

# Jianfei Chen

---

CONTACT INFORMATION 1-508, FIT Building Phone: +86-185-1831-6949  
Tsinghua University, Beijing, China, 100084 E-mail: chenjian14@mails.tsinghua.edu.cn  
<http://ml.cs.tsinghua.edu.cn/~jianfei> <http://github.com/cjf00000>

RESEARCH My current research interest is *large scale machine learning*, particularly for deep generative models and topic models, by optimizing the algorithm and system jointly. I am currently working on stochastic training of graph convolutional networks, and fast GPU dense-dense-sparse matrix multiplication for matrix factorization models. Previously, I focused on combining distributed computing, stochastic algorithms and heterogeneous computing techniques to build fast and scalable learning systems for topic models, handling hundreds of billions of tokens and millions of topics.

## Selected Publications (full list is at the back):

**Jianfei Chen**, Chongxuan Li, Yizhong Ru and Jun Zhu. Population Matching Discrepancy and Applications in Deep Learning. To Appear in *Neural Information Processing Systems (NIPS)*, 2017.

**Jianfei Chen**, Jun Zhu, Jie Lu, and Shixia Liu. Scalable Inference for Nested Chinese Restaurant Process Topic Models. arXiv 1702.07083.

Kaiwei Li, **Jianfei Chen**, Wenguang Chen, and Jun Zhu. SaberLDA: Sparsity-Aware Learning of Topic Models on GPUs. *Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, 2017.

**Jianfei Chen**, Kaiwei Li, Jun Zhu, and Wenguang Chen. WarpLDA: a Cache Efficient O(1) Algorithm for Latent Dirichlet Allocation. *Very Large Data Bases (VLDB)*, 2016.

**Jianfei Chen**, Jun Zhu, Zi Wang, Xun Zheng, and Bo Zhang. Scalable Inference for Logistic-Normal Topic Models, *Advances in Neural Information Processing Systems (NIPS)*, 2013.

EDUCATION **Tsinghua University**, Beijing, China  
*PhD Candidate* (Computer Science) **September 2014 – July 2019 (expected)**  
• Advisor: Jun Zhu

**Tsinghua University**, Beijing, China  
*Bachelor of Engineering* (Computer Science) **August 2010 – July 2014**  
• Outstanding graduates of Tsinghua University and Beijing city;  
• GPA: 89.3/100, ranked 14th / 129 in Tsinghua Computer Science Department

VISITING AND INDUSTRIAL EXPERIENCE University of Oxford, with Yee Whye Teh **Mar. 2018 – Oct. 2018 (scheduled)**  
*Research on interpretable and scalable deep topic models.*

Big Data Lab, Baidu Inc., with Tong Zhang **Feb. 2014 – July 2014**  
*Designed a hierarchical noisy-or model for feature extraction that resembles to Google's Phil.*

Stanford U., with Andrew Ng **July 2013 – Sep. 2013**  
*Worked on a state-of-the-art deep speech recognition system, optimized the GPU code.*

SELECTED AWARDS

- **Programming Contents**
  - TopCoder SRM rating 2591 (top 0.3%), **8th** among 1000+ competitors in TopCoder High-School contest, 2009.
  - Gold medal (rank 16th) in National Olympiad in Informatics, China, 2009
  - Ranked **5th** in Baidu Astar 2010, **Top 500** in Google Code Jam 2010 / 2011.
  - Champion in Tsinghua AI competition, 2010.
- **Scholarship and Miscellaneous**
  - National Scholarship (offered for top 3% students in Tsinghua University)
  - Tsinghua Scholarship for distinguished freshman, Shengda Scholarship for 2 times, Technology Innovation Scholarship for 3 times.

SKILLS

- Language: English (fluent, TOEFL/iBT score 105), Chinese (native)
- Language: C/C++ (proficient), Python (Tensorflow), MatLab, Bash, L<sup>A</sup>T<sub>E</sub>X, GoLang, C#, Scala

## Full list of publications

### Peer reviewed conference papers

- **Jianfei Chen**, Chongxuan Li, Yizhong Ru and Jun Zhu. Population Matching Discrepancy and Applications in Deep Learning. To Appear in *Neural Information Processing Systems (NIPS)*, 2017.
- Kaiwei Li, **Jianfei Chen**, Wenguang Chen, and Jun Zhu. SaberLDA: Sparsity-Aware Learning of Topic Models on GPUs. *Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, 2017.
- **Jianfei Chen**, Kaiwei Li, Jun Zhu, and Wenguang Chen. WarpLDA: a Cache Efficient O(1) Algorithm for Latent Dirichlet Allocation. *Very Large Data Bases (VLDB)*, 2016.
- Arnab Bhadury, **Jianfei Chen**, Jun Zhu, and Shixia Liu. Scaling up Dynamic Topic Models. *World Wide Web Conference (WWW)*, 2016.
- Chengtao Li, Jun Zhu, and **Jianfei Chen**. Bayesian Max-Margin Multitask Learning with Data Augmentation. *International Conference on Machine Learning (ICML)*, 2014.
- Ning Chen, Jun Zhu, **Jianfei Chen** and Bo Zhang. Dropout Training for Support Vector Machines. *Association for the Advancement of Artificial Intelligence (AAAI)*, 2014.
- Shixia Liu, Xiting Wang, **Jianfei Chen**, Jun Zhu, and Baining Guo. TopicPanorama: a Full Picture of Relevant Topics. *IEEE Visualization (IEEE Vis)*, 2014.
- **Jianfei Chen**, Jun Zhu, Zi Wang, Xun Zheng, and Bo Zhang. Scalable Inference for Logistic-Normal Topic Models, *Advances in Neural Information Processing Systems (NIPS)*, 2013.

### Journal papers

- Jun Zhu, **Jianfei Chen**, and Wenbo Hu. Big Learning with Bayesian Methods. *National Science Review*, in press.
- Xiting Wang, Shixia Liu, Junlin Liu, **Jianfei Chen**, Jun Zhu, and Baining Guo. TopicPanorama: A Full Picture of Relevant Topics. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 2016.

### Technical reports

- **Jianfei Chen**, Jun Zhu, Jie Lu, and Shixia Liu. Scalable Inference for Nested Chinese Restaurant Process Topic Models. arXiv 1702.07083.
- Ning Chen, Jun Zhu, **Jianfei Chen**, and Ting Chen. Dropout Training for SVMs with Data Augmentation. arXiv 1508.02268.

### Services

- Reviewer for IJCAI (2017), ICML (2017/2015), NIPS (2017);
- Web master for ICML 2014;
- Student volunteer for NIPS 2013.

### TA experience

- TA for Tsinghua-Duke Kunshan University Machine Learning Summer school, 2016.
- TA for Statistical Machine Learning, Tsinghua University, 2015-2016.

### Software

- WarpLDA: cache efficient implementation for Latent Dirichlet Allocation;
- ScaCTM: Scalable inference for Correlated Topic Model;
- ScaNCRP: Scalable inference for nested Chinese Restaurant Process Topic Models;
- BigTopicModel: A general purpose framework for topic models. Easy, Efficient and Elastic.