Zhaonan Sun

1101 Route 134 Kitchawan Rd
Yorktown Heights, NY 10598.

Email: zsun@us.ibm.com
Phone: 765-464-4751

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

1. Zhang, P., **Sun, Z.**, Wang, F., and Hu, J. (2015). Towards computational drug repositioning: a comparative study of single-task and multi-task learning. *American Medical Informatics Association Annual Symposium (AMIA)*. To Appear.
2. **Sun, Z.**, Wang, F., and Hu, J. (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.
3. 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.
4. **Sun, Z.**, Kuczek, T. and Zhu, Y. (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.
5. **Sun, Z.**, Wu, H., Qin, Z. and Zhu, Y. (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.
6. **Sun, Z.** and Zhu, Y. (2012). Systematic Comparison of RNA-Seq Normalization Methods Using Measurement Error Models. *Bioinformatics*. 28:20. Pages 2584-2591.
7. Vishwanathan, S.V.N., **Sun, Z.**, Theera-Ampornpunt, N. and Varma, M. (2010). Multiple Kernel Learning and the SMO Algorithm. *NIPS*. Pages 3311-3325.
8. Wang, X., Wang, X. and **Sun, Z.** (2009). Comparison on confidence bands of decision boundary between SVM and Logistic Regression. *Proceedings of fifth international joint conference on INC, IMS and IDC*.

PROFESSIONAL EXPERIENCE

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

- Developing disease progression model with EHR systems

- Developing machine learning method for drug repositioning

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

- Developed large-scale risk screening method for comprehensive geriatric care
- Developed multi-task learning method for evidence boost in drug repositioning

Statistical Consultant, Purdue University Aug 2013 - May 2014

- Provided data analysis and statistical software support for Purdue community.

Research Assistant, Purdue University June 2011 - July 2013

- Developed statistical method for assessing and improving measurement errors in RNA-Seq gene expression data
- Developed statistical method for identify differential gene expression using RNA-Seq data
- Involved in extending SMO algorithm in multiple kernal learning

TALKS

1. *Joint Statistical Meetings*, Seattle, WA, 2015. Comprehensive Risk Prediction Using Interactive Graph-Guided Fussed Lasso Penalty.
2. *IBM Research Health Informatics PICs*, Yorktown Heights, NY, 2015. Multi-task learning approach for comprehensive risk prediction.
3. *Eastern North American Region Meetings*, Baltimore, MD, 2014. Statistical calibration of qRT-PCR, microarray and RNA-Seq gene expression data with measurement error models.
4. *Purdue Bioinformatics Seminar*, West Lafayette, IN, 2014. Statistical calibration of high-throughput gene expression data using measurement error models.
5. *Joint Statistical Meetings*, San Diego, CA, 2012. Differential gene expression pattern analysis using exon-level RNA-Seq data.
6. *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 Joint Conference on Artificial Intelligence 2015
2. IEEE International Conference on Healthcare Informatics 2015
3. KDD 2015 Workshop on ”BigCHat: Connected Health at Big Data”
4. 1st Workshop on Matrix Computations for Biomedical Informatics 2015

PROGRAMMING SKILLS

python, R, SQL, SAS, MATLAB, LaTeX