GHC-8120, CMU
4902 Forbes Ave
Pittsburgh, PA15213
☎ *+1-412-251-6037*
✉ *yuxiangw@cs.cmu.edu*
www.cs.cmu.edu/ yuxiangw/

Yu-Xiang Wang

Briefly about me

I am a Statistics and Machine Learning PhD candidate at the School of Computer Science, Carnegie Mellon University. I work with Steve Fienberg, Alex Smola, Ryan Tibshirani and Jing Lei. My work focuses on developing provable and practical methods for various challenging learning regimes (e.g., high dimensional, heterogeneous, privacy-constrained, sequential, parallel and distributed) and often involves exploiting hidden structures in data (generalized sparsity, union-of-subspace, graph or network structures), balancing various resources (model complexity, statistical power and privacy budgets) and developing scalable optimization tools. These work lead to applications in different domains including computer vision, recommender systems, web security, urban taxi analytics and even political campaigns.

Education

2013–Present **Ph.D. in Statistics and Machine Learning**, *Carnegie Mellon University*, Pittsburgh.
GPA: 4.21/4. Expected Date of Graduation: Summer 2017

2012–2013 **M.Eng in Electrical Engineering**, *National University of Singapore*, Singapore.
GPA: 4.88/5. Thesis: Robust Learning of Low-dimensional Structures: Theory, Algorithms and Applications

2007–2011 **B.Eng in Electrical Engineering (1st Class Honours) with Minor in Math.**, *National University of Singapore*, Singapore.
GPA:*4.91/5*. Rank: 2nd out of 325 students

Work/Research Experience

2016 summer **Research intern**, *Microsoft Research*, NYC.
- Worked on the optimal off-policy evaluation for contextual bandit, with potential application to Bing search, Cortana and other web services that interacts with users.

2014 summer **Research intern**, *Centre for Urban Science and Progress*, NYC.
- Worked on data science projects involving NYC taxi data/NYPD data.

2013–Present **Research assistant**, *Machine Learning Department*, CMU.
- Locally adaptive nonparametric regression via trend filtering.
- Practical differential privacy for machine learning.
- Large-scale/parallel and distributed optimization for machine learning.

2011–2013 **Research engineer**, *Control and Mechatronics Lab, Mechanical Engineering*, NUS.
- Studied the theory and applications of machine learning. In particular collaborative filtering/recommendation systems/matrix completion, matrix factorization and subspace clustering.

2012–2013 **Software engineer (part-time)**, *Mozat (a mobile social network company)*.
- Designed and developed an intelligent friend ranking system.
- Web development with Python/Django.

2010–2011 **Final Year Project**, *Unmanned System Research Group*, ECE, NUS.

- Built the vision navigation system for an indoor unmanned aerial vehicle (UAV).
- Won two gold medals in the SAFMC Competition.

Publications

Preprint

[1] Yu-Xiang Wang, Alekh Agarwal, Miro Dudik. Optimal and Adaptive Off-policy Evaluation in Contextual Bandit. In preparation.

[2] Yu-Xiang Wang, Jing Lei, Steve Fienberg. A Minimax Theory for Adaptive Data Analysis. In preparation.

[3] Ziqi Liu, Alex Smola, Kyle Soska, Yu-Xiang Wang. Attributing Hacks. Submitted.

[4] Seth Flaxman, Dougal Sutherland, Yu-Xiang Wang, and Yee Whye Teh . Understanding the 2016 US Presidential Election using ecological inference and distribution regression with census microdata . In preparation.

[5] Yu-Xiang Wang, Huan Xu. Provable Subspace Clustering: When LRR meets SSC. Under review at IEEE Trans. on Information Theory.

Journal

[6] Yu-Xiang Wang, Jing Lei, Steve Fienberg. Learning with Differential Privacy: Stability, Learnability and the Sufficiency and Necessity of ERM Principle. In *JMLR, 2016.*

[7] Yu-Xiang Wang, James Sharpnack, Alex Smola and Ryan Tibshirani. Trend Filtering on Graphs. In *JMLR, 2016 (Short version appeared at AISTATS'2015)*

[8] Yu-Xiang Wang and Huan Xu. Noisy Sparse Subspace Clustering. In *JMLR, 2016 (Short version appeared at ICML'2013)*

[9] Yu-Xiang Wang, Choon Meng Lee, Loong Fah Cheong. Practical Matrix Completion and Corruptions Recovery using Alternating Robust Subspace Minimization. In *International Journal on Computer Vision (IJCV), 2014*

[10] Zhi Gao, Yu-Xiang Wang, Loong Fah Cheong. Block RPCA for Consistent Moving Object Detection. in *IEEE TPAMI, 2013*

Conference

(* denotes "co-first authors")

[11] Veeranjaneyulu Sadhanala*, Yu-Xiang Wang*, Ryan Tibshirani. Total Variation Classes Beyond 1d: Minimax Rates, and the Limitations of Linear Smoothers. In *Advances in Neural Information Processing Systems (NIPS)*, 2016

[12] Yu-Xiang Wang, Jing Lei, Steve Fienberg. On-Average KL-Privacy and its equivalence to Generalization for Max-Entropy Mechanisms. In *Privacy in Statistical Databases, 2016.*

[13] Yu-Xiang Wang, Veeru Sadhanala, Wei Dai, Willie Neiswanger, Eric Xing. Parallel and Distributed Block-Coordinate Frank-Wolfe. In *International Conference of Machine Learning (ICML)*, 2016

[14] Veeranjaneyulu Sadhanala*, Yu-Xiang Wang*, Ryan Tibshirani. Graph Sparsification Approaches for Large-Scale Laplacian Smoothing.. In *AISTATS'2016*.

[15] Yining Wang, Yu-Xiang Wang, Aarti Singh. Graph Connectivity in Noisy Sparse Subspace Clustering. In *AISTATS'2016*.

[16] Mu Li, Ziqi Liu, Alex Smola, Yu-Xiang Wang. DiFacto — Distributed Factorization Machines. In *ACM International Conference on Web Search and Data Mining (WSDM)*, 2016 [Best Paper Honorable Mention]

[17] Yining Wang, Yu-Xiang Wang and Aarti Singh. Differentially Private Subspace Clustering. In *Advances in Neural Information Processing Systems (NIPS)*, 2015

[18] Yu-Xiang Wang, Steve Fienberg and Alex Smola. Privacy for Free: Posterior Sampling and Stochastic Gradient Monte Carlo. In *International Conference of Machine Learning (ICML)*, 2015

[19] Ziqi Liu, Yu-Xiang Wang, Alex Smola. Differential Private Matrix Factorization. In *RecSys'15*.

[20] Seth Flaxman, Yu-Xiang Wang, Alex Smola. Who Supported Obama in 2012? Ecological inference through distribution regression. In *KDD'15*. [Best Student Paper Award]

[21] Yining Wang, Yu-Xiang Wang and Aarti Singh. A Deterministic Analysis for Sparse Subspace Clustering for Dimension-reduced data. In *International Conference of Machine Learning (ICML)*, 2015

[22] Yu-Xiang Wang, Alex Smola, Ryan Tibshirani. The Falling Factorial Basis and Its Statistical Applications. In *International Conference of Machine Learning (ICML)*, 2014

[23] Yu-Xiang Wang, Huan Xu and Chenlei Leng. Provable Subspace Clustering: When LRR meets SSC. In *Advances in Neural Information Processing Systems (NIPS)*, 2013

[24] Yu-Xiang Wang and Huan Xu. Noisy Sparse Subspace Clustering. In *International Conference of Machine Learning (ICML)*, 2013.

[25] Yu-Xiang Wang and Huan Xu. Stability of Matrix Factorization for Collaborative Filtering. In *International Conference of Machine Learning (ICML)*, 2012

[26] Yu-Xiang Wang, An Efficient Algorithm for UAV Indoor Pose Estimation using Vanishing Geometry. In *IAPR Conference on Machine Vision Application (MVA)*, 2011

Honors and Awards

2016 **WSDM'16 Best Paper Honorable Mention**.
For our paper: "Difacto – Distributed Factorization Machines".

2015 **MSR Fellowship finalist**.
Last round of the selective fellowship (only 30 among all math/CS/ECE PhD students in US).

2015 **Baidu Scholarship**.
Awarded to only 10 worldwide. Competed globally among CS PhD students of Chinese origin.

2015 **KDD'15 Best Student Paper Award**.
For our paper: "Who voted for Obama in 2012?".

2014 **NIPS'14 Outstanding Reviewer Award**.
For the outstanding quality in the reviews.

2014–2015 **ICML Travel Scholarship**.
To subsidize student travel. Sponsored by NSF, Google and etc.

2011 **1st Prize, IEEE Control Chapter (Singapore) Prize**.
For the best undergraduate thesis in control.

2011 **Gold Medal, Singapore Amazing Flying Machine Competition**.
Awarded to the winner of the task-based UAV competition organized by Defense Science Organization.

2008 **IET Singapore Scholarship**, *Institute of Engineering and Technology*.
Awarded to students in Singapore for outstanding academic performance and extra-curriculum activities.

2007–2011 **Dean's List (every eligible semester)**, *NUS Faculty of Engineering*.
Awarded to top 5% students in Electrical Engineering each semester.

2007–2011 **MOE Undergraduate Scholarship for PRC Students**.
A highly selective scholarship awarded to selected fresh undergraduates students from China, covering full undergraduate tuition and living stipend

2005 **1st Prize, China's National Physics Olympiad (NPhO)**.
Awarded to top 0.5% NPhO participants (only 15 in 2005) in Shaanxi Province

Talks

2013-Present **Invited talks**.

- 2016 JSM Chicago: "Learning with Differential Privacy"
- 2016 ICML Privacy Workshop: "Learning with Differential Privacy and On-Average KL-privacy"
- 2016 Columbia U: "Trend Filtering and Optimal TV-Denoising"
- 2016 Office of Financial Research, D.C.: "Practical machine learning with Differential Privacy and Beyond"
- 2015 Google Pittsburgh: "Practical machine learning with Differential Privacy and Beyond"
- 2015 Carnegie Mellon University ML seminar: "Trend Filtering, Falling Factorial Basis and Adaptive Statistical Estimation on Graphs"
- 2015 National University of Singapore: "Trend Filtering: Some Recent Advances and Challenges"
- 2015 Singapore Management University: "On Trend filtering and Differential Privacy"
- 2015 CMU Math seminar: "Trend Filtering: Some Recent Advances and Challenges"
- 2014 Center for Urban Science and Progress, NYC: "Learning with Differential Privacy"
- 2013 Columbia U: "Noisy Sparse Subspace Clustering"

2012-Present **Conference/workshops contributed talks**.

- NIPS 2016 Workshop on Adaptive Data Analysis: "Gaussian Adaptive Data Analysis"
- NIPS 2016 Barcelona, "What-If?" Workshop: "Optimal and Adaptive Off-Policy Evaluation in Contextual Bandits"
- PSD 2016 Dubrovnik: "On-Average KL-Privacy and its equivalence to generalization
- ICML 2016 NYC: "Parallel and Distributed Block-Coordinate Frank-Wolfe
- NIPS 2015 Montreal: "Learning with Differential Privacy"
- ICML 2015 Lille: "Privacy for Free: Posterior Sampling and Stochastic gradient MCMC"
- ICML 2015 Workshop on Privacy and incomplete information: "Learning with Differential Privacy"
- ICML 2014 Beijing: "Falling Factorial Basis and its Statistical Applications"
- NIPS 2013 Tahoe: "Provable subspace clustering: When LRR meets SSC"
- ICML 2012 Edinburgh: "Stability of Matrix Factorization for Collaborative Filtering"

Professional activities

2013-Present **Reviewer for conferences and journals in ML, Statistics, ECE and CS**.

- International Conference on Machine Learning (ICML)
- Neural Information Processing Systems (NIPS) [Outstanding Reviewer Award 2014]
- Conference on Learning Theory (COLT)
- International Conference on Artificial Inteligence and Statistics (AISTATS)
- Annals of Statistics
- Biometrika
- Annals of the Institute of Statistical Mathematics (AISM)
- Journal of Privacy and Confidentiality
- Privacy for Statistical Databases (PSD'2016)
- IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)
- IEEE Trans. on Control of Network Systems (TCNS)
- SIAM Journal of Computing (SICOMP)
- Theory of Computing (ToC)

2014-Present **Program commitee for the following workshops**.

- NIPS'15 Workshop on Learning with Incomplete Information and Privacy
- AAAI'15 Workshop on AI for Cities
- CMU Machine Learning Symposium'15.

2014-2015 **Teaching assistant for these courses**.

- 10-801 Randomized Algorithms and Advanced Optimization [Suvrit Sra and Alex Smola]
- 10-725 Convex Optimization [Ryan Tibshirani]

Advanced technical courses taken (selected)

2013-2014 **CMU Courses**.

- Introduction to Machine Learning
- Intermediate Statistics
- Probabilistic Graphical Models
- Advanced Optimization and Randomized Algorithms
- Advanced Data Analysis I & II
- Statistical Machine Learning
- Advanced Statistics Theory
- Advanced Probability Overview
- Data Privacy
- Spectral graph theory (audit)
- Foundations of ML and Data Science (audit)

2011-2013 **NUS Courses**.

- 3D Computer Vision
- Pattern Recognition
- Biological Perception
- Visual Computing
- Advanced Topics in Vision and Machine Learning
- Advanced Control Systems
- Multivariable Control Systems

Language and Computer Skills

Language Chinese Mandarin(native), English(fluent), German (basic)

Programming C/C++, C#, Java, R, Python, Matlab, Shell

References

Stephen E. Fienberg
Maurice Falk University Professor of Statistics and Social Science
Department of Statistics, Machine Learning Department, CyLab, and i-Lab
Carnegie Mellon Univeristy
Pittsburgh, PA 15213
✉ mmv@cs.cmu.edu
☎ (412) 268-1474

Ryan J. Tibshirani
Associate Professor
Department of Statistics and Machine Learning Department
Carnegie Mellon University
Pittsburgh, PA 15213
✉ ryantibs@cmu.edu
☎ (412) 268-1884

Jing Lei
Assistant Professor
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213
✉ jinglei@andrew.cmu.edu
☎ (412) 268-8781

Alexander J. Smola
Director at Amazon
Professor
Machine Learning Department
Carnegie Mellon University
✉ alex@smola.org
☎ (408) 759-1044

Other references upon request

Huan Xu
Assistant Professor
School of Industrial & Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332
✉ huan.xu@isye.gatech.edu
☎ (404) 385-5385

Suvrit Sra
Principal Research Scientist
Laboratory for Information & Decision Systems (LIDS), Data Science Center
Massachusetts Institute of Technology
Cambridge, MA 02139
✉ suvrit@mit.edu
☎ (617) 253-3816

Alekh Agarwal
Researcher
Microsoft Research NYC
New York , NY 10011
☎ alekha@microsoft.com

Miroslav Dudík
Senior Researcher
Microsoft Research NYC
New York, NY 10011
☎ mdudik@cmu.edu