LUWAN ZHANG

Postdoctoral Research Fellow, Harvard School of Public Health

EDUCATION

- **University of Wisconsin-Madison**
  Ph.D. in Statistics; GPA 4; Advisor: Professor Ming Yuan   *2017*
  M.S. in Computer Sciences; GPA 3.95   *2015*
- **Nankai University**
  B.S. in Mathematics; GPA 93/100 (Top 3 out of 50)   *2012*

RESEARCH INTERESTS

- Clustering, network analysis, unsupervised learning methods
- Consensus learning by combining multiple data sources
- Methods for low-dimensional embeddings, kernel methods
- Applications in Electronic Health Records, general biomedical and biological data

COMPUTER SKILLS

R, Python, SQL, MATLAB, Java, HTML

RESEARCH EXPERIENCES

Research Fellow at Harvard School of Public Health   *07/2017 – present*

- Conduct methodological research driven by Electronic Health Records (EHR) applications; faculty mentor: Professor Tianxi Cai.
- Developed a novel clustering algorithm to automate clinical codes grouping highly in line with human annotations.
- Built a consensus graph, by combining multiple data sources, in discovery of disease-symptom-treatment network; have successfully applied to identify coronary artery disease network.
- Lead phenotypic efforts in Phenowide Association Studies (PheWAS) and phenotyping prediction.
- Collaborative research on natural language processing (NLP) and word embeddings (`word2vec`) with Harvard Medical School.
- Work closely with Partners HealthCare, Brigham and Women's Hospital, VA Boston.
- Help mentor Ph.D. and Master students; coordinate various research projects; manage lab meetings for research updates and journal club.

Research Assistannt at University of Wisconsin-Madison   *07/2014 – 05/2017*

- Developed a graphical model based unsupervised ensemble learning method; established theoretical guarantees for high-dimensional settings; significantly boosted the predictive performance on Rheumatoid Arthritis on AUC from 80% to 93%.
- Developed an efficient algorithm in pursuit of a shrinkage Euclidean distance matrix estimator; successfully visualized the diversity of Vpu protein sequences in a high-profile HIV-1 study; substantially reduced errors in molecular 3D structure reconstruction from classical multidimensional scaling approach.
- Collaborated with two Computer Science professors on finding a near-optimal design matrix in Danzig Selector with applications in Reinforcement Learning.

INDUSTRIAL EXPERIENCE

Data Science Internship at IBM Commerce   *06/2015 – 09/2015*

- Led an anomaly detection project on B2B file transfer in collaboration with two product teams.

- Constucted a feed forward neural network (FFNN) with three hidden layers to predict daily file transfer volume.
- Built a gradient boosting decision tree to detect transaction failure and risk factors.
- Presented as team representative to IBM Distinguished Engineers
- Awarded the exceptional summer intern of the year.

TEACHING EXPERIENCES

Guest Lecturer at Harvard University   *Fall 2017*

- Gave special lectures on recent developments in clustering to graduate students in Biostatistics.

Instructor at University of Wisconsin-Madison   *Spring 2017*

- Gave weekly lectures to a class of 100+ students majoring in engineering.
- Improved in-class engagement for low-vision or hearing-impaired students by arranging notetakers.

Teaching Assistant at University of Wisconsin-Madison   *Fall 2012, Sping 2013, Fall 2016*

- Led two 75-min discussion sessions per week; held office hours; homework and exam grading.
- Ranked at the top 3% in TA evaluation.

PUBLIC SPEAKING EXPERIENCES

- 33rd New England Statistics Symposium: Statistical Data Science, University of Connecticut. *05/2019*
- IMS-China International Conference on Statistics and Probability, Naning, China.   *07/2017*
- Annual Retreat, the Center for Predictive Computational Phenotyping (CPCP), University of Wisconsin-Madison.   *06/2015, 06/2017*
- Annual Board Meeting, Morgridge Institute for Research, University of Wisconsin-Madison.   *06/2015*

RESEARCH PAPERS

- **L. Zhang**, G. Wahba, and M. Yuan. Distance shrinkage and euclidean embedding via regularized kernel estimation. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 78(4):849867, 2016.
- B. Liu, **L. Zhang**, and J. Liu. Dantzig Selector with an Approximately Optimal Denoising Matrix and its Application in Sparse Reinforcement Learning, Proceedings of Uncertainty in Artificial Intelligence,487-496, 2016.
- **L. Zhang** and T. A. Cai. Unsupervised Ensemble Learning via Ising Model Approximation with Application to Phenotyping Prediction, *Submitted*.
- **L. Zhang**, K. Liao, I. Kohane, and T. Cai. Multi-view Banded Spectral Clustering with Application to ICD9 Clustering, *Submitted*.
- **L. Zhang**, Y. Zhang, T. Cai, Y. Ahuja, Y.-L. Ho, A. Beam, K. Cho, I. Kohane, K. Liao, and T. Cai. Automated Data-driven Grouping on ICD Codes with Multi-view Banded Spectral Clustering, *Submitted*.
- T. A. Cai, T. Cai, N. Yang, **L. Zhang**, K. K. Kumamaru, F. J. Rybicki, and K. P. Liao. EXTraction of EMR Numerical Data: A Novel Tool to EXTEND Clinical Research Using Automated Numerical Data Collection, *Submitted*.