

# Wei Dai (David)

---

CONTACT INFORMATION	Machine Learning Department, Carnegie Mellon University Homepage: <a href="http://www.davidwd.org">http://www.davidwd.org</a>	Email: <a href="mailto:david@davidwd.org">david@davidwd.org</a>
RESEARCH INTERESTS	Large-scale machine learning; Parallel machine learning algorithms and theory; Deep learning and medical imaging	
EDUCATION	<b>Carnegie Mellon University</b> , Pittsburgh, PA	
	<i>Ph.D. Candidate in Machine Learning</i>	<b>2012 – 2018</b>
	• Research Advisor: Eric P. Xing	
	<i>M.Sc. in Machine Learning</i>	<b>2012 – 2016</b>
	<b>California Institute of Technology</b> , Pasadena, CA	
	<i>B.Sc. with Honor in Computer Science</i>	<b>2010 – 2012</b>
	• Research Advisor: Andreas Krause	
	<b>Wesleyan University</b> , Middletown, CT	
	<i>B.A. with High Honor in Physics and Mathematics</i>	<b>2007 – 2010</b>
	• Research Advisor: Francis W. Starr	
EMPLOYEMENT	<b>Petuum Inc</b> , Pittsburgh, PA	
	<i>Senior Director of Engineering</i>	<b>Sept 2017 – Present</b>
	<i>Director of Product Development</i>	<b>May 2017 – Sept 2017</b>
	<i>Senior Manager</i>	<b>Jul 2016 – May 2017</b>
	<ul style="list-style-type: none"><li>• As part of the founding team, I help grow the engineering team from &lt;10 in 2016 to over 60 full-time engineers in 2Q18. Both a trench worker and a manager, I influence product design as a domain expert in machine learning, and deliver quarterly internal platform product milestones from 3Q17 - 1Q18. Currently I lead the engineering division, marshaling the work of 8 teams and 4 front-line managers.</li><li>• I led the former medical imaging team and personally developed best in class deep learning methods around chest x-ray to assist radiologists in diagnosis. The project led to the first contract for the company. Petuums healthcare solution assist doctor workflows by leveraging image, textual, and structured data such as chest x-rays, CT scan, admission note, lab test, and others. The team has grown to cover other non-healthcare industries.</li><li>• During a product pivot, I served as tech lead and co-designed and implemented a unified framework that flexibly combines all data and machine learning functions developed across platform and solution teams.</li><li>• Proposed and implemented org adjustment within platform teams with broad feedback and support, resulting in a simpler org structure that empowers engineering managers and tech leads.</li><li>• Fostering engineering culture by hosting bi-weekly tech talks, performing 1-on-1, planning and executing performance review and promotions for all engineering. Recommended engineering ladders and support long-term efforts in formalizing engineering and management titles.</li></ul>	
	<b>Bosch Research</b> , Pittsburgh, PA	
	<i>Research Intern</i>	<b>May 2016 – Aug 2016</b>
	<ul style="list-style-type: none"><li>• Used Tensorflow on a GPU cluster to create state-of-the-art convolutional neural networks for environmental sound analysis. Our works were published in ICASSP 2017.</li></ul>	
	<b>Facebook</b> , Menlo Park, CA	
	<i>Software Engineering Intern</i>	<b>May 2015 – Aug 2015</b>

- Developed a distributed machine learning backend for large-scale logistic regression solved by LBFGS using Petuum parameter server.
- Benchmarked Petuum against Facebook's internal system and open source Vowpal Wabbit; showed that Petuum achieves high system throughput and produces comparable to better models.

**Google**, Pittsburgh, PA

*Software Engineering Intern*

**May 2013 – Aug 2013**

- Contributed to the Ad Quality backend; developed a hyperparameter tuning framework to optimize SmartAds training system with convex and non-convex optimization algorithms; built a web frontend for other teams to interface with the framework.

**LinkedIn**, Mountain View, CA

*Software Developer Intern*

**Jun 2012 – Aug 2012**

- Implemented several background tasks in the payment backend using Java, Oracle SQL, Python, and Spring Framework.

**OpenX**, Pasadena, CA

*Software Developer Intern*

**Apr 2012 – Jun 2012**

- Simulated a large number of users to load-test several internal servers using Erlang and Tsung; developed Tsung modules to enable Thrift protocols.

**Caltech Computer Science Department**, Pasadena, CA

*Research Assistant*

**Jun 2011 – Sept 2011**

- Contributed to the Community Seismic Network project which applies machine learning to detect earthquakes using smartphones.
- Applied *coreset* to training Gaussian mixture model using smartphone acceleration sensor data.

*Teaching Assistant*

**Mar 2011 – Mar 2012**

- Held weekly Office Hour and grade students' work for Computer Language Shop on C programming language.
- Translated and organized historical Chinese documents for research purposes.

#### SELECTED PUBLICATIONS

Z. Wang, N. Dong, **W. Dai**, S. D. Rosario, E. P. Xing

“Classification of Breast Cancer Histopathological Images using Convolutional Neural Networks with Hierarchical Loss and Global Pooling”

*To appear in International Conference on Image Analysis and Recognition (ICIAR)*, 2018. [Oral Presentation]

N. Dong, Z. Wang, **W. Dai**, E. P. Xing

“Unsupervised Domain Adaptation for Automatic Estimation of Cardiothoracic Ratio”

*To appear in Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2018.

H. Zhang, S. Xu, G. Neubig, **W. Dai**, Q. Ho, G. Yang, E. P. Xing

“Cavs: A Vertex-centric Programming Interface for Dynamic Neural Networks”

*To appear in Annual Technical Conference (ATC)*, 2018. [Oral Presentation]

H. Zhang, Z. Zheng, S. Xu, **W. Dai**, Q. Ho, X. Liang, Z. Hu, J. Wei, P. Xie, E. P. Xing

“Poseidon: An Efficient Communication Architecture for Distributed Deep Learning on GPU Clusters”

*Annual Technical Conference (ATC)*, 2017. [Oral Presentation]

X. Liang, L. Lee, **W. Dai**, E. P. Xing

- “Dual Motion GAN for Future-Flow Embedded Video Prediction”  
*International Conference on Computer Vision (ICCV)*, 2017.
- I. E.H. Yen, X. Huang, **W. Dai**, P. Ravikumar, I. Dhillon, E. P. Xing  
“PPDSparse: A Parallel Primal-Dual Sparse Method for Extreme Classification”  
*Knowledge Discovery and Data Mining (KDD)*, 2017.
- Y. Zhou, Y. Yu, **W. Dai**, Y. Liang, E. P. Xing  
“Distributed Proximal Gradient Algorithm for Partially Asynchronous Computer Clusters”  
*arXiv:1704.03540*, 2017.
- W. Dai**, J. Doyle, X. Liang, H. Zhang, N. Dong, Y. Li, E. P. Xing  
“SCAN: Structure Correcting Adversarial Network for Organ Segmentation in Chest X-rays”  
*arXiv:1703.08770*, 2017.
- W. Dai\***, C. Dai\*, S. Qu, J. Li, S. Das  
“Very Deep Convolutional Neural Networks for Raw Waveforms”  
*International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.
- J. Li, **W. Dai**, F. Metze, S. Qu, S. Das  
“A Comparison of Deep Learning Methods for Environmental Sound Detection”  
*International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.
- A. Harlap, H. Cui, **W. Dai**, J. Wei, G. R. Ganger, P. B. Gibbons, G. A. Gibson, E. P. Xing  
“Addressing the Straggler Problem for Iterative Convergent Parallel ML”  
*ACM Symposium on Cloud Computing (SoCC)*, 2016.
- Y. Wang, V. Sadhanala, **W. Dai**, W. Neiswanger, S. Sra, E. P. Xing  
“Parallel and Distributed Block-Coordinate Frank-Wolfe Algorithms”  
*International Conference of Machine Learning (ICML)*, 2016.
- J. K. Kim, Q. Ho, S. Lee, X. Zheng, **W. Dai**, G. Gibson, E. P. Xing.  
“STRADS: A Distributed Framework for Scheduled Model Parallel Machine Learning”  
*European Conference on Computer Systems (EuroSys)*, 2016.
- Y. Zhou, Y. Yu, **W. Dai**, Y. Liang, E. P. Xing  
“On Convergence of Model Parallel Proximal Gradient Algorithm for Stale Synchronous Parallel System”  
*Artificial Intelligence and Statistics (AISTATS)*, 2016.
- E. P. Xing, Q. Ho, P. Xie, **W. Dai**  
“Strategies and Principles of Distributed Machine Learning on Big Data”  
*Engineering, Volume:2, pp. 179 - 95*, 2016.
- J. Wei, **W. Dai**, A. Qiao, H. Cui, Q. Ho, G. R. Ganger, P. B. Gibbons, G. A. Gibson, E. P. Xing  
“Managed Communication and Consistency for Fast Data-Parallel Iterative Analytics”  
*ACM Symposium on Cloud Computing (SoCC)*, 2015. [**Best Paper**]
- E. P. Xing, Q. Ho, **W. Dai**, J. K. Kim, J. Wei, S. Lee, X. Zheng, P. Xie, A. Kumar, Y. Yu  
“Petuum: A New Platform for Distributed Machine Learning on Big Data”  
*Knowledge Discovery and Data Mining (KDD)*, 2015. [Oral Presentation]
- J. Yuan, F. Gao, Q. Ho, **W. Dai**, J. Wei, X. Zheng, E. P. Xing, T. Liu, and W. Ma  
“LightLDA: Big Topic Models on Modest Compute Cluster”  
*International World Wide Web Conference (WWW)*, 2015. [Oral Presentation]
- W. Dai**, A. Kumar, J. Wei, Q. Ho, G. Gibson, E. P. Xing  
“Analysis of High-Performance Distributed ML at Scale through Parameter Server Consistency Models”  
*AAAI Conference on Artificial Intelligence (AAAI)*, 2015. [Oral Presentation]

H. Cui, A. Tumanov, J. Wei, L. Xu , **W. Dai**, J. Haber-Kucharsky, Q. Ho, G. R. Ganger, P. B. Gibbons, G. A. Gibson, E. P. Xing  
 “Exploiting Iterative-ness for Parallel ML Computations”  
*Symposium on Cloud Computing (SoCC)*, 2014.

H. Cui, J. Cipar, Q. Ho, J. K. Kim, S. Lee, A. Kumar, J. Wei, **W. Dai**, G. R. Ganger, P. B. Gibbons, G. A. Gibson, E. P. Xing  
 “Exploiting Bounded Staleness to Speed Up Big Data Analytics”  
*Annual Technical Conference (ATC)*, 2014.

**W. Dai**, J. Wei, X. Zheng, J. K. Kim, S. Lee, J. Yin, Q. Ho, E. P. Xing  
 “Petuum: A Framework for Iterative-Convergent Distributed ML”  
*NIPS, Big Learning Workshop*, 2013.

**W. Dai**, S. K. Kumar, F. W. Starr  
 “Universal two-step crystallization of DNA-Functionalized Nanoparticles”  
*Soft Matter*, Vol. 6, pp. 6130-6135, 2010.

**W. Dai**, C. W. Hsu, F. Sciortino, F. W. Starr  
 “Valency Dependence of Polymorphism and Polyamorphism in DNA-Functionalized Nanoparticles”  
*Langmuir*, Vol. 26, pp. 3601-3608, 2010.

**W. Dai**  
 “Effect of Valency on the Dynamics and Thermodynamics of DNA-linked Nanoparticles Materials”  
*Bachelor of Arts Honor Thesis: Wesleyan University*, 2010.

#### INVITED AND CONTRIBUTED TALKS

*AAAI Conference on Artificial Intelligence*, January 2015. Title: Analysis of High-Performance Distributed ML at Scale through Parameter Server Consistency Models

*Carnegie Mellon Univ.*, 2015 Spring: Invited guest lecturer for 10-605 Machine Learning with Large Datasets on Parameter Server. Title: Parameter Server and Stuff that Makes Large-scale Machine Learning Work.

*California Institute of Technology*: Summer Undergraduate Research Fellowship Seminar Day, October 2011. Title: A Smartphone that Learns: Toward Adaptive Earthquake Detection on Smartphones. (Advanced to final round in Peripall Speaking Competition.)

*American Physical Society*, March 2010 in Seattle, USA. Title: Phase Behavior of DNA-Functionalized Nanoparticles: Dependence on Number and Orientation of Attached DNA strands.

#### AWARDS AND HONORS

Best Paper Award, ACM Symposium on Cloud Computing (SoCC), 2015.

High Honors from Wesleyan University Physics Department: Awarded for my undergraduate honor thesis work, 2010.

Phi Beta Kappa Admission: A selective academic honor society for distinguished students at the nation’s institutions of higher learning, 2010.

Student Prizes at Wesleyan University: Bertman Prize (Physics), Karl Van Dyke Prize (Physics), 2010; Sherman Prize (Mathematics), 2007.

Freeman Asian Scholarship: A four-year full scholarship awarded to two students per country from eleven Asian countries for outstanding scholastic and leadership achievements, 2007.

#### PROFESSIONAL SERVICE

Reviewer for IEEE Transaction on Medical Imaging, 2017, 2018.

Reviewer for Science Advance, 2017

Reviewer for PLOS ONE, 2017, 2018.

Reviewer for IEEE Transactions on Big Data, 2016.

Reviewer for AAAI Conference on Artificial Intelligence, 2016.

PROGRAMMING    C/C++, Python, Matlab, Java, Linux, L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub>