Pengzhi Gao

CONTACT INFORMATION	Petuum, Inc. 2555 Smallman Street, Suite 120, Pittsburgh, Pennsylvania 15222	*Mobile:* 215-696-0238 *Email:* gpengzhi@gmail.com *Homepage:* https://gpengzhi.github.io/

EDUCATION

Rensselaer Polytechnic Institute, Troy, NY

Ph.D., Electrical Engineering, August 2013 - December 2017

- Advisor: Professor Meng Wang
- Thesis: High-dimensional Data Analysis by Exploiting Low-dimensional Models with Applications in Synchrophasor Data Analysis in Power Systems

University of Pennsylvania, Philadelphia, PA

M.S., Electrical Engineering, August 2011 - May 2013

Xidian University, China

B.S. (with honors), Electronic and Information Engineering, August 2007 - May 2011

WORK EXPERIENCE

Data Scientist   February 2018 to present
Core Machine Learning Team,
Petuum, Inc.
Supervisor: Dr. Tong Wen and Zhiting Hu

- Designed and implemented the machine learning library (based on TensorFlow, DyNet, and LightGBM) for Petuum AI Builder Platform.
- Designed and developed Texar-PyTorch (https://github.com/asyml/texar-pytorch, gaining over 520 stars), an open-source machine learning and text generation toolkit based on PyTorch.
- Maintained and contributed to Texar-TensorFlow (https://github.com/asyml/texar, gaining over 1820 stars), an open-source machine learning and text generation toolkit based on TensorFlow.
- Designed and developed Forte (https://github.com/asyml/forte), a toolkit for building natural language processing pipelines, featuring cross-task interaction, adaptable data-model interfaces and many more.

Research Intern   December 2010 to May 2011
Internet Media Group,
Microsoft Research Asia
Supervisor: Dr. Feng Wu and Dr. Chong Luo

- Analyzed the data collected from 54 sensors deployed in Intel Berkeley Research Lab (150 MB of data) to exploit the temporal correlations in sensor readings. Developed a joint source network coding scheme for approximate data gathering in wireless sensor network.

SKILL SETS

- Proficiency with MATLAB, Python, Dynet, PyTorch, and TensorFlow
- Experienced in Java, R, C/C++, C#, AMPL

HONORS AND AWARDS

- North America Finalist of IBM Watson Build Challenge	2017
- Paper selected as the runner-up of the Best Paper in Electric Energy Systems Track of Hawaii International Conference on System Sciences	2015
- Founders Award of Excellence (top 1%)	2015
- Paper selected as one of the Best Conference Papers on Power System Analysis and Modeling of IEEE Power & Energy Society General Meeting	2014
- Excellent Graduate of Xidian University (top 1%)	2011
- National Scholarship (top 1%)	2010
- First prize of the College Academic and Technological Scholarship (top 2%)	2008-2010
- Excellent Student Awards (top 1%)	2008

	1. **P. Gao**, R. Wang, and M. Wang. "Robust Matrix Completion by Exploiting Dynamic Low-dimensional Structures." *submitted to IEEE Transactions on Signal Processing*, 2019.

Journal Publications

1. **P. Gao**, R. Wang, and M. Wang. "Robust Matrix Completion by Exploiting Dynamic Low-dimensional Structures." *submitted to IEEE Transactions on Signal Processing*, 2019.

2. **P. Gao**, R. Wang, M. Wang, and J. H. Chow. "Low-rank Matrix Recovery from Noisy, Quantized and Erroneous Measurements." *IEEE Transactions on Signal Processing*, 2018, 66 (11): 2918-2932.

3. **P. Gao**, M. Wang, J. H. Chow, M. Berger, and L. M. Seversky. "Missing Data Recovery for High-dimensional Signals with Nonlinear Low-dimensional Structures." *IEEE Transactions on Signal Processing*, 2017, 65 (20): 5421-5436.

4. **P. Gao**, M. Wang, J. H. Chow, S. G. Ghiocel, B. Fardanesh, G. Stefopoulos, and M. P. Razanousky. "Identification of Successive "Unobservable" Cyber Data Attacks in Power Systems Through Matrix Decomposition." *IEEE Transactions on Signal Processing*, 2016, 64 (21): 5557-5570.

5. **P. Gao**, M. Wang, S. G. Ghiocel, J. H. Chow, B. Fardanesh, and G. Stefopoulos. "Missing Data Recovery by Exploiting Low-dimensionality in Power System Synchrophasor Measurements." *IEEE Transactions on Power Systems*, 2016, 31 (2): 1006-1013.

Conference Publications

1. Z. Liu, A. Bukkittu, M. Gupta, **P. Gao**, S. Singhavi, A. Ahmed, W. Wei, Z. Hu, H. Shi, E. Xing and Z. Hu. "Forte: Composing Diverse NLP tools For Text Retrieval, Analysis and Generation." *submitted to 2020 Annual Conference of the Association for Computational Linguistics (ACL)*, 2020.

2. M. Wang, J. H. Chow, Y. Hao, S. Zhang, W. Li, R. Wang, **P. Gao**, C. Lackner, E. Farantatos, and M. Patel. "A Low-rank Framework of PMU Data Recovery and Event Identification." *Proc. of the First IEEE International Conference on Smart Grid Synchronized Measurements and Analytics (SGSMA)*, College Station, Texas, May, 2019.

3. G. Mijolla, S. Konstantinouplos, **P. Gao**, J. H. Chow, and M. Wang. "An Evaluation of Low-Rank Matrix Completion Algorithms for Synchrophasor Missing Data Recovery." *Proc. of the 20th Power Systems Computation Conference (PSCC)*, Dublin, Ireland, Jun. 2018.

4. **P. Gao**, and M. Wang. "Dynamic Matrix Recovery from Partially Observed and Erroneous Measurements." *Proc. of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, Canada, Apr. 2018.

5. M. Wang, J. H. Chow, **P. Gao**, Y. Hao, W. Li, and R. Wang. "Recent Results of PMU Data Analytics by Exploiting Low-dimensional Structures." *Proc. of the 10th Bulk Power Systems Dynamics and Control Symposium (IREP)*, Espinho, Portugal, Aug. 2017.

6. **P. Gao**, R. Wang, and M. Wang. "Quantized Low-rank Matrix Recovery with Erroneous Measurements: Application to Data Privacy in Power Grids." *Proc. of Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2016.

7. **P. Gao**, M. Wang, and J. H. Chow. "Matrix Completion with Columns in Union and Sums of Subspaces." *Proc. of IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Orlando, FL, Dec. 2015.

8. M. Wang, J. H. Chow, **P. Gao**, X. T. Jiang, Y. Xia, S. G. Ghiocel, B. Fardanesh, G. Stefopoulos, Y. Kokai, N. Saito, and M. P. Razanousky. "A Low-Rank Matrix approach for the Analysis of Large Amounts of Synchrophasor Data." *Proc. of Hawaii International Conference on System Sciences (Runner-up of Best Paper in Electric Energy Systems Track)*, Kauai, Hawaii, Jan. 2015.

9. M. Wang, **P. Gao**, S. G. Ghiocel, J. H. Chow, B. Fardanesh, G. Stefopoulos, and M. P. Razanousky. "Identification of "Unobservable" Cyber Data Attacks on Power Grids." *Proc. of IEEE SmartGridComm*, Venice, Italy, Nov. 2014.

10. **P. Gao**, M. Wang, S. G. Ghiocel, and J. H. Chow. "Modeless Reconstruction of Missing Synchrophasor Measurements." *Proc. of IEEE Power & Energy Society General Meeting (selected in Best Conference Paper sessions)*, Washington, DC, Jul. 2014.

TECHNICAL REPORTS	1. Z. Hu, **P. Gao**, A. Bukkittu, and Z. Hu. "Introducing Texar-PyTorch: An ML Library integrating the best of TensorFlow into PyTorch." Octorber, 2019.

PATENTS	1. M. Wang, **P. Gao**, and J. H. Chow. "A low-rank-based missing PMU data recovery method." Application No.: 62/445305, Filed January 12, 2017.

PROJECTS

DyNet: The Dynamic Neural Network Toolkit
Core Machine Learning Team,
Petuum, Inc.

- DyNet is a neural network library developed by Carnegie Mellon University, Petuum, and many others. It is written in C++ (with bindings in Python) and is designed to be efficient when run on either CPU or GPU, and to work well with networks that have dynamic structures that change for every training instance. I contributed the example and tutorial part of DyNet repository.

IBM Watson Build Challenge 2017
ECSE Department,
Rensselaer Polytechnic Institute

- We developed a web application with IBM Watson APIs based on IBM Bluemix. We implemented multiple missing data recovery algorithms in Python for large-scale Phasor Measurement Unit (PMU) data analysis.

Online Algorithm for PMU Data Processing (OLAP)
ECSE Department,
Rensselaer Polytechnic Institute

- We implemented OLAP by C# based on Project Alpha for the real-time application. Project Alpha is the elite version of Open PDC. The code developed on Project Alpha can be run on Open PDC as an action adapter.

Mobile Eye Gaze Estimation with Deep Learning
ECSE Department,
Rensselaer Polytechnic Institute

- We implemented a deep convolutional neural network based on TensorFlow for eye gaze estimation. The original dataset comes from the GazeCapture project. Our model follows the architecture introduced in CVPR paper "Eye Tracking for Everyone", and we changed and tuned the hyper parameters due to the different image size in our training dataset.

PROFESSIONAL ACTIVITIES & SERVICE

- Student Member of IEEE, 2013 - 2017. Member of IEEE, 2018 - present.
- RPI Student Representative at the Center for Ultra-wide-area Resilient Electric Energy Transmission Networks (CURENT), 2015 - 2016.
- Teaching Assistant (Rensselaer Polytechnic Institute):
  Modeling and Analysis of Uncertainty, Fall 2017,
  Distributed Systems and Sensor Networks, Fall 2017.
- Program Committee Member:
  Conference on Uncertainty in Artificial Intelligence (UAI) 2018.
- Reviewer:
  IEEE Transactions on Smart Grid,
  IEEE Transactions on Automatic Control,
  IEEE/ACM Transactions on Networking,
  IEEE Signal Processing Letters,
  Annals of Mathematics and Artificial Intelligence,
  American Control Conference,
  IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm),
  International Symposium on Antennas and Propagation.