Pengtao Xie

Petuum Inc
2555 Smallman St. Suite 120
Pittsburgh, PA 15222

Tel: (412) 320-6230
Email: pengtaoxie2008@gmail.com
Web: http://www.cs.cmu.edu/~pengtaox

Research Interests

Diversity-promoting Machine Learning

Machine Learning for Healthcare

Natural Language Processing

Large-scale Distributed Machine Learning

Probabilistic Graphical Models

Deep Learning

Positions

Petuum Inc

Senior Director of Data Solutions and Service, Aug 2018 – Present

Director of Data Solutions and Service, Jan 2018 – Jul 2018

Research Scientist, Dec 2016 – Present

Senior Engineering Manager, Sept 2017 – Dec 2017

Engineering Manager, Dec 2016 – Aug 2017

Consultant, Jul 2016 – Nov 2016

Microsoft Research Redmond

Research Intern, May 2014 – Aug 2014

Microsoft Research Asia

Research Intern, Aug 2009 – Jun 2010

Education

Carnegie Mellon University, 2013 — 2018

Ph.D. in Machine Learning, M.S in Natural Language Processing (2015)

Advisor: Prof. Eric Xing

Tsinghua University, China, 2010 — 2013

M.E. in Computer Science

Advisor: Prof. Mingsheng Ying

Sichuan University, China, 2006 — 2010

B.E. in Computer Science

Selected Awards and Honors

Innovator Award, 2018 (presented by the Pittsburgh Business Times).

1st Place (out of 400+ participating teams) in both Defenses and Targeted Attacks, 3rd Place in Untargeted Attacks, in NIPS Adversarial Vision Challenge, 2018.

Siebel Scholarship, 2014 (85 graduate students from around the world).

National Scholarship of China, 2009.

National First Prize in China Undergraduate Mathematical Contest of Modeling, 2008.

Goldman Sachs Global Leader Scholarship, 2008 (150 undergraduate students from around the world).

Publications

Pengtao Xie, Wei Wu, Yichen Zhu, and Eric P. Xing. Orthogonality-promoting Distance Metric Learning: Convex Relaxation and Theoretical Analysis. *The 35th International Conference on Machine Learning (ICML), 2018.* **(Long Oral Presentation)**

Pengtao Xie, Hongbao Zhang, Yichen Zhu, and Eric P. Xing. Nonoverlap-promoting Variable Selection. *The 35th International Conference on Machine Learning (ICML), 2018.* **(Short Oral Presentation)**

Pengtao Xie, Haoran Shi, Ming Zhang, and Eric P. Xing. A Neural Architecture for Automated ICD Coding. *The 56th Annual Meeting of the Association for Computational Linguistics (ACL), 2018.* **(Oral Presentation)**

Baoyu Jing, **Pengtao Xie**, and Eric P. Xing. On the Automatic Generation of Medical Imaging Reports. *The 56th Annual Meeting of the Association for Computational Linguistics (ACL), 2018.*

Pengtao Xie, Jun Zhu, and Eric P. Xing. Diversity-promoting Bayesian Learning of Latent Variable Models. *To appear in the Journal of Machine Learning Research (JMLR), 2018.*

Pengtao Xie, Jin Kyu Kim, Qirong Ho, Yaoliang Yu, and Eric P. Xing. Orpheus: Efficient Distributed Machine Learning via System and Algorithm Co-design. *Symposium of Cloud Computing (SOCC), 2018.* **(Oral Presentation)**

Devendra Sachan, **Pengtao Xie**, and Eric P. Xing. Effective Use of Bidirectional Language Modeling for Medical Named Entity Recognition. *Conference on Machine Learning for Healthcare (MLHC), 2018.*

Zeya Wang, Nanqing Dong, Sean Rosario, Min Xu, **Pengtao Xie**, and Eric P. Xing. Ellipse Detection of Optic Disc-and-Cup Boundary in Fundus Image with Unsupervised Domain Adaption. *IEEE International Symposium on Biomedical Imaging (ISBI).*

Xiangan Liu, Keyang Xu, **Pengtao Xie**, and Eric P. Xing. Unsupervised Pseudo-labeling for Extractive Summarization on Electronic Health Records. *NIPS ML for Healthcare Workshop, 2018.* **(Spotlight Presentation)**

Pengtao Xie, Ruslan Salakhutdinov, Luntian Mou, and Eric P. Xing. Deep Conditional Determinantal Point Process for Large-scale Multi-label Classification. *International Conference on Computer Vision (ICCV), 2017.*

Pengtao Xie, Barnabas Poczos, and Eric P. Xing. Near-orthogonality Regularization in Kernel Methods. *The Conference on Uncertainty in Artificial Intelligence (UAI), 2017.* **(Plenary Oral Presentation)**

Pengtao Xie, Aarti Singh, and Eric P. Xing. Uncorrelation and Evenness: A New Diversity-Promoting Regularizer. *The 34th International Conference on Machine Learning (ICML), 2017.* **(Oral Presentation)**

Pengtao Xie, Yuntian Deng, Yi Zhou, Abhimanu Kumar, Yaoliang Yu, James Zou, and Eric P. Xing. Learning Latent Space Models with Angular Constraints. *The 34th International Conference on Machine Learning (ICML), 2017.* **(Oral Presentation)**

Jianxin Li, Haoyi Zhou, **Pengtao Xie**, and Yingchun Zhang. Improving the Generalization Performance of Multi-class SVM via Angular Regularization. *The 26th International Joint Conference on Artificial Intelligence (IJCAI), 2017.* **(Oral Presentation)**

Pengtao Xie and Eric P. Xing. A Constituent-centric Neural Architecture for Reading Comprehension. *The 55th Annual Meeting of the Association for Computational Linguistics (ACL), 2017.*

Hao Zhang, Zeyu Zheng, Shizhen Xu, Wei Dai, Qirong Ho, Xiaodan Liang, Zhiting Hu, Jinliang Wei, **Pengtao Xie**, and Eric P. Xing. Poseidon: An Efficient Communication Interface for Distributed Deep Learning on GPU Clusters. *USENIX Annual Technical Conference (ATC), 2017.* **(Oral Presentation)**

Ying Zhou, Xumin Ni, Kai Yuan, Yaoliang Yu, **Pengtao Xie**, Eric P. Xing, and Shuhua Xu. Inference of Multiple-Wave Population Admixture by Modeling Decay of Linkage Disequilibrium with Polynomial Functions. *Heredity, 2017.*

Eric P. Xing, Qirong Ho, **Pengtao Xie**, and Wei Dai. Strategies and Principles of Distributed Machine Learning on Big Data. *Transactions of Chinese Academy of Engineering, 2016.*

Pengtao Xie, Jun Zhu, and Eric P. Xing. Diversity-promoting Bayesian Learning of Latent Variable Models. *International Conference on Machine Learning (ICML), 2016.* **(Oral Presentation)**

Pengtao Xie, Jin Kyu Kim, Yi Zhou, Qirong Ho, Abhimanu Kumar, Yaoliang Yu, and Eric P. Xing. Lighter-communication Distributed Machine Learning via Sufficient Factor Broadcasting. *Conference on Uncertainty in Artificial Intelligence (UAI), 2016.*

Eric P. Xing, Qirong Ho, Wei Dai, Jin Kyu Kim, Jinliang Wei, Seunghak Lee, Xun Zheng, **Pengtao Xie**, Abhimanu Kumar, and Yaoliang Yu. Petuum: A New Platform for Distributed Machine Learning on Big Data. *IEEE Transactions on Big Data, 2015.*

Pengtao Xie. Learning Compact and Effective Distance Metrics with Diversity Regularization. *European Conference on Machine Learning (ECML), 2015.* **(Oral Presentation)**

Pengtao Xie, Yuntian Deng, and Eric Xing. Diversifying Restricted Boltzmann Machine for Document Modeling. *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2015.* **(Oral Presentation)**

Eric P. Xing, Qirong Ho, Wei Dai, Jin Kyu Kim, Jinliang Wei, Seunghak Lee, Xun Zheng, **Pengtao Xie**, Abhimanu Kumar, and Yaoliang Yu. Petuum: A New Platform for Distributed Machine Learning on Big Data. *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2015.* **(Oral Presentation)**

Pengtao Xie, Diyi Yang, and Eric P. Xing. Incorporating Word Correlation Knowledge into Topic Modeling. *Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2015.*

Pengtao Xie, Yulong Pei, Yuan Xie, and Eric P. Xing. Mining User Interests from Personal Photos. *The 29th AAAI Conference on Artificial Intelligence (AAAI), 2015.*

Pengtao Xie and Eric P. Xing. Integrating Image Clustering and Codebook Learning. *The 29th AAAI Conference on Artificial Intelligence (AAAI), 2015.* **(Oral Presentation)**

Pengtao Xie and Eric P. Xing. Integrating Document Clustering and Topic Modeling. *The 29th International Conference on Uncertainty in Artificial Intelligence (UAI), 2013.*

Pengtao Xie and Eric P. Xing. Multi-modal Distance Metric Learning. *The 23rd International Joint Conference on Artificial Intelligence (IJCAI), 2013.* **(Oral Presentation)**

Teaching

Guest Lecturer at Carnegie Mellon University

10708, Probabilistic Graphical Models. *Instructor*: Prof. Eric Xing, Spring 2017.

Teaching Assistants at Carnegie Mellon University

10708, Probabilistic Graphical Models. *Instructor*: Prof. Eric Xing, Spring 2015.

10701, Machine Learning. *Instructors*: Prof. Barnabas Poczos and Prof. Aarti Singh, Spring 2014.

10601, Machine Learning. *Instructors*: Prof. William Cohen and Prof. Eric Xing, Fall 2013.

Advising

Ph.D. Students: Seojin Bang (CMU, 2018), Hongyang Zhang (CMU, 2018), Susu Xu (CMU, 2018), Biwei Huang (CMU, 2018), Yaodong Yu (University of Virginia, 2018), Yuan Xie (Indiana University Bloomington, 2015)

Master Students: Devendra Sachan (CMU, 2017), Hongbao Zhang (CMU, 2017), Hakim Sidahmed (CMU, 2015), Yuntian Deng (CMU, 2015)

Undergraduate Students: Alex Tomala (University of Waterloo, 2018), Darshan Patil (CMU, 2017), Rishub Jain (CMU, 2017), Haoran Shi (Peking University, 2017), Yichen Zhu (Peking University, 2017), Binghong Chen (Tsinghua University, 2016)

External Grants and Awards

I wrote the following funded proposals together with Professor Eric Xing.

NSF IIS1617583: "III: RI: Small: A New Approach to Latent Space Learning with Diversity-inducing Mutual Angular Regularization, with Applications to Healthcare Data Analytics", 09/01/2016 - 8/31/2019, $499,361.

NSF CCF1629559: "XPS: FULL: Broad-purpose, Aggressively Asynchronous and Theoretically Sound Parallel Large-scale Machine Learning", 09/01/2016 - 8/31/2020, $625,379.

Under review by NIH: "A Study of Machine Learning Systems for Medical Decision-making Support Based on Electronic Health Records".

Patents

Gilad-Bachrach Ran, Thomas W. Finley, Mikhail Bilenko, and **Pengtao Xie**. Neural Networks for Encrypted Data. *US Patent, NO.US9946970B2, 2018.*

Pengtao Xie and Eric Xing. An Efficient Peer-to-peer Architecture for Distributed Machine Learning. *US Patent, NO.US20180302498A1, 2018.*

Pengtao Xie and Eric Xing. A Constituent Centric Architecture for Reading Comprehension. *US Patent, NO.US20180300314A1, 2018.*

Pengtao Xie and Eric Xing. A Machine Learning System for Disease, Patient, Drug Co-embedding, and Multi-drug Recommendation. *US Patent, NO.US20180330808A1, 2018.*

Pengtao Xie and Eric Xing. A Machine Learning System for Measuring Patient Similarity. *Filed, 2018.*

Pengtao Xie and Eric Xing. Systems and Methods for Medical Topic Discovery Based on Large-scale Machine Learning. *Filed, 2018.*

Pengtao Xie and Eric Xing. Systems and Methods for Automatically Tagging Concepts to, and Generating Text Reports for, Medical Images Based on Machine Learning. *Filed, 2018.*

Pengtao Xie and Eric Xing. Systems and Methods for Automatically Generating International Classification of Disease Codes for Patients Based on Machine Learning. *Filed, 2018.*

Professional Service

Reviewer for International Conference on Machine Learning (ICML), 2014, 2018-2019.

Reviewer for Neural Information Processing Systems (NIPS), 2016, 2018

Program Committee Member for International Conference on Artificial Intelligence and Statistics (AISTATS), 2017-2019

Program Committee Member for Conference on Uncertainty in Artificial Intelligence (UAI), 2018

Reviewer for International Conference on Learning Representations (ICLR), 2019

Program Committee Member for AAAI Conference on Artificial Intelligence (AAAI), 2019

Program Committee Member for IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016-2019

Reviewer for International Conference on Computer Vision (ICCV), 2015, 2017

Reviewer for European Conference on Computer Vision (ECCV), 2016, 2018

Program Committee Member for Annual Meeting of the Association for Computational Linguistics (ACL), 2015-2018

Reviewer for Conference on Empirical Methods in Natural Language Processing (EMNLP), 2015

Reviewer for ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2015

Reviewer for IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2018

Reviewer for IEEE Transactions on Knowledge and Data Engineering (TKDE), 2015-2018

Reviewer for IEEE Transactions on Multimedia (TMM), 2016-2017

Reviewer for PLOS ONE, 2017-2018

Reviewer for IEEE Transactions on Neural Networks and Learning Systems (TNNLS), 2015-2016, 2018

Reviewer for Journal of the American Statistical Association (JASA), 2015

Program Committee Member for European Conference on Machine Learning (ECML), 2016-2017

Program Committee Member for Asian Conference on Computer Vision (ACCV), 2016

Reviewer for British Machine Vision Conference (BMVC), 2017

Membership: ACM, IEEE.

Talks

Symposium of Cloud Computing, Carlsbad, Oct 2018. Orpheus: Efficient Distributed Machine Learning via System and Algorithm Co-design.

The 35th International Conference on Machine Learning, Stockholm, Jul 2018. Orthogonality-promoting Distance Metric Learning: Convex Relaxation and Theoretical Analysis.

The 35th International Conference on Machine Learning, Stockholm, Jul 2018. Nonoverlap-promoting Variable Selection.

The 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Jul 2018. A Neural Architecture for Automated ICD Coding.

The 34th International Conference on Machine Learning, Sydney, Aug 2017. Uncorrelation and Evenness: A New Diversity-promoting Regularizer.

The 34th International Conference on Machine Learning, Sydney, Aug 2017. Learning Latent Space Models with Angular Constraints.

Conference on Uncertainty in Artificial Intelligence, Sydney, Aug 2017. Near-orthogonality Regularization in Kernel Methods.

The 33rd International Conference on Machine Learning, New York, Jun 2016. Diversity-promoting Bayesian Learning of Latent Variable Models.

Database Seminar, Carnegie Mellon University, Mar 2016. Sufficient Factor Broadcasting for Distributed Machine Learning.

The 11th CSL Student Conference, University of Illinois Urbana-Champaign, Feb 2016. Latent Variable Modeling with Diversity-inducing Mutual Angular Regularization.

Artificial Intelligence Seminar, Carnegie Mellon University, Feb 2016. Diversity-inducing Learning of Latent Variable Models: Frequentist and Bayesian Perspectives.

VALSE Webinar, Oct 2015. Diversity Regularization of Latent Variable Models: Theory, Algorithm, and Applications.

Machine Learning Lunch Seminar, Carnegie Mellon University, Sept 2015. Mutual Angular Regularization of Latent Variable Models: Theory, Algorithm, and Applications.

Beihang University, Aug 2015. Diversifying Restricted Boltzmann Machine for Document Modeling.

ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Sydney, Aug 2015. Diversifying Restricted Boltzmann Machine for Document Modeling.

VASC Seminar, Carnegie Mellon University, April 2015. Integrating Image Clustering and Codebook Learning.

Machine Learning Lunch Seminar, Carnegie Mellon University, April 2015. Integrating Data Clustering and Representation Learning.

SCS Student Seminar, Carnegie Mellon University, April 2015. Incorporating Word Correlation Knowledge into Topic Modeling.

CL+NLP Seminar, Carnegie Mellon University, Apr 2015. Incorporating Word Correlation Knowledge into Topic Modeling.

Database Seminar, Carnegie Mellon University, Mar 2015. Mining User Interests from Personal Photos.

The 29th AAAI Conference on Artificial Intelligence, Austin, Jan 2015. Integrating Image Clustering and Codebook Learning.

Database Seminar, Carnegie Mellon University, Sep 2014. CryptGraph: Privacy-preserving Graph Analytics on Encrypted Graph.

Machine Learning Lunch Seminar, Carnegie Mellon University, Sep 2014. Privacy-preserving Neural Network Prediction on Encrypted Data.

Cylab Student Seminar, Carnegie Mellon University, Sep 2014. Privacy-preserving Neural Network Prediction on Encrypted Data.

SDI/ISTC Seminar, Carnegie Mellon University and Intel, Sep 2014. Privacy-preserving Neural Network Prediction on Encrypted Data.

Cloud Machine Learning Team and Machine Learning Group, Microsoft Research, Aug 2014. Privacy-preserving Neural Network Prediction on Encrypted Data.

Cryptography Group, Microsoft Research, Jul 2014. Privacy-preserving Neural Network Prediction on Encrypted Data.

The 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013), Beijing, Aug 2013. Multi-Modal Distance Metric Learning.