Jinmian Ye

Master Student
School of Computer Science and Engineering
University of Electronic Science and Technology of China (**UESTC**)

Email: jinmian.y@gmail.com
Phone: (86) 177 0811 2046
Github: https://github.com/ay27

EDUCATION & INTERNSHIP

Intern (Feb. 2018 - Aug. 2018)   Advisor: Chen Qiang, Shuicheng Yan
- AI Institute, **360 Inc.**

M.S. (Sept. 2016 - Jun. 2019)   Advisor: Zenglin Xu
- Computer Science and Technology, **UESTC**

Academic Visiting (2017 Summer)   Advisor: Irwin King, Michael R. Lyu
- Department of Computer Science and Engineering , *The Chinese University of Hong Kong* (**CUHK**)

B.S. (Sept. 2012 - Jun. 2016)
- Computer Science and Engineering, **UESTC**
- *Thesis*: Solving ADMM-based distributed matrix factorization on Parameter Server

RESEARCH EXPERIENCES

Compressing Embedding Layer (Feb. 2018 - Now)
- I'm trying to optimize and compress the huge embedding layer in natural language processing area, achieving hundreds of times the compression ratio while maintaining the model performance.

Neural Networks Model Pruning (May 2017 - Nov. 2017)
- Using Block-Term tensor decomposition to prune the fully-connected input-to-hidden weight matrix in RNN, reducing thousands time of learnable parameters while increasing up to 15.6% accuracy comparing to vanilla LSTM.

Deep Learning Framework Development (Jan. 2017 - Sept. 2017)
- A dynamic GPU memory management runtime that enables training super deep and wide models. Some intuitive rules such as liveness analysis, re-computation, etc, are proposed to enhance memory utilization efficiency.

RESEARCH INTERESTS

- Model Pruning and Accelerating in Neural Networks
- Parallel and High Performance Computing
- Tensor Decomposition and Compression

SKILLS

- Proficient in Python, C/C++, Java
- Knowledgeable in TensorFlow, PyTorch, Caffe, etc.

PUBLICATIONS

1. Zhonghui You, **Jinmian Ye**, Kunming Li, Ping Wang. Adversarial Noise Layer: Regularize Neural Network By Adding Noise. (arXiv preprint)
2. **Jinmian Ye**, Linnan Wang, Guangxi Li, Di Chen, Shandian Zhe, Xinqi Chu, Zenglin Xu. Learning Compact Recurrent Neural Networks with Block-Term Tensor Decomposition. (CVPR-2018)
3. Guangxi Li, **Jinmian Ye**, Haiqin Yang, Di Chen, Zenglin Xu: BT-Nets: Simplifying Deep Neural Networks via Block Term Decomposition. (arXiv preprint)
4. Linnan Wang, **Jinmian Ye**, Yiyang Zhao, Wei WU, Ang Li, Shuaiwen Leon Song, Zenglin Xu, Tim Kraska. SuperNeurons:Dynamic GPU Memory Management for Training Deep Nonlinear Neural Networks. (PPoPP-2018)