Yizhe Zhang

Information Initiative at Duke (iiD), Gross Hall,
140 Science Dr. Durham, NC 27708, U.S.A.
Email: yz196@cs.duke.edu
Web: http://people.duke.edu/~yz196/
+1 9195368858

Personal Summary

A Ph.D. student with expertise on a wide range of computational biology, machine learning and Bayesian statistics topics. Able of implementing complicated models with various programming languages.

Education

Duke University, Durham, NC, *Ph.D. in Computational Biology and Bioinformatics* *— 2013 - present*
Advisor: Dr. Alexander Hartemink and Dr. Lawerence Carin.

Nanjing University, Nanjing, China, *B.Sc. in Physics* *— 2007 - 2011*
Department for Intensive Instruction, Kuang Yaming Honors school

Research Interest

- **Markov Chain Monte Carlo:** *Scalable and parallel MCMC; Hamiltonian Monte Carlo; the interplay between optimization and MCMC methodology.*
- **Probabilistic graphic models**: *Dynamic Bayesian networks, conditional random fields, hidden Markov models.*
- **Scalable Bayesian inference:** *Bayesian inference methods for deep models, recurrent neural networks, and Bayesian factor models.*

Research Experience

Electronical and Computer Engineering, Duke University *— 2015 - present*

- **Monomial Gamma Hamilton Monte Carlo.**
  Demonstrate the equivalence of HMC and slice sampler, the resulting method have theoretical advantage over standard HMC.
- **Non-parametric Bayesian factor model with efficient Gaussian process.**
  Use GP to characterize spatial dependency. Applications to image denoising, inpainting and depth channel reconstruction demonstrate performance improvements over other methods.
- **Deep dynamic Poisson factor analysis for large scale medical records data.**
  Scalable Bayesian topic modeling for large scale time series data. By taking advantage of parallel computation and data-augmentation strategy, it is fast and accurate.

Computational Biology and Bioinformatics, Duke University *—2013 - 2015*

- **Causal inference and dynamic Bayesian network modeling for gene regulatory network .**
- **Coupled sequence regression and annotation inference using Viterbi path integration.**
- **Enhancer-promoter association study using multilingual Relational Topic Modeling.**

Shanghai Center for Bioinformation Technology, Shanghai, China *—2011 - 2013*

- **MOST+: A fast and accurate motif discovering algorithm using suffix tree.**
- **CTF: An algorithms predicting protein binding sites using Conditional Random Field.**
- **Genome-wide association study of gut microbiome in Hepatitis B patients.**

PUBLICATIONS

CONFERENCES

- **Yizhe Zhang**, Ricardo Henao, Chunyuan Li, Lawrence Carin. Learning Spatial Dependent Dictionary with Efficient Multicative Gaussian Process. *submitted (2015)*.
- **Yizhe Zhang**, Ricardo Henao, Jianling Zhong, Lawrence Carin, Alexander Hartemink. Learning a Hybrid Architecture for Sequence Regression and Annotation. *to appear on AAAI (2016)*.
- Kai Fan, **Yizhe Zhang**, Lawrence Carin, Katherine Heller. Stochastic Gradient Langevin Dynamics for Noisy Variational Auto-Encoder. *submitted* (2015).
- **Yizhe Zhang**, Lawrence Carin. Learning Dictionary with Spatial and Inter-dictionary Dependency. *NIPS workshop* (2015).

JOURNALS

- **Yizhe Zhang**, Yupeng He and Chaochun Wei (2015). MOST+: a Motif Finding Approach Combining Genomic Sequence and Heterogeneous Genome-wide Signatures. *BMC Genomics*.
- Yupeng He, **Yizhe Zhang**, Guangyong Zheng and Chaochun Wei (2012). CRF-based Transcription Factor Binding Site Finding System. *BMC Genomics*.
- Jiemeng Liu, Haifeng Wang, Hongxing Yang, **Yizhe Zhang**, Jinfeng Wang, Fangqing Zhao and Ji Qi. (2012). Composition-based Classification of Short Metagenomic Sequences Elucidates the Landscapes of Taxonomic and Functional Enrichment of Microorganisms. *Nucleic Acids Research*.

COURSES

- Statistics: Advanced statistical computing (STA863), Generalize linear model (STA841), Bayesian statistics (STA601), Statistical inference (STA732) , Advanced Machine Learning (STA571).
- Computer science: Probabilistic graphical models (CS590), Computational systems biology (CS662).
- Biology: Molecular Biology (BIO201), Computational sequence biology (CBB561).
- Teaching experience: Advanced Machine Learning (STA571).

REWARDS

- Travel award for NIPS —*2015*
- DataFest (Stat@Duke). Best use of data reward —*2014*
- Travel scholarship for ICIBM —*2014*
- Department two-years fellowship —*2013*
- National Excellent Graduate Scholarship（top 1%） —*2012*

SKILL AND PROFICIENCIES

- Language: Mandarine, English
- Programming: C/C++, Python, Java, MATLAB and R.
- Misc. : Latex, SAS, GIT, CUDA, Object-C and SQL