

# AMEET TALWALKAR

Machine Learning Department  
Carnegie Mellon University  
talwalkar@cmu.edu

---

## Experience

- 2018 – Assistant Professor at Carnegie Mellon University, Machine Learning Department
- 2017 – Co-founder and Chief Scientist at Determined AI
- 2016 – 2017 Affiliated Assistant Professor at University of California, Los Angeles, Statistics Department
- 2015 – 2017 Assistant Professor at University of California, Los Angeles, Computer Science Department
- 2014 – 2017 Technical Advisor at Databricks
- 2014 – 2015 Visiting Assistant Professor at University of California Berkeley, EECS Department
- 2010 – 2014 Postdoctoral Scholar at UC Berkeley, Computer Science Division (Advisor: Michael I. Jordan)
- 2007 – 2009 Intern at Google New York
- 2004 – 2006 Researcher in Paul Greengard's Neuroscience Laboratory at The Rockefeller University
- 2003 – 2004 Software Developer at Wireless Generation
- 2002 – 2003 Consultant at Oliver, Wyman & Company

## Education

- 2006 – 2010 M.S./Ph.D in Computer Science, Courant Institute, New York University  
Thesis: *Matrix Approximation for Large-scale Learning* (Advisor: Mehryar Mohri)
- 1998 – 2002 B.S. in Computer Science, summa cum laude, Yale University

## Awards and Honors

- 2016 edX Prize Finalist for Exceptional Contributions in Online Teaching and Learning
- 2016 Microsoft Azure Research Award
- 2015 Google Research Award
- 2015 Bloomberg Research Grant Award
- 2015 Amazon Web Services Research Award
- 2011 NSF Office of Cyberinfrastructure (OCI) Postdoctoral Fellowship
- 2011 Genentech Innovation Postdoctoral Fellowship
- 2011 Janet Fabri Prize for best doctoral dissertation in NYU's Computer Science Department
- 2009 Best Student Paper at New York Academy of Sciences' Symposium on Machine Learning
- 2008 Henning Biermann Award for exceptional service to NYU's Computer Science Department
- 2002 Yale Computer Science Undergraduate Prize
- 2001 Member of Yale's Phi Beta Kappa and Tau Beta Pi Honor Societies
- 2000 Yale Science and Engineering Association Junior High Scholarship Award

- 1998 Robert C. Byrd Scholarship  
1998 Westinghouse Science Talent Search Finalist

## **Publications** (\* alphabetical, † equal contribution)

### WORK IN SUBMISSION

- P. Chaudhari, C. Baldassi, R. Zecchina, S. Soatto, A. Talwalkar. *Parle: parallelizing stochastic gradient descent*.  
L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, A. Talwalkar. *A Novel Bandit-Based Approach to Hyperparameter Optimization*.

### BOOK AND BOOK CHAPTERS

- M. Mohri\*, A. Rostamizadeh\*, A. Talwalkar\*. *Foundations of Machine Learning*. The MIT Press, 2012.  
S. Kumar\*, M. Mohri\*, A. Talwalkar\*. “Ensemble Nyström.” In *Ensemble Machine Learning: Methods and Applications*. Editors: C. Zhang and Y. Ma. Springer, 2012.  
A. Talwalkar, S. Kumar, M. Mohri, H. Rowley. “Large-Scale Manifold Learning.” In *Manifold Learning Theory and Applications*. Editors: Y. Ma and Y. Fu. CRC Press, 2011.

### PEER-REVIEWED JOURNAL PUBLICATIONS

- X. Meng, J. Bradley, B. Yuvaz, E. Sparks, S. Venkataraman, D. Liu, J. Freeman, D. Tsai, M. Amde, S. Owen, D. Xin, R. Xin, M. Franklin, R. Zadeh, M. Zaharia, A. Talwalkar. *MLlib: Machine Learning in Apache Spark*. In *Journal of Machine Learning Research (JMLR)*, 2016.  
L. Mackey†, A. Talwalkar†, M.I. Jordan. *Distributed Matrix Completion and Robust Factorization*. In *Journal of Machine Learning Research (JMLR)*, 2015.  
A. Talwalkar†, J. Liptrap†, J. Newcomb, C. Hartl, J. Terhorst, K. Curtis, M. Bresler, Y. S. Song, M. I. Jordan, D. Patterson. *SMA\$H: A Benchmarking Toolkit for Variant Calling*. In *Bioinformatics*, 2014.  
N. Gong, A. Talwalkar, L. Mackey, L. Huang, R. Shin, E. Stefanov, E. Shi, D. Song. *Joint Link Prediction and Attribute Inference using a Social-Attribute Network*. In *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2014.  
A. Talwalkar, S. Kumar, M. Mohri, H. Rowley. *Large-scale SVD and Manifold Learning*. In *Journal of Machine Learning Research (JMLR)*, 2013.  
A. Kleiner, A. Talwalkar, P. Sarkar, M.I. Jordan, *A Scalable Bootstrap for Massive Data*. In *Journal of the Royal Statistical Society, Series B (JRSS-B)*, 2013.  
S. Kumar\*, M. Mohri\*, A. Talwalkar\*. *Sampling methods for the Nyström Method*. In *Journal of Machine Learning Research (JMLR)*, 2012.  
P. Radivojac, W. Clark, T. Oron, A. Schnoes, T. Wittkop, A. Sokolov, K. Graim, C. Funk, K. Verspoor, A. Ben-Hur, G. Pandey, J. Yunes, A. Talwalkar, et. al. *A Large-scale Evaluation of Computational Protein Function Prediction*. *Nature Methods*, 2012.

### PEER-REVIEWED CONFERENCE PUBLICATIONS

- V. Smith, C. Chiang, M. Sanjabi, A. Talwalkar. *Federated Multi-task Learning*. *Neural Information Processing Systems (NIPS)*, 2017.

- A. Amini, S. Kazemitabar, A. Bloniarz, A. Talwalkar. *Variable Importance Using Decision Trees*. Neural Information Processing Systems (NIPS), 2017.
- A. Das, I. Upadhyaya, X. Meng, A. Talwalkar. *Collaborative Filtering as a Case-Study for Model Parallelism on Bulk Synchronous Systems*. International Conference on Information and Knowledge Management (CIKM), 2017.
- L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, A. Talwalkar. *Hyperband: Bandit-Based Configuration Evaluation for Hyperparameter Optimization*. In International Conference on Learning Representations (ICLR), 2017.
- H. Qi, E. Sparks, A. Talwalkar. *Paleo: A performance model for deep neural networks*. In International Conference on Learning Representations (ICLR), 2017.
- F. Abuzaid, J. Bradley, F. Liang, A. Feng, L. Yang, M. Zaharia, A. Talwalkar. *Yggdrasil: An Optimized System for Training Deep Decision Trees at Scale*. In Neural Information Processing Systems (NIPS), 2016.
- A. Bloniarz, C. Wu, B. Yu, A. Talwalkar. *Supervised neighborhoods for distributed nonparametric regression*. In International Conference on Artificial Intelligence and Statistics (AISTATS), 2016.
- K. Jamieson, A. Talwalkar. *Non-stochastic Best Arm Identification and Hyperparameter Optimization*. In International Conference on Artificial Intelligence and Statistics (AISTATS), 2016.
- E. Sparks, A. Talwalkar, D. Haas, M. Franklin, M. I. Jordan, T. Kraska. *Automating Model Search for Large Scale Machine Learning*. In Symposium on Cloud Computing (SOCC), 2015.
- S. Agarwal, H. Milner, A. Kleiner, A. Talwalkar, B. Mozafari, M. I. Jordan, S. Madden, and I. Stoica. *Knowing When You're Wrong: Building Fast and Reliable Approximate Query Processing Systems*. In Special Interest Group on Management of Data (SIGMOD), 2014.
- A. Bloniarz<sup>†</sup>, A. Talwalkar<sup>†</sup>, J. Terhorst<sup>†</sup>, M. I. Jordan, D. Patterson, B. Yu, Y. S. Song. *Changepoint Analysis for Efficient Variant Calling*. In International Conference on Research in Computational Molecular Biology (RECOMB), 2014.
- E. Sparks, A. Talwalkar, V. Smith, J. Kottalam, X. Pan, J. Gonzalez, M. Franklin, M. I. Jordan, T. Kraska. *MLI: An API for Distributed Machine Learning*. In International Conference on Data Mining (ICDM), 2013.
- A. Talwalkar<sup>†</sup>, L. Mackey<sup>†</sup>, Y. Mu, S. Chang, M. I. Jordan. *Distributed Low-rank Subspace Segmentation*. In International Conference on Computer Vision (ICCV), 2013.
- A. Kleiner, A. Talwalkar, S. Agarwal, I. Stoica, M.I. Jordan. *A General Bootstrap Performance Diagnostic*. In Conference on Knowledge, Discovery and Data Mining (KDD oral), 2013.
- T. Kraska<sup>†</sup>, A. Talwalkar<sup>†</sup>, J. Duchi, R. Griffith, M. Franklin, M.I. Jordan. *MLbase: A Distributed Machine Learning System*. In Conference on Innovative Data Systems Research (CIDR), 2013.
- A. Kleiner, A. Talwalkar, P. Sarkar, M.I. Jordan, *The Big Data Bootstrap*. In International Conference on Machine Learning (ICML oral), 2012.
- L. Mackey<sup>†</sup>, A. Talwalkar<sup>†</sup>, M.I. Jordan. *Divide-and-Conquer Matrix Factorization*. In Neural Information Processing Systems (NIPS), 2011.
- M. Mohri<sup>\*</sup>, A. Talwalkar<sup>\*</sup>. *Can Matrix Coherence be Efficiently and Accurately Estimated?* In International Conference on Artificial Intelligence and Statistics (AISTATS oral), 2011.
- A. Talwalkar, A. Rostamizadeh. *Matrix Coherence and the Nyström Method*. In Conference on Uncertainty in Artificial Intelligence (UAI oral), 2010.
- C. Cortes<sup>\*</sup>, M. Mohri<sup>\*</sup>, A. Talwalkar<sup>\*</sup>. *On the Impact of Kernel Approximation on Learning Accuracy*. In International Conference on Artificial Intelligence and Statistics (AISTATS), 2010.
- S. Kumar<sup>\*</sup>, M. Mohri<sup>\*</sup>, A. Talwalkar<sup>\*</sup>. *The Ensemble Nyström Method*. In Neural Information Processing Systems (NIPS), 2010.
- S. Kumar<sup>\*</sup>, M. Mohri<sup>\*</sup>, A. Talwalkar<sup>\*</sup>. *On Sampling-based Approximate Spectral Decomposition*. In International Conference on Machine Learning (ICML), 2009.

- S. Kumar\*, M. Mohri\*, A. Talwalkar\*. *Sampling Techniques for the Nyström Method*. In International Conference on Artificial Intelligence and Statistics (AISTATS), 2009.
- C. Allauzen\*, M. Mohri\*, A. Talwalkar\*. *Sequence Kernels for Predicting Protein Essentiality*. In International Conference on Machine Learning (ICML), 2008.
- A. Talwalkar, S. Kumar, H. Rowley. *Large-Scale Manifold Learning*. In International Conference on Vision and Pattern Recognition (CVPR oral), 2008.

#### PEER-REVIEWED WORKSHOP PUBLICATIONS

- N. Gong, A. Talwalkar, L. Mackey, L. Huang, E. Shin, E. Stefanov, E. Shi, D. Song. *Jointly Predicting Links and Inferring Attributes using a Social-Attribute Network*. In ACM Workshop on Social Network Mining and Analysis (SNA-KDD), 2012.
- L. Boucher, R. Sekuler, A. Talwalkar, A. B. Sekuler. *Motion Perception is Influenced by Sound: Two- and Three-Dimensional Motion*. Association for Research in Vision and Ophthalmology, 1998.

#### TECHNICAL REPORTS

- K. Curtis, A. Talwalkar, M. Zaharia, A. Fox, D. Patterson. *Hybrid Genomic Processing Via Similar Regions*. 2015.

### Invited Talks

#### *Scalable Deep Learning*

- O'Reilly AI Conference, San Francisco, Sep. 2017.
- MIT, Boston, MA, Sep. 2017.

#### *Scalable Machine Learning Pipelines*

- CMU Machine Learning Department Seminar, Pittsburgh, PA, May 2017.
- Yale University CS Department Seminar, New Haven, CT, Mar. 2017.
- Brown University CS Department Seminar, Providence, RI, Mar. 2017.
- Netflix, Los Gatos, CA, Mar. 2017.

#### *A Novel Bandit-Based Approach to Hyperparameter Optimization*

- NIPS Workshop on Optimizing the Optimizers, Barcelona, Spain, Dec. 2016.
- Google Research, New York City, NY, Sep. 2016.
- UCLA Statistics Seminar Series, Los Angeles, CA, Feb. 2016.
- Information Theory and Applications Workshop, San Diego, CA, Feb. 2016.

#### *Paleo: A Performance Model for Deep Neural Networks*

- NIPS Workshop on Machine Learning Systems, Barcelona, Spain, Dec. 2016.

#### *Scalable and Accurate Local Learning via Random Forests*

- Joint Statistical Meetings, Chicago, IL, Aug. 2016.
- Bloomberg R&D Seminar, New York City, NY, June 2016.

#### *Machine Learning in Apache Spark*

- UT Austin Wireless Networking and Communications Group, Austin, TX, May 2015.
- Columbia Machine Learning Seminar, New York City, NY, Apr. 2015.
- USC CS Colloquium, Los Angeles, CA, Mar. 2015.
- Bloomberg R&D Seminar, New York City, NY, Mar. 2015.
- Information Theory and Applications Workshop, San Diego, CA, Feb. 2015.

#### *MLlib: Apache Spark's Machine Learning Library*

- AMPCamp5, Berkeley, CA, Nov. 2014.

- Spark User Meetup, Sunnyvale, CA, July 2014.
- Spark Summit 2014, San Francisco, CA, June 2014.

*Machine Learning in the Wild*

- Facebook AI Lab, Menlo Park, CA, Apr. 2014.
- UCSD CS Seminar, San Diego, CA, Apr. 2014.
- Wisconsin ECE Seminar, Madison, WI, Apr. 2014.
- Princeton CS Seminar, Princeton, NJ, Mar. 2014.
- UCLA CS Seminar, Los Angeles, CA, Mar. 2014.
- UCSD ECE Seminar, San Diego, CA, Feb. 2014.
- Harvard Biostatistics Seminar, Boston, MA, Feb. 2014.
- UCSC CS Seminar, Santa Cruz, CA, Feb. 2014.

*Addressing Speed and Accuracy in Variant Calling*

- Illumina, San Diego, CA, Jan. 2014.

*MLbase: A User-friendly System for Distributed Machine Learning*

- NIPS Workshop on Distributed ML and Matrix Computations, Montreal, Canada, Dec. 2014.
- The Machine Learning Conference (MLconf), San Francisco, CA, Nov. 2014.
- Stanford Spark Class, Palo Alto, CA, Aug. 2014.
- DIMACS Workshop on Systems and Analytics of Big Data, Piscataway, NJ, Mar. 2014.
- Information Theory and Applications Workshop, San Diego, CA, Feb. 2014.
- Strata Conference, Santa Clara, CA, Feb. 2014.
- Silicon Valley Machine Learning Meetup, Mountain View, CA, Jan. 2014.
- AMPCamp3, Berkeley, CA, Aug. 2013.
- Spark User Meetup, San Francisco, CA, Aug. 2013.
- SAP, Palo Alto, CA, July 2013.
- IBM Research Almaden, San Jose, CA, Apr. 2013.

*Changepoint Analysis for Efficient Variant Calling*

- UCSC Center for Biomolecular Science and Engineering Seminar, Santa Cruz, CA, Nov. 2013.

*MLI: An API for Distributed Machine Learning*

- Bay Area Machine Learning Symposium, Menlo Park, CA, Aug. 2013.

*SMaSH: A Benchmarking Suite for Variant Calling*

- UCSC Center for Biomolecular Science and Engineering Seminar, Santa Cruz, CA, May 2013.

*Scalable and User-Friendly Machine Learning*

- Netflix, Los Gatos, CA, Sep. 2013.
- Microsoft Silicon Valley (CISL Group), Mountain View, CA, Aug. 2013.
- SF Bay Area Machine Learning Meetup, San Francisco, CA, May 2013.
- AT&T, San Francisco, Apr. 2013.
- Brown University CS Department Seminar, Providence, RI, Mar. 2013.

*Divide-and-Conquer Learning for Big Data*

- Boston University ECE Seminar, Boston, MA, Feb. 2013.

*Divide-and-Conquer Matrix Factorization*

- The Machine Learning Conference (MLconf), San Francisco, CA, Nov. 2013.
- Information Theory and Applications Workshop, San Diego, CA, Feb. 2013.
- INRIA, Paris, France, Dec. 2012.
- NIPS Workshop on Sparse and Low-rank Approximation, Sierra Nevada, Spain, Dec. 2011.

*Matrix Approximation for Large-scale Learning*

- NYU CS Seminar, New York City, NY, May 2010.
- Statistical Artificial Intelligence Lab Seminar, Berkeley, CA, Mar. 2010.
- Google Research, New York City, NY Feb. 2010.

### *The Ensemble Nyström Method*

- New York Academy of Sciences, New York City, NY, Nov. 2009.

### *Improved Bounds for the Nyström Method*

- New York Academy of Sciences, New York City, NY, Oct. 2008.

### *Sequence Kernels for Predicting Protein Essentiality*

- Google Research, New York City, NY, Aug. 2008.

## **Presentations**

*MLlib: A Distributed Machine Learning Library*. NIPS Workshop on Machine Learning Open Source Software, Lake Tahoe, CA, Dec. 2013.

*MLbase: A User-friendly System for Distributed Machine Learning*. GraphLab Workshop, San Francisco, CA, July 2013.

*Research Highlight - MLbase*. NSF Expeditions in Computing PI Meeting, Arlington, VA, May 2013.

*MLbase: A Distributed Machine Learning Wrapper*. NIPS Workshop on Big Learning, Lake Tahoe, CA, Dec. 2012.

*Distributed Pipeline for Genomic Variant Calling*. NIPS Workshop on Big Learning, Lake Tahoe, CA, Dec. 2012.

*Divide-and-Conquer Subspace Segmentation*. NIPS Workshop on Big Learning, Lake Tahoe, CA, Dec. 2012.

*Bootstrapping Big Data*. In NIPS Workshop on Big Learning, Sierra Nevada, Spain, Dec. 2011.

*Genome-Scale Phylogenetic Function Annotation of Large and Diverse Protein Families*. In Scientific Discovery through Advanced Computing (SciDAC) Conference, Denver, CO, July 2011.

*Very Large-scale Low-rank Approximation*. In NIPS Workshop on Large-Scale Machine Learning: Parallelism and Massive Datasets, Whistler, Canada, Dec. 2009.

## **Professional Activities**

### REVIEWING

*Conferences*: Neural Information Processing Systems (NIPS), International Conference on Machine Learning (ICML), International Conference on Artificial Intelligence and Statistics (AISTATS), Conference on Learning Theory (COLT), Algorithmic Learning Theory (ALT), International Workshop on Mining and Learning with Graphs (MLG), International Conference on Pattern Recognition Applications and Methods (ICPRAM), Special Interest Group on Management of Data (SIGMOD)

*Journals*: Journal of Machine Learning Research (JMLR), IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), Neural Networks, IEEE Transactions on Information Theory, Foundations and Trends in Machine Learning (FnTML), IEEE Transactions on Neural Networks and Learning Systems, Data Mining and Knowledge Discovery

### SEMINAR / WORKSHOP ORGANIZATION

2015 – 2016 UCLA Computer Science Weekly Seminar Series

2014 – 2016 Information Theory and Applications Workshop Session Organizer

2015 Co-organizer, 3-day course on *Distributed Analytics and Machine Learning with Apache Spark* at Berkeley Institute for Data Science, UC Berkeley

2014 Co-organizer, Workshop on *Distributed Machine Learning and Matrix Computations* at Neural Information Processing Systems, Montreal, Canada

2014 Organizer, AMPCamp 5, UC Berkeley

- 2013 Organizer, Systems and Machine Learning (SysML) Seminar, AMPLab, UC Berkeley
- 2011 Co-organizer, Workshop on *Sparse Representation and Low-rank Approximation* at Neural Information Processing Systems, Sierra Nevada, Spain
- 2010 Co-organizer, Workshop on *Low-rank Methods for Large-scale Machine Learning* at Neural Information Processing Systems, Whistler, Canada
- 2009 Co-organizer, Machine Learning Seminar, Courant Institute, NYU

OTHER

- 2017 NSF Panelist
- 2010 NSF Panelist
- 2006 – 2008 Courant Institute Graduate Student Representative