# Zhenwei Dai

🛛 +1 (832) 329 1018 🔹 🖂 zd11@rice.edu 🔹 🚱 www.github.com/DAIZHENWEI **S**Google Scholar

## Education

#### **Rice University**

Ph.D. in Statistics, GPA: 3.93/4.00

o Advisor: Dr. Anshumali Shrivastava and Dr. Reinhard Heckel

### The Chinese University of Hong Kong

M.Phil. in Bioinformatics

#### The Chinese University of Hong Kong

B.S. in Statistics (First Honor)

- Hong Kong, China Asia-Pacific Economic Cooperation Scholarship (2016)
- Hong Kong Special Administrative Region Government Scholarship (2015)
- Advantage Trust Statistics Scholarship (2015)
- Dean's List, The Chinese University of Hong Kong (2015, 2016)
- Shaw College Head's List Honor, The Chinese University of Hong Kong (2014, 2015, 2016)

## **Professional Skills**

Languages: Python, R, C/C++, MATLAB

Software: Pytorch, LATEX, Ubuntu

Reseach: Machine Learning, Probabilistic Algorithm, Optimization, Bayesian Statistics, Multivariate Analysis

## **Professional Experience**

#### TOTAL E&P Research & Technology USA, LLC

Machine Learning Research Intern

- Develop a reinforcement learning model to optimize the quality and synthesizability of TOTAL antioxidant molecules
- Propose a multi-task reinforcement learning structure to speed up the training of our molecule optimization model, which reduces the number of training episodes by over 80%
- Build parallel computing framework to improve the scalability of the multi-task reinforcement learning models

#### Department of Statistics, Rice University

#### Research Assistant

- Propose new algorithms and corresponding data structures for large scale machine learning and statistical inference
- o Develop theorems to characterize the optimization process in deep learning

# Conferences

- o Active Sampling Count Sketch (ASCS) for Online Sparse Estimation of a Trillion Scale Covariance Matrix, ACM SIGMOD 2021
- o Adaptive learned Bloom filter (Ada-BF): Efficient Utilization of the Classifier, NeurIPS 2020
- o Channel Normalization in Convolutional Neural Network avoids Vanishing Gradients, ICML 2019 Workshop Deep Phenomena
- o Batch effects correction for microbiome data with Dirichlet-multinomial regression, ISMB 2018 (Oral)

Houston, TX Aug 2018 - Present

Hong Kong, China Aug 2016 - July 2018

Hong Kong, China Sept 2012 - June 2016

Oct 2020 – April 2021, Houston, TX

Aug 2018 – Present, Houston, TX

## **Machine Learning Projects**

Large Scale Inference: Online Sparse Estimation of a Trillion Scale Covariance Matrix 2020 Spring & Summer

- Propose an online and one-pass sketching algorithm, ASCS, to estimate and infer the large entries of an ultra-high dimensional covariance matrix using sublinear memory cost
- o Perform experiments to evaluate our algorithm on large scale web and DNA datasets
- o ASCS achieves over 1 million times memory reduction compared to the brute force method

Probabilistic Algorithm: Adaptive Learned Bloom Filter for Membership Test 2019 Spring & Summer

- Propose two adaptive learned Bloom filters by utilizing the full spectrum of learner's prediction score to enhance the performance of Bloom filters
- Perform exhaustive experiments on large scale web datasets to evaluate the performance and robustness of our algorithms, where our algorithms reduce the FPR by over 80% compared with the state-of-art methods

#### Deep Learning: Batch Normalization in Convolutional Neural Network (CNN) 2019 Spring & Summer

- o Develope theorems to explain the vanishing gradients during the training of deep CNN
- o Formally explain why batch normalization controls the gradients scale at initialization for deep CNN

Bayesian Statistics: Batch effects correction for microbiome data 2017 Summer – 2018 Fall

- Propose a Bayesian variable selection method to correct batch effects and detect microbial taxa associated with the input variable
- o Develope a hybrid MCMC algorithm to estimate the posterior distribution of parameters

## **Publications**

- Zhenwei Dai, Aditya Desai, Reinhard Heckel, and Anshumali Shrivastava. Active Sampling Count Sketch (ASCS) for Online Sparse Estimation of a Trillion Scale Covariance Matrix. In *Proceedings of the 2021 ACM SIGMOD International Conference on Management of Data*, 2021.
- [2] Zhenwei Dai and Anshumali Shrivastava. Adaptive Learned Bloom Filter (Ada-BF): Efficient Utilization of the Classifier with Application to Real-Time Information Filtering on the Web. In Advances in Neural Information Processing Systems, volume 33, pages 11700–11710, 2020.
- [3] **Zhenwei Dai** and Reinhard Heckel. Channel Normalization in Convolutional Neural Network Avoids Vanishing Gradients. *ICML Workshop Deep Phenomena*, 2019.
- [4] **Zhenwei Dai**, Sunny H Wong, Jun Yu, and Yingying Wei. Batch Effects Correction for Microbiome Data with Dirichlet-multinomial Regression. *Bioinformatics*, 35(5):807–814, 2018.
- [5] Zhenwei Dai, Olabisi O Coker, Geicho Nakatsu, William KK Wu, Liuyang Zhao, Zigui Chen, Francis KL Chan, Karsten Kristiansen, Joseph JY Sung, Sunny Hei Wong, et al. Multi-cohort Analysis of Colorectal Cancer Metagenome Identified Altered Bacteria Across Populations and Universal Bacterial Markers. *Microbiome*, 6(1):70, 2018.
- [6] Olabisi O Coker\*, Zhenwei Dai\*, Yongzhan Nie, Guijun Zhao, Lei Cao, Geicho Nakatsu, William KK Wu, Sunny Hei Wong, Zigui Chen, Joseph JY Sung, et al. Mucosal Microbiome Dysbiosis in Gastric Carcinogenesis. *Gut*, 67(6):1024– 1032, 2018.
- [7] Jeffery Ho\*, **Zhenwei Dai**\*, Thomas NY Kwong, et al. Disease burden of clostridium difficile infections in adults, Hong Kong, China, 2006–2014. *Emerging infectious diseases*, 23(10):1671, 2017.
- [8] Geicho Nakatsu, Haokui Zhou, William Ka Kei Wu, Sunny Hei Wong, Olabisi O Coker, **Zhenwei Dai**, et al. Alterations in enteric virome are associated with colorectal cancer and survival outcomes. *Gastroenterology*, 155(2):529–541, 2018.
- [9] Thomas NY Kwong, Xiansong Wang, Geicho Nakatsu, Tai Cheong Chow, Timothy Tipoe, **Zhenwei Dai**, et al. Association between bacteremia from specific microbes and subsequent diagnosis of colorectal cancer. *Gastroenterology*, 155(2):383–390, 2018.