarily for presenting web search results.
- ◇ Experimented with an Integer Linear Programming based prototype for producing a readable summary which is query relevant as well as spans the document description.
- ◇ Patent on stopword-based Clause Identification pending.

Research Theses

PhD Thesis

Algorithms for Structural Learning with Decompositions

- ◇ Proposed supervised, unsupervised, and latent variable-based machine learning algorithms for structured output settings commonly occurring in natural language processing.
- ◇ Developed efficient learning algorithms that decompose a larger structured output space into smaller and tractable spaces.

MS Research Thesis

Learning Multi-linear Representations for Efficient Inference

- ◇ Proposed a new representation for probability distributions which provides faster inference than the commonly used graphical models approach.
- ◇ Developed a learning technique based on convex optimization and provided a closed-form approximation.
- ◇ Demonstrated faster and more accurate inference than the popular Bayesian networks on real world data.

Awards and Achievements

- ◇ Received the State Farm Fellowship for the year 2012-13.
- ◇ Received the David J. Kuck Outstanding M.S. Thesis Award for the year 2009.
- ◇ Received the prestigious Siebel Scholarship for the year 2009.
- ◇ Secured All India Rank 5 amongst approximately 172, 000 appearing candidates in the Joint Entrance Examination (JEE) for the IIT schools, which is one of the most competitive engineering exams in India.
- ◇ Secured All India Rank 2 in the All India Engineering Entrance Examination (AIEEE) amongst approximately 200, 000 students who took the exam.
- ◇ Amongst the top 25 gold medalists in the Indian National Physics Olympiad, 2003. Received special prize for the "Best Solution to the Most Tricky Problem".
- ◇ Received Presidents certificate for being listed as one of the top 1% students in the Higher Secondary Certificate Examination of Mathematics in India.

Patents

Hadar Shemtov and Tapas Kanungo and Rajhans Samdani and Donald Metzler, *Phrase Identification Using Break Points*. Application number 12334725.

Program Committee and Reviewing

International Conference on Machine Learning (ICML), 2010, 2011, 2013, 2014, 2015 (best reviewer award); Neural Information Processing Systems (NIPS) 2015; North American Association for Computational Linguistics (NAACL) 2015; Association for Computational Linguistics (ACL) 2013; American Association for Artificial Intelligence (AAAI) 2015; Conference on Computational Natural Language Learning (CoNLL) 2013, 2014, 2015

Programming Skills

- ◇ **Fluent:** C++, Java, Python,
- ◇ **Working knowledge:** C, Perl, Awk, FORTRAN, Prolog, Javascript, SQL, MySQL, JDBC, Matlab, Bash, Scheme