

HUA OUYANG

Curriculum Vitae

701 First Ave
Sunnyvale, CA 94089
☎ +1-678-860-4067
✉ huaouyang@gmail.com
🌐 <http://learningr.com>

OBJECTIVE

Solve big data problems with advanced and scalable machine learning techniques.

HIGHLIGHTS OF EXPERTISE

- Enthusiasm and intuitions in bringing research methodologies to abstract engineering problems, leading to high precision solutions.
- Industrial experience in developing novel machine learning models, algorithms and systems for mobile search, mail and web security products.
- Experience in leading a industrial research team and conducting applied science at a fast pace for product development.
- Industrial experience in designing web-scale machine learning platforms. Developed and deployed a Hadoop-based parallel Gradient Boosting tool on thousands of nodes in Yahoo. Deployed a deep learning platform on multiple-GPUs.
- Years of rigorous training and research experience in statistical and computational machine learning theory, design and analysis of scalable learning and optimization algorithms, optimal stochastic optimization and distributed learning, large-scale applications for text, image, video and speech processing.

EDUCATION

Ph.D. in Computer Science

2007–2013 **Georgia Institute of Technology**, *School of Computational Science and Engineering*, Atlanta.

Area: Machine Learning Theory and Algorithms

Thesis: Optimal Stochastic and Distributed Algorithms for Machine Learning

Advisors: Alexander G. Gray, Hongyuan Zha

Thesis Committee: Alexander G. Gray, Hongyuan Zha,

Arkadi Nemirovski, Haesun Park, Maria Balcan

M.Phil. in Electronic Engineering

2004–2007 **The Chinese University of Hong Kong**, *Department of Electronic Engineering*, Hong Kong.

Area: Speech and Video Signal Modeling and Inference, Pattern Recognition

Thesis: Classification and Fusion Methods for Multimodal Biometric Authentication

B.Eng. in Electronic and Information Engineering

1999–2003 **Huazhong University of Science and Technology**, *Department of Electronics and Information Engineering*, Wuhan.

Thesis: Telephone Environment Speaker Verification

WORK EXPERIENCE

Research Scientist

- 2014-Now **Yahoo Labs**, Sunnyvale, CA.
Leading Yahoo Labs' deep learning research team. Project design and goal planning with managers and tech leads in production teams. Hunting talents from both academia and industry and scheduling interviews. Organized the Deep Learning workshop in Yahoo Labs' Science Week.
- 2014-Now **Yahoo Labs**, Sunnyvale, CA.
Designed and deployed deep learning (feed-forward and LSTM recurrent neural networks) platform over multiple-GPUs, in collaboration with Flickr. This system is built over Yahoo's Hadoop grids. GPU machines can directly access Yahoo's huge data on HDFS, enabling large-scale training. We implemented fast device-to-device communication for exchanging large neural networks.
- 2013-Now **Yahoo Labs**, Sunnyvale, CA.
Designed feed-forward and recurrent neural networks (LSTM) for query rewriting, query classification and search relevance ranking. Improved online relevance metric (DCG5) by 7% for tail queries.
Deployed a new learning-to-rank loss function for Yahoo mobile search (patent filed).
Developed a multi-armed-bandit algorithm for query auto suggestion (patent filed).
Deployed a graph-based effort propagation for commercial spam detection (patent filed).
Deployed a real-time automatic fraud detection system for Yahoo Mail.
- 2013-2014 **Yahoo Labs**, Sunnyvale, CA.
Developed and deployed a parallel Gradient Boosting tool. This GBDT tool can run on both shared memory machines and thousands of distributed Hadoop nodes, leading to over 100x speedup. It is being used on a daily basis by many production teams at Yahoo.

Research Assistant

- 2007-2013 **Skytree Inc.**, Atlanta, GA.
Main contributor to the open-source machine learning tool MLPACK (<http://www.mlpack.org/>). Its commercial version evolved as the core product of the San Jose based machine learning start-up Skytree Inc. (<http://www.skytree.net/>), co-founded by my advisor Prof. Alexander G. Gray

Research Assistant

- 2007-2013 **Fundamental Algorithmic and Statistical Tools Laboratory (FASTLab)**, *Computational Science and Engineering, Georgia Tech*, Atlanta, GA.
Conducted research on fast algorithms for nearest neighbor search on large-scale astronomical survey data, supervised ranking and rank-based dimension reduction for visualization, stochastic kernel machines, optimal stochastic and noise-adaptive online learning algorithms, data-distributed and communication-robust learning and matrix factorization algorithms. Collaborating with Prof. Arkadi Nemirovski on the analysis of optimal stochastic algorithms for optimizing separable and structural functions.

Research Staff Intern

- Summer 2010 **Multimedia Research Group**, *IBM Thomas J. Watson Research Center*, Hawthorne, NY.
Collaborated with Lexing Xie, John Smith and their group on supervised metric learning for web-scale content-based image retrieval. Participated in the ImageNet Large Scale Visual Recognition Challenge 2010 as a representative of IBM Research. Helped writing a research grant proposal on video-based event detection from social network.

Research Assistant

- 2004-2007 **Digital Signal Processing and Speech Technology Laboratory (DSPSTL)**, *Electronic Engineering, The Chinese University of Hong Kong*, Hong Kong.
Conducted research with Prof. Tan Lee on statistical modeling and inference of speech and video data, multi-modal temporal feature representation and robust classifier-fusion algorithms for large-scale audio-visual speaker verification in mobile environment.

REFEREED PUBLICATIONS

- 2015 Acar Tamersoy, Hua Ouyang, Duen Horng Chau. **Effort-based Detection of Comment Spammers**, *36th IEEE Symposium on Security and Privacy (S & P 2015)*, San Jose, CA, 2015

- 2013 Hua Ouyang, Niao He, Long Q. Tran, Alexander Gray. **Stochastic Alternating Direction Method of Multipliers**, In *Proceedings of the 30th International Conference on Machine Learning (ICML 2013)*, Atlanta, GA, 2013
- 2012 Hua Ouyang, Niao He, Alexander Gray. **Stochastic ADMM for Nonsmooth Optimization**, *NIPS 5th International Workshop on "Optimization for Machine Learning" (OPT 2012)*, Lake Tahoe, NV, 2012
- Hua Ouyang, Alexander Gray. **Stochastic Smoothing for Nonsmooth Minimizations: Accelerating SGD by Exploiting Structure**, In *Proceedings of the 29th International Conference on Machine Learning (ICML 2012)*, Edinburgh, United Kingdom, 2012
- Hua Ouyang, Alexander Gray. **NASA: Achieving Lower Regrets and Faster Rates via Adaptive Stepsizes**, In *Proceedings of the 18th ACM Conference on Knowledge Discovery and Data Mining (KDD 2012)*, Beijing, China, 2012
- 2010 Apostol Natsev, John R. Smith, Matthew Hill, Gang Hua, Bert Huang, Michele Merler, Lexing Xie, Hua Ouyang, Mingyuan Zhou. **IBM Research TRECVID 2010 Video Copy Detection and Multimedia Event Detection System**, *NIST TRECVID Workshop*, 2010
- Hua Ouyang, Alexander Gray. **Fast Stochastic Frank-Wolfe Algorithms for Nonlinear SVMs**, In *Proceedings of the SIAM Intl. Conf. on Data Mining (SDM 2010)*, Columbus, Ohio, 2010
- Winner of the **Best Student Paper Award** from the American Statistical Association
- Hua Ouyang, Alexander Gray. **Stochastic Mirror Descent Algorithm for L1-Regularized Risk Minimizations**, In *Proceedings of the IEEE International Conference on Computer and Information Technology (CIT 2010)*, Bradford, United Kingdom, 2010
- 2009 Parikshit Ram, Dongryeol Lee, Hua Ouyang, Alexander Gray. **Rank-Approximate Nearest Neighbor Search: Retaining Meaning and Speed in High Dimensions**, In *Advances in Neural Information Processing Systems 22 (NIPS 2009)*, Vancouver, Canada, 2009
- Rajesh Narasimha, Hua Ouyang, Alexander Gray, Steven W. McLaughlin, Sriram Subramaniam. **Automatic Joint Classification and Segmentation of Whole Cell 3D Images**, In *Pattern Recognition*, vol. 42, issue 6, 2009
- 2008 Hua Ouyang, Alexander Gray. **Learning Nearest-Neighbor Classifiers with Hyperkernels**, *NIPS Workshop on "Kernel Learning"*, Whistler, Canada, 2008
- Hua Ouyang, Alexander Gray. **Learning Dissimilarities by Ranking: From SDP to QP**, In *Proceedings of the 25th International Conference on Machine Learning (ICML 2008)*, Helsinki, Finland, 2008
- 2007 Rajesh Narasimha, Hua Ouyang, Alexander Gray. **Automatic Mining of Whole Cell Tomograms for Cancer Detection**, *NIPS workshop on Machine Learning in Computational Biology (MLCB 2007)*, Vancouver, Canada, 2007
- 2006 Hua Ouyang, Tan Lee. **Feature Extraction from Talking Mouths for Video-Based Bi-Modal Speaker Verification**, In *Proceedings of the International Conference on Acoustic Speech and Signal Processing (ICASSP 2006)*, Toulouse, France, 2006
- W.N. Chan, Tan Lee, Nengheng Zheng and Hua Ouyang, Tan Lee. **Use of Vocal Source Features in Speaker Segmentation**, In *Proceedings of the International Conference on Acoustic Speech and Signal Processing (ICASSP 2006)*, Toulouse, France, 2006
- 2005 Hua Ouyang, Tan Lee. **A New Lip Feature Representation Method for Video-based Bimodal Authentication**, In *Proceedings of the NICTA International Multimodal User Interaction Workshop (MMUI 2005)*, Sydney, Australia, 2005

OTHER PUBLICATIONS

- 2014 Makoto Yamada, Avishek Saha, Hua Ouyang, Dawei Yin, Yi Chang. **N³LARS: Minimum Redundancy Maximum Relevance Feature Selection for Large and High-dimensional Data**, *ArXiv:1411.2331*, 2014
- Ph.D. Dissertation Hua Ouyang. **Optimal Stochastic and Distributed Algorithms for Machine Learning**, *Georgia Institute of Technology*, Atlanta, GA 2013
- 2011 Hua Ouyang, Alexander Gray. **Data-Distributed Weighted Majority and Online Mirror Descent**, *ArXiv:1105.2274*, 2011
- M.Phil. Thesis Hua Ouyang. **Classification and Fusion Methods for Multimodal Biometric Authentication**, *The Chinese University of Hong Kong*, Hong Kong, 2007

INVITED TALKS

- April 2015 **Nvidia GPU Conference GTC 2015**, *Deep Learning at Yahoo Labs*, San Jose, CA.
- August 2012 **Baidu Inc., Web Search & NLP Groups**, *Stochastic Smoothing with Applications to Big Data*, Beijing, China.
- May 2012 **UC Berkeley "From Data to Knowledge Workshop"**, *Stochastic Smoothing for Nonsmooth Minimizations: Accelerating SGD by Exploiting Structure*, Berkeley, CA.
- August 2010 **Best Student Paper Presentation at the Joint Statistical Meetings (JSM 2010)**, *Stochastic Frank-Wolfe Algorithms for Nonlinear SVMs*, Vancouver, Canada.
- July 2010 **IBM Thomas J. Watson Research Center**, *Rank-Approximate Nearest Neighbor Search: Retaining Meaning and Speed in High Dimensions*, Hawthorne, NY.

HONORS & AWARDS

- 2014 Yahoo Faculty, Research and Engagement Program Awards (FREP Award 2014), Yahoo Labs
- 2013 Yahoo Labs Excellence Awards (LEAP Award 2013), Yahoo Labs
- 2012 Student Travel Award, ACM Conference on Knowledge Discovery and Data Mining (KDD 2012)
- 2012 Travel Award for Invited Speakers, UC Berkeley "From Data to Knowledge Workshop"
- 2010 Best Paper Finalist, SIAM International Conference on Data Mining (SDM 2010)
- 2012 Student Travel Award, SIAM International Conference on Data Mining (SDM 2010)
- 2010 Travel Award for ASA Best Student Paper Award Recipient, Joint Statistical Meetings (JSM 2010)
- 2010 Winner of the Best Student Paper Competition, American Statistical Association (ASA)- Statistical Computing
- 2010 Nominated by the College of Computing of Georgia Tech for Google PhD Fellowship
- 2008 Student Travel Award, International Conference on Machine Learning (ICML 2008)
- 2006 Merit Award, IEEE Signal Processing Hong Kong Chapter Postgraduate Forum
- 2005 1st Runner-Up, Microsoft Imagine Cup 2005, Hong Kong
- 2004-2006 Graduate Student Fellowship, The Chinese University of Hong Kong
- 1999-2003 Top Academic Excellency Scholarship, Huazhong University of Science and Technology
- 1999 Honors at Entrance Scholarship, Huazhong University of Science and Technology
- 1999 1st Class Prize of Hubei Province in China Mathematics Olympiad, Hubei, China