

Pengtao Xie

Jacobs School of Engineering
University of California San Diego
9500 Gilman Dr, La Jolla, CA 92093

Tel: (412) 320-6230
Email: plxie@eng.ucsd.edu
Web: pengtaoxie.github.io/

Research Interests

Machine learning inspired by humans' learning skills, such as learning by passing tests, interleaving learning, etc.

Machine learning for healthcare

Positions

University of California San Diego

Assistant Professor in the Electrical and Computer Engineering Department, July 2020 –

Petuum Inc

Associate Vice President, July 2019 – April 2020

Senior Director of Data Solutions and Service, Aug 2018 – Jun 2019

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 the Machine Learning Department, School of Computer Science

M.S in Natural Language Processing (2015)

Advisor: Prof. Eric Xing

Tsinghua University, China, 2010 — 2013

M.E. in Computer Science Department

Advisor: Prof. Mingsheng Ying

Sichuan University, China, 2006 — 2010

B.E. in School of Computer Science

Selected Awards and Honors

Finalist (top 5) for AMIA Doctoral Dissertation Award, 2020.

Amazon AWS Research Award, 2020.

Tencent AI-Lab Faculty Award, 2020.

Tencent WeChat Faculty Award, 2020.

Google Cloud Research Credits, 2020.

Top Reviewer for ICML-2020, 2020.

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

Jiaqi Zeng and **Pengtao Xie**. Contrastive Self-supervised Learning for Graph Representation Learning. *AAAI Conference on Artificial Intelligence (AAAI)*, 2021.

Seojin Bang, **Pengtao Xie**, Heewook Lee, Wei Wu, Eric Xing. Explaining Black-box Models Using A Deep Variational Information Bottleneck Approach. *AAAI Conference on Artificial Intelligence (AAAI)*, 2021.

G. Zeng, W. Yang, Z. Ju, Y. Yang, S. Wang, R. Zhang, M. Zhou, J. Zeng, X. Dong, R. Zhang, H. Fang, P. Zhu, S. Chen and **Pengtao Xie**. MedDialog: Large-scale Medical Dialogue Datasets. *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020.

Congzheng Song, Shanghang Zhang, Najmeh Sadoughi, **Pengtao Xie**, Eric Xing. Generalized Zero-shot ICD Coding. *International Joint Conference on Artificial Intelligence (IJCAI)*, 2020.

Zeya Wang, Baoyu Jing, Yang Ni, Nanqing Dong, **Pengtao Xie**, Eric P Xing. Adversarial Domain Adaptation Being Aware of Class Relationships. *European Conference on Artificial Intelligence (ECAI)*, 2020.

Biwei Huang, Kun Zhang, **Pengtao Xie**, Mingming Gong, Eric Xing, and Clark Glymour. Specific and Shared Causal Relation Modeling and Mechanism-based Clustering. *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

Keyang Xu, Mike Lam, Jingzhi Pang, Xin Gao, Charlotte Band, Piyush Mathur, Frank Papay, Ashish Khanna, Jacek Cywinski, Kamal Maheshwari, **Pengtao Xie**, and Eric Xing. Multimodal Machine Learning for Automated ICD Coding. *Conference on Machine Learning for Healthcare (MLHC)*, 2019.

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. *The IEEE International Symposium on Biomedical Imaging (ISBI)*, 2019.

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)*.

Xiangnan 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

Lecturer at UC San Diego

ECE285 Deep Generative Models, Winter 2021.

ECE269 Linear Algebra, Winter 2021.

Guest Lecturer at Carnegie Mellon University

10708, Probabilistic Graphical Models. *Instructor*: Prof. Eric Xing, Spring 2017, Spring 2020.

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.

External Grants and Awards

Amazon AWS Research Award, 2020

Tencent AI-Lab Faculty Award, 2020

Tencent WeChat Faculty Award, 2020

Google Cloud Research Credits, 2020

Patents

Gilad-Bachrach Ran, Thomas W. Finley, Mikhail Bilenko, and **Pengtao Xie**. Neural networks for encrypted data. *US Patent 9946970*, 2018.

Pengtao Xie and Eric Xing. Constituent centric architecture for reading comprehension. *US Patent 10706234*, 2020.

Pengtao Xie and Eric Xing. Efficient Peer-to-Peer Architecture for Distributed Machine Learning. *US Patent App. 15849663*, 2018.

Pengtao Xie and Eric Xing. Machine learning system for disease, patient, and drug co-embedding, and multi-drug recommendation. *US Patent App. 15946482*, 2018.

Pengtao Xie and Eric Xing. Machine Learning System for Patient Similarity. *US Patent App. 16038895*, 2019

Pengtao Xie and Eric Xing. Systems and Methods for Medical Topic Discovery Based on Large-Scale Machine Learning. *US Patent App. 16207115*, 2020.

Pengtao Xie and Eric Xing. Systems and Methods for Automatically Tagging Concepts to, and Generating Text Reports for, Medical Images Based On Machine Learning. *US Patent App. 16207117*, 2020

Pengtao Xie and Eric Xing. Systems and Methods for Automatically Generating International Classification of Diseases Codes for a Patient Based on Machine Learning. *US Patent App. 16207119*, 2020.

Pengtao Xie and Eric Xing. Systems and Methods for Predicting Medications to Prescribe to a Patient Based on Machine Learning. *US Patent App. 16207114*, 2020.

Advising

PhD

Ramtin Hosseini (ECE, 2020 -)

Yueyu Jiang (ECE, 2020 -)

Sai Somayajula (ECE, 2020 -)

Master

Arda Bati (MS, ECE, 2020 -)

Xuehai He (MS, ECE, 2019 -)

Jiachen Li (ECE, 2020 -),

Yuxiao Liang (ECE, 2020)

Kai Wang (ECE, 2020)

Kai Wang (CSE, 2020)

Xingyi Yang (ECE, 2019 -)

Yi Yu (ECE, 2020)

Shengzhe Zhang (ECE, 2020)

Jinyu Zhao (ECE, 2020)

Undergraduate

Yuhong Chen (Math, 2020)

Rongrong Liu (Cognitive Science, 2020)

Gaotong Wu (ECE, 2020)

Jiawen Zeng (ECE, 2020)

Professional Service

Co-organizers

Co-organizer for NeurIPS 2020 Workshop on “Self-Supervised Learning: Theory and Practice”

Co-organizer for AAAI 2021 Workshop on “Trustworthy AI for Healthcare”

Co-organizer for ICLR 2021 workshop on “Machine Learning for Preventing and Combating Pandemics”

Area Chairs

Area Chair for ICML 2021

Area Chair for CVPR 2021

Area Chair for ICCV 2021

Area Chair for NAACL 2021

Area Chair for AAAI 2021

Area Chair for IJCAI 2021

Program Committee Members and Reviewers

Reviewer for Nature Machine Intelligence, 2020

Reviewer for International Conference on Machine Learning (ICML), 2014, 2018-2020

Reviewer for Neural Information Processing Systems (NIPS), 2016, 2018-2019

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, 2021

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, 2020

Reviewer for IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2020-2021

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

Reviewer for Management Science, 2019

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

Member

Machine Learning for Signal Processing Technical Committee

Association for Computing Machinery

Institute of Electrical and Electronics Engineers

American Medical Informatics Association

University Service

Member of PhD Admission Committee, ECE, 2020

Member of Diversity & Inclusion Committee, ECE, 2020

Faculty Advisor for Summer Research Internship Program (SRIP), ECE, 2020-2021

Member of PhD Preliminary Exam Committee, ECE, 2020

Moderator for Student Research Conference, UCSD, 2020

Faculty Advisor for Undergraduate Summer Research, Division of Physical Sciences, UCSD, 2020

Interviewer for Faculty Recruiting, ECE, 2020

Faculty member of the Seventh College, UCSD, 2020 -

Selected Talks

Machine Learning for Medical Decision Support

Mar 2021, University of Massachusetts Amherst

Mar 2020, Allen Institute for Artificial Intelligence

Jan 2020, Language Technologies Institute, Carnegie Mellon University

Nov 2019, Department of Biomedical Informatics, University of Pittsburgh

Oct 2019, AI NEXTCon Developer Conference

Apr 2019, New York University

Apr 2019, University of Massachusetts Amherst

Mar 2019, University of California San Diego

Feb 2019, Columbia University

Feb 2019, Johns Hopkins University

Feb 2019, University of California Los Angeles

Feb 2019, RISELab, University of California, Berkeley

Jan 2019, University of Wisconsin-Madison

Diversity-promoting and Large-scale Distributed Learning for Healthcare

Nov 2018, University of California San Diego

Nov 2018, University of California, Los Angeles

Nov 2018, The University of Texas at Austin

Nov 2018, Georgia Institute of Technology

Oct 2018, University of Illinois Urbana-Champaign

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.