

Meghana Kshirsagar

Memorial Sloan Kettering Cancer Center – 1275 York Avenue, NY 10021

✉ meghana.ksagar@gmail.com • 🌐 www.meghanak.net

Education

Carnegie Mellon University, School of Computer Science 2010–2015

PhD from Language Technologies Institute (LTI), GPA: 3.98

Advisors: Jaime Carbonell and Judith Klein-Seetharaman

Thesis: *Combine and conquer: Methods for Multitask Learning in Biology and Language*

Indian Institute of Technology, Bombay 2004–2007

Master of Technology, Computer Science Dept., CPI: 9.3/10

Advisor: S. Sudarshan

Thesis: *Graph Algorithms for Keyword Search on External Memory Data Graphs*

Vasavi Engg. College, Osmania University, Hyderabad 2000–2004

Bachelor of Engineering, Computer Science Dept., CPI: 8.45/10

Work & Research Experience

Memorial Sloan Kettering Cancer Center, NY Jun 2016–

Research Scholar

Building machine learning models to combine knowledge from diverse, large scale, next-gen sequencing data arising in epigenetic studies (part of ENCODE consortium).

IBM T.J Watson Research, Yorktown Heights Sept 2015–Apr 2016

Postdoctoral researcher, Machine Learning

Multimodal data analysis involving time-series data from sources such as genomic information, hyperspectral images, ground measurements in the context of the TERRA project (ARPA-e funded grant)

Yahoo! Labs, Bangalore 2007–2009

Research Engineer, Search Relevance & Information Extraction

Applying and extending algorithms and machine learning techniques for large scale classification and information extraction from the Web such as web-page segmentation and classification

Microsoft Research, Redmond (Internship) summer 2012

Genome wide association studies for heritable diseases like Crohn's disease, with the goal of identifying causative genes by incorporating gene-gene dependencies from biological pathway networks

Industrial Scientific, Pittsburgh (Consulting) fall 2011

IBM Research, Delhi (Internship) summer 2005

Awards & Achievements

- Richard King Mellon Presidential Fellow of Life Sciences, Carnegie Mellon University, 2011-2014
- Ray Ozzie Fellowship awarded by Computer Science Dept at University of Illinois, Urbana Champaign, 2009
- Best Paper award at the Conference on Management of Data, 2010
- Best Poster prize at the CMU Student Research Symposium, 2013
- Won the Carnegie Mellon University Social Innovations Challenge, 2011
- Awards for topping the Computer Science Dept, 2001, 2002

- Selected for the meritorious Pratibha scholarship by the Govt. of Andhra Pradesh (India) for academic excellence in higher secondary education, 2000

Ongoing projects

- **Decoding epigenetic regulation:** inferring transcription factor binding from joint models trained on *in vivo* chromatin accessibility data and *in vitro* binding studies. We have developed hierarchical graphical models and pre-trained neural networks, to incorporate SELEX *in vitro* binding preferences into in-vivo models trained on ATAC-seq peaks. These models allow us to learn representations for transcription factors across several cell types and over similar members of a transcription factor family.
- **Identifying potential regulatory factors and maturation signatures in neuronal maturation** (collaboration with Lorenz Studer lab): The goal of this work is to combine time-course experimental data from chromatin accessibility (ATAC-seq) and gene expression (RNA-seq) studies of the neuronal maturation process in order to identify regulatory factors that control the transition of young neurons to mature neurons.

Publications

Meghana Kshirsagar, Han Yuan, Lee Zamparo, and Christina Leslie. Learning interpretable models of transcription factor binding from atac-seq data. *In preparation*.

Lee Zamparo, Han Yuan, Meghana Kshirsagar, and Christina Leslie. Better decoding of tf signals in accessible chromatin with learned embeddings and neural networks. *In preparation*.

Han Yuan, Meghana Kshirsagar, Lee Zamparo, Yuheng Lu, and Christina Leslie. Bindspace: decoding transcription factor binding signals by large-scale joint embedding. *Nature Methods (under review)*, 2018.

Meghana Kshirsagar, Eunho Yang, and Aurélie Lozano. Learning task structure via sparsity grouped multitask learning. *European Conference on Machine Learning (ECML 2017)*, 2017.

Meghana Kshirsagar, Jaime Carbonell, Judith Klein-Seetharaman, and Keerthiram Murugesan. Multitask matrix completion for learning protein interactions across diseases. *International Conference on Research in Computational Molecular Biology (RECOMB), Journal of Computational Biology (2017 issue)*, pages 53–64, 2016.

Sylvia Schleker, Meghana Kshirsagar, and Judith Klein-Seetharaman. Comparing human–salmonella with plant–salmonella protein–protein interaction predictions. *Frontiers in Microbiology*, 6(36), 2015.

Meghana Kshirsagar, Sam Thomson, Nathan Schneider, Jaime Carbonell, Noah A Smith, and Chris Dyer. Frame-semantic role labeling with heterogeneous annotations. In *Association for Computational Linguistics (ACL)*, 2015.

Meghana Kshirsagar, Sylvia Schleker, Jaime Carbonell, and Judith Klein-Seetharaman. Techniques for transferring host-pathogen protein interactions knowledge to new tasks. *Frontiers in Microbiology*, 6, 2015.

Meghana Kshirsagar, Jaime Carbonell, and Judith Klein-Seetharaman. Multitask learning for host–pathogen protein interactions. *Bioinformatics*, 29(13):i217–i226, 2013.

Meghana Kshirsagar, Jaime Carbonell, and Judith Klein-Seetharaman. Techniques to cope with missing data in host–pathogen protein interaction prediction. *Bioinformatics*, 28(18):i466–i472, 2012.

Zhongming Zhao, Junfeng Xia, Ozgur Tostan, Irtisha Singh, Meghana Kshirsagar, Jaime Carbonell, and Judith Klein-Seetharaman. Virus interactions with human signal transduction pathways. *International journal of computational biology and drug design*, 4(1):83–105, 2011.

Meghana Kshirsagar, Rajeev Rastogi, Sandeep Satpal, Sengamedu Srinivasan, and Venu Satuluri. High-precision web extraction using site knowledge. *Proceedings of the Conference on Management of Data (COMAD)*, 2010 (**Best Paper Award**).

Bhavana Bharat Dalvi*, Meghana Kshirsagar*, and S Sudarshan. Keyword search on external memory data graphs. *Proceedings of the Very Large Data Bases (VLDB)*, 1(1):1189–1204, 2008.

Workshop papers

- Inferring transcription factor binding profiles jointly from SELEX and ATAC-seq, *Cold Spring Harbor Labs (CSHL) workshop for Quantitative Biology, 2017*
- Iteratively Regrouped Lasso: learning group structures in genome wide studies of crops. M. Kshirsagar, E. Yang and A. C. Lozano, *Data Science for Food, Energy and Water at Conference on Knowledge Discovery and Data Mining (KDD) 2016*
- Automated Sorghum Phenotyping and Trait Development Platform. M. Tuiinstra, C. Weil, A. Thompson, C. Boomsma, M. Crawford, A. Habib, E. Delp, K. Cherkauer, M. Kshirsagar, E. Yang, P. Olsen, K. Natesan and A. C. Lozano, *Data Science for Food, Energy and Water at Conference on Knowledge Discovery and Data Mining (KDD) 2016*
- Leveraging Heterogeneous Data Sources for Relational Semantic Parsing. M. Kshirsagar, N. Schneider and C. Dyer, *Assoc. for Computational Linguistics (ACL) workshop on Semantic Parsing 2014*
- Multisource transfer learning for host-pathogen protein interaction prediction in unlabeled tasks, M. Kshirsagar, J. Carbonell and J. Klein-Seetharaman, *NIPS Workshop on Machine Learning for Computational Biology 2013*
- Confident prediction of Salmonella-human protein-protein interactions. S. Schleker, I. Nouretdinov, M. Kshirsagar, J. Klein-Seetharaman, A. Gammerman et al., *European Conf. Computational Biology 2012*
- Transfer learning based methods for new hosts: discovering host-pathogen protein-protein interactions. M. Kshirsagar, J. Carbonell and J. Klein-Seetharaman, *Intelligent Systems for Molecular Biology (ISMB) 2012*

Other professional activities

- **Reading groups:** Organized the matrix factorization reading group at CMU, Machine Learning reading group at IBM Research, Deep Learning reading group at MSKCC
- **Mentoring:** graduate students on rotation and undergrad interns at MSKCC
- **Patents:** Three patents on information extraction techniques (USPTO Publication # 20100223214, 20100257440, 20090216739)
- **Program Committee:** PLoS Computational Biology, Neural Computation, BMC Genomics 2013, IJCAI 2016, WWW Posters 2017-2018, Workshop for ML in Comp Bio 2016-2018, Biotechnology Journal 2017, NIPS 2016-2017, ICML 2017-2018, ICLR 2018, Co-organizer of ICML Workshop for Computational Biology (WCB 2017-2018)
- **White papers:** Wrapper Induction for automatic extraction, TechPulse 2008; Site-Specific Conditional Random Fields, TechPulse 2008; Web-Scale Information Extraction, TechPulse 2009
- **Invited talks:** Machine Learning seminar, Carnegie Mellon University (2015), Pro-active Learning and applications to Computational Biology, University of Pittsburgh (2013)
- **Posters:** Poster at Grad Expo 2010 at Univ. of Illinois, LTI Student Research Symposia 2012, 2013
- **Teaching Assistantship:** Machine Learning, Data-Mining and Information Retrieval
- **Organizational:** LTI Colloquium organization: identifying, inviting and hosting speakers. LTI Student committee: allocating student funds, organizing zero-waste events. At IIT Bombay: Elected as Cultural Secretary & Publications Coordinator, Alumni Secretary, Systems Admin for Hostel-11 and Mechanical Engg.

Outreach

- Co-founder of LaptopRehab, a campaign to donate phased out computers at Carnegie Mellon, and personal laptops to schools <http://sites.google.com/site/cmulaptoprehab>
- Sessions on CS and Machine Learning at Technights, a women@SCS workshop for school girls organized by Carnegie Mellon
- Organized Roadshows on Computer Science and Machine Learning at Pittsburgh schools

References

Jaime Carbonell

Judith Klein-Seetharaman

Chris Dyer

jgc@cs.cmu.edu

j.klein-seetharaman@warwick.ac.uk

cdyer@cs.cmu.edu