Zhaonan Sun

Yorktown Heights, NY 10598
www.linkedin.com/in/zhaonansun
Email: zhaonan.sun@gmail.com
Phone: 765-464-4751

PROFESSIONAL EXPERIENCES

Associate Director, Biogen Aug 2021 - present

Research Staff Member, IBM T. J. Watson Research Center Aug 2015 - Aug 2021

- Team lead for IBM-CHDI collaboration on Huntington's Disease progression modeling
- Co-Chair, IBM Health Informatics Professional Interest Community
- Global Technical Lead for Disease Progression Modeling

Postdoctoral Researcher, IBM T. J. Watson Research Center Aug 2014 - Jul 2015

- Developed statistical and machine learning algorithms for risk prediction and drug effect analysis

EDUCATION

Ph.D in Statistics, Purdue University 2008 - 2014

M. Sc. in Statistics, Renmin University of China 2006 - 2008

B. Sc. in Statistics, Renmin University of China 2002 - 2006
Magna Cum Laude with Honor

PUBLICATIONS

Disease Progression Modeling

1. Bum Chul Kwon, Vibha Anand, Kristen A Severson, Soumya Ghosh, **Zhaonan Sun**, Brigitte I Frohnert, Markus Lundgren, Kenney Ng. (2020) DPVis: Visual analytics with hidden markov models for disease progression pathways. *IEEE transactions on visualization and computer graphics*.
2. **Zhaonan Sun**, Soumya Ghosh, Ying Li, Yu Cheng, Amrita Mohan, Cristina Sampaio, Jianying Hu. (2019) A probabilistic disease progression modeling approach and its application to integrated Huntington's disease observational data. *Journal of American Medical Informatics Association Open*. 1(1), 123-130.
3. Zach Shahn, Ying Li, **Zhaonan Sun**, Amrita Mohan, Cristina Sampaio, Jianying Hu. (2019) G-Computation and Hierarchical Models for Estimating Multiple Causal Effects From Observational Disease Registries With Irregular Visits. *AMIA Summits on Translational Science Proceedings*. 789.
4. **Zhaonan Sun**, Ying Li, Soumya Ghosh, Yu Cheng, Amrita Mohan, Cristina Sampaio, Jianying Hu. (2017) A Data-Driven Method for Generating Robust Symptom Onset

Indicators in Huntington's Disease Registry Data. *American Medical Informatics Association Annual Symposium(AMIA)*. 1635-1644.

Risk Prediction

1. Bin Liu, Ying Li, Soumya Ghosh, **Zhaonan Sun**, Kenney Ng, Jianying Hu. (2019) Complication risk profiling in diabetes care: A bayesian multi-task and feature relationship learning approach. *IEEE Transactions on Knowledge and Data Engineering*. 32 (7), 1276-1289.
2. Sandra Liu, Jie Chen, **Zhaonan Sun**, Yu Zhu. (2018) From good to great: nonlinear improvement of healthcare service. *International Journal of Pharmaceutical and Healthcare Marketing*. 12(4), 391-408.
3. Jaehee Shim, **Zhaonan Sun**, Amos Cahan. (2018) Patient specific Vancomycin Dose Recommendation with Baseline Information at the Time of the First Dose. *Journal of Pharmacokinetics and Pharmacodynamics*. 45, S48-S49.
4. Bin Liu, Ying Li, **Zhaonan Sun**, Soumya Ghosh, Kenney Ng. (2018) Early Prediction of Diabetes Complications from Electronic Health Records: A Multi-task Survival Analysis Approach. *AAAI Conference on Artificial Intelligence(AAAI)*. 32 (1).
5. Xiang Li, **Zhaonan Sun**, Xin Du, Haifeng Liu, Gang Hu, Guotong Xie. (2017) Bootstrap-based Feature Selection to Balance Model Discrimination and Predictor Significance: A Study of Stroke Prediction in Atrial Fibrillation. *American Medical Informatics Association Annual Symposium(AMIA)*. 1130-1139.
6. **Zhaonan Sun**, Fei Wang, and Jianying Hu. (2015): LINKAGE: An Approach for Comprehensive Risk Prediction for Care Management. *21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*. Pages 1145-1154.

Computational Phenotyping

1. Zhengping Che, Yu Cheng, Shuangfei Zhai, **Zhaonan Sun**, Yan Liu. (2017) Boosting Deep Learning Risk Prediction with Generative Adversarial Networks for Electronic Health Records. *The IEEE International Conference on Data Mining(ICDM)*. 787-792.
2. Soumya Ghosh, Yu Cheng, **Zhaonan Sun**. (2017) Deep State Space Models for Computational Phenotyping. *IEEE International Conference on Health Informatics*. 399-402.

Drug Effect Analysis

1. Ying Li, Ping Zhang, **Zhaonan Sun**, and Jianying Hu. (2016) Data-Driven Prediction of Beneficial Drug Combinations in Spontaneous Reporting Systems. *American Medical Informatics Association Annual Symposium (AMIA)*. 808-817.
2. Ping Zhang, **Zhaonan Sun**, Fei Wang, and Jianying Hu. (2015). Towards computational drug repositioning: a comparative study of single-task and multi-task learning. *American Medical Informatics Association Annual Symposium (AMIA)*. Abstract.

Others Topics

1. Ben Li, **Zhaonan Sun**, Qing He, Yu Zhu, and Zhaohui Qin. (2016): Bayesian inference with historical data-based informative priors improves detection of differentially expressed genes. *Bioinformatics.* 32 (5): 682-689.

2. Jiang, Y., Frankenberger, J. R., Bowling, L. C., and **Sun, Z.** (2014). Quantification of Uncertainty in Estimated Nitrate-N Loads in Agricultural Watersheds. *Journal of Hydrology*. 519:106-116.

3. **Zhaonan Sun**, Thomas Kuczek, and Yu Zhu. (2014). Statistical calibration for qRT-PCR, microarray and RNA-Seq expression data with measurement error models. *The Annals of Applied Statistics*. 8(2):1022-1044.

4. **Zhaonan Sun**, Han Wu, Zhaohui Qin, and Yu Zhu. (2013). Model-Based Methods for Transcript Expression Level Quantification in RNA-Seq in *Advances in Statistical Bioinformatics: Models and Integrative Inference for High-Throughput Data*, edited by Do, K-A., Qin, S. and Vannucci, M. Cambridge University Press.

5. **Zhaonan Sun** and Yu Zhu. (2012). Systematic Comparison of RNA-Seq Normalization Methods Using Measurement Error Models. *Bioinformatics*. 28:20. Pages 2584-2591.

6. S.V.N. Vishwanathan, **Zhaonan Sun.**, Theera-Ampornpunt, N. and Varma, M. (2010). Multiple Kernel Learning and the SMO Algorithm. *NIPS*. Pages 3311-3325.

7. Xing Wang, Xi Wang and **Zhaonan Sun** (2009). Comparison on confidence bands of decision boundary between SVM and Logistic Regression. *Proceedings of fifth international joint conference on INC, IMS and IDC*.

TALKS

1. *16th Annual HD Therapeutics Conference*, Palm Spring, CA, 2021. DSI - A Disease Status Index for Huntington's Disease.
2. *IBM Research Got Science Seminar*, Yorktown Heights, NY, 2020. Disease Progression Modeling.
3. *14th Annual HD Therapeutics Conference*,Palm Spring, CA, 2019. Multi-modal HD progression model with clinical and morphometric data.
4. *Colloquium Seminars at Biostatistics Department of Columbia University*, New York, NY, 2019. Disease Progression Modeling with Large-Scale Observational Data in Huntington's Disease.
5. *Colloquium Seminars at Biostatistics Department of Columbia University*, New York, NY, 2019. Disease Progression Modeling with Large-Scale Observational Data in Huntington's Disease.
6. *4th International Conference on Big Data and Information Analytics*, Houston, TX, 2018. Disease Progression Modeling with Large-Scale Observational Data.
7. *International Chinese Statistical Association Applied Statistics Symposium*, New Brunswick, NJ, 2018. Disease Progression Modeling with Large-Scale Observational Data.
8. *1st Enroll-HD Congress*, Quebec City, Quebec, 2018. Understanding Huntington's Disease Progression: A Probabilistic Modeling Approach.
9. *American Medical Informatics Association Annual Symposium*, Washington, DC, 2017.A Method for Generating Robust Symptom Onset Indicators in Huntington's Disease Registry Data.

10. *New England Statistics Symposium*, Storrs, CT, 2017.Exploiting Convolutional Neural Network for Risk Prediction with Medical Feature Embedding.
11. *Joint Summits of American Medical Informatics Association*, San Francisco, CA, 2017.Exploring Factors that Contribute to Missing Values in Observational Huntington's Disease Study Data.
12. *The Second Statistical Forum on Huntington's Disease*, Princeton, NJ, 2016. Machine learning for disease progression models.
13. *American Medical Informatics Association Annual Symposium*, San Francisco, CA, 2015. A graph based methodology for temporal signature identification from EHR.
14. *Joint Statistical Meetings*, Seattle, WA, 2015. Comprehensive Risk Prediction Using Interactive Graph-Guided Fussed Lasso Penalty.
15. *IBM Research Health Informatics PICs*, Yorktown Heights, NY, 2015. Multi-task learning approach for comprehensive risk prediction.
16. *Eastern North American Region Meetings*, Baltimore, MD, 2014. Statistical calibration of qRT-PCR, microarray and RNA-Seq gene expression data with measurement error models.
17. *Purdue Bioinformatics Seminar*, West Lafayette, IN, 2014. Statistical calibration of high-throughput gene expression data using measurement error models.
18. *Joint Statistical Meetings*, San Diego, CA, 2012. Differential gene expression pattern analysis using exon-level RNA-Seq data.
19. *Joint Statistical Meetings*, Miami, FL, 2011. An integrative approach to comparing and normalizing gene expression data generated from RNA-Seq, Microarray and RT-PCR technologies.

PROGRAM COMMITTEE

1. International Conference on Machine Learning (ICML) 2021
2. International Joint Conference on Artificial Intelligence(IJCAI) 2021
3. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) 2021
4. NIPS Machine Learning for Health (ML4H) workshop 2019
5. International Joint Conference on Artificial Intelligence(IJCAI) 2019
6. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) 2019
7. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) 2018
8. NIPS Machine Learning for Health (ML4H) workshop 2017
9. IEEE International Conference on Healthcare Informatics 2016
10. International Joint Conference on Artificial Intelligence(IJCAI) 2016
11. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) 2016
12. International Joint Conference on Artificial Intelligence 2015
13. IEEE International Conference on Healthcare Informatics 2015
14. KDD 2015 Workshop on "BigCHat: Connected Health at Big Data"
15. 1st Workshop on Matrix Computations for Biomedical Informatics 2015