Lu Tang

(*Updated on December 17, 2017*)

University of Michigan
Department of Biostatistics
1415 Washington Heights
Ann Arbor, MI 48109

E-Mail: lutang@umich.edu
Phone: 434-466-6323
Website: `www-personal.umich.edu/~lutang`

Education

2013 - Present	University of Michigan Ph.D. Biostatistics (Advisor: Peter X.K. Song)	Ann Arbor, MI
2012 - 2013	University of Virginia M.S. Statistics	Charlottesville, VA
2010 - 2012	University of Virginia B.A. Mathematics, minor in Computer Science	Charlottesville, VA
2008 - 2010	Sun Yat-sen University Major in Information and Computational Science	Guangzhou, China

Research Interests

Methodology: Data integration and harmonization, high-dimensional statistical inference, regression parameter clustering, statistical computing, optimization

Applications: Metabolomics, environmental health, epigenetics, bioinformatics, children's health, statistical quality control

Experience

Professional Experience

2014 - Present — Department of Biostatistics, University of Michigan — Ann Arbor, MI
Doctoral Dissertation Research

- Dissertation research: developed statistical learning methods that enable to perform clustering and estimation of regression coefficients for integration of heterogeneous data; developed divide-and-combine methods in the Hadoop paradigm to fit regularized generalized linear models for Big Data.
- Created and maintained open-source software packages for the developed methods resulted from the dissertation research, including an R package `metafuse` and Python MapReduce function `modac`.

2013 - Present — Children's Environmental Health and Disease Prevention Research Center, School of Public Health, University of Michigan — Ann Arbor, MI
Data Manager and Analyst

- Being a data manager of the center, I provide various technical support in data requests, storage, documentation and quality control to investigators for over 50 research projects, using both R and SAS.

- Collaborate with investigators and perform statistical analysis to examine long-term effects of prenatal environmental exposures on child's growth and development; analyze metabolites and DNA methylation data. In particular, I developed a data pretreatment procedure for metabolomics array data to reduce measurement biases caused by batch effects, and developed a network partitioning algorithm to identify metabolite clusters.

2017 Summer — A9.com, Inc. (A subsidiary of Amazon.com, Inc.) — Palo Alto, CA
Graduate Data Science Intern

- Employed by the Advertising Technology team to improve bidding strategies and ranking models for Amazon online display advertisements on third party websites; used Spark MLlib machine learning library on Hadoop to analyze data with 20 million entries and more than 100 million unique features.
- Responsible for the development and implementation of divide-and-conquer algorithms for ordinal categorical regression models for large scale online advertisement data, enabling parallel computation and reducing model training time by at least 50%. This work has been submitted for a patent application.

2012 - 2013 — Predictive Technology Laboratory, University of Virginia — Charlottesville, VA
Research Assistant

- Applied machine learning and natural language processing techniques to help the Virginia Department of Transportation develop an internal data quality control system by borrowing text information from police-filed traffic accident reports. Specifically, I built a multinomial classification model based on abstracted language features to predict vehicular crash types and achieved 84% classification accuracy.

Other Experience and Activities

2015 - 2016 — Rackham Student Government, University of Michigan — Ann Arbor, MI

- Served as a representative of the Biological and Health Sciences Division, and as a member of Student Life Committee and Budgetary Committee.
- Organized happy hour, sport and family-friendly events for graduate students.

2015 — Information Retrieval Class Project, University of Michigan — Ann Arbor, MI

- Implemented an NLP system to improve Amazon search engine by providing check boxes to help customers refine product search results based on free-text product reviews. The system was built using Amazon electronic product reviews from year 1995-2013.

2011 — Study Abroad Program, University of Virginia — India

- Studied comparative politics abroad in India during the winter semester of 2011; the program was partially funded by the Rai Foundation. Visited local non-profit organizations and participated in community service activities focused on promoting gender equality.

Honors and Awards

2017	Kaggle Prediction Challenge Third Place at Michigan Student Symposium for Interdisciplinary Statistical Sciences (MSSISS).
2016	Poster Award for Most Likely Transformative Scientific Impact at Michigan Institute for Data Science (MIDAS) Annual Symposium.
2015	Best Poster Award at the International Biometric Society Eastern North American Region (ENAR) Conference in Miami, FL.
2014, 15, 16	Conference Travel Grant by University of Michigan Rackham Graduate School.
2012	Outstanding Student Scholarship by University of Virginia Department of Statistics (the Best Master Student Award).
2012	Member of Pi Mu Epsilon Mathematics Society (inducted in the Spring 2012).
2010 - 2012	University of Virginia College of Arts and Sciences Dean's List (every semester).
2009	Sun Yat-sen University Merit-based Scholarship (top 10%).

Publications

1. **Tang, L.** and Song, P.X.K. (2016). Fused LASSO approach in regression coefficients clustering – Learning parameter heterogeneity in data integration. *Journal of Machine Learning Research*, 17(113), 1-23.

2. Marchlewicz, E.H., Dolinoy, D.C., **Tang, L.**, Milewski, S., Jones, T.R., Goodrich, J.M., Soni, T., Domino, S.E., Song, P.X.K., Burant, C. and Padmanabhan, V. (2016). Lipid metabolism is a key mediator of developmental epigenetic programming. *Scientific Reports*, doi:10.1038/srep34857.

3. Zhou, L., **Tang, L.**, Song, A.T., Cibrik, D. and Song, P.X.K. (2016). A LASSO method to identify protein signature predicting post-transplant renal graft survival. *Statistics in Biosciences*, 9(2), 431-452.

4. Gerber, M.S. and **Tang, L**. (2013). Automatic quality control of transportation reports using statistical language processing. *IEEE Transactions on Intelligent Transportation Systems*, 14(4), 1681-1689.

Manuscripts Submitted:

5. **Tang, L.**, Zhou, L. and Song, P.X.K. (2016). Fusion learning algorithm to combine partially heterogeneous Cox models. *Computational Statistics* (invited revision).

6. Wang, F., Zhou, L., **Tang, L.** and Song, P.X.K. (2017). Method of contraction-expansion (MOCE) for simultaneous inference in linear models. *Annals of Statistics* (under review).

7. **Tang, L.**, Zhou, L. and Song, P.X.K. (2017). Method of divide-and-combine in regularized generalized linear models for big data. *Journal of Machine Learning Research* (submitted).

8. **Tang, L.**, Chaudhuri, S., Bagherjeiran, A., and Zhou, L. (2017). Learning ordinal regression model via divide-and-conquer technique. *The International World Wide Web Conference, WWW2018* (under review).

Manuscripts In Preparation:

9. **Tang, L.** and Song, P.X.K. Fusion of pattern-mixture GEE models with nonignorable missing data.

10. **Tang, L.**, LaBarre, J., Song, P.X.K., and Peterson, K.E. Delayed effect of *in utero* lead exposure on adolescent untargeted metabolic profile.

Presentations

1. Predicting the Payment of Blight Tickets in the City of Detroit. *Michigan Student Symposium for Interdisciplinary Statistical Sciences*, Ann Arbor, MI. March 2017. (Poster). **Won the third place in the Kaggle Prediction Challenge**.
2. Method of Divide-and-Combine in Regularized Generalized Linear Models for Big Data. *International Biometrics Society ENAR Spring Meeting*, Washington, DC. March 2017.
3. Method of Divide-and-Combine in Regularized Generalized Linear Models for Big Data. *Michigan Institute for Data Science Annual Symposium*, Ann Arbor, MI. November 2016. (Poster). **Won the Poster Award for Most Likely Transformative Scientific Impact**.
4. Fused Lasso Approach in Regression Coefficients Clustering. *Joint Statistical Meetings*, Chicago IL. August 2016. (Poster).
5. Regression Coefficients Clustering in Data Integration – Learning Data Heterogeneity. *Sun Yat-sen University Precision Medicine Workshop*, Guangzhou, China. June 2016.
6. Learning Parameter Heterogeneity in Data Integration. *International Biometrics Society ENAR Spring Meeting*, Austin, TX. March 2016.
7. Regularized Lasso Approach for Parameter Fusion in Data Harmonization. *International Biometrics Society ENAR Spring Meeting*, Miami, FL. March 2015. (Poster). **Won the Best ENAR Poster Award**.
8. Text Mining for Vehicular Crash Report Narratives. *University of Virginia Department of Systems and Information Engineering*, Charlottesville, VA. October 2012.

Technical Skills

Programming	R, Python, Java, SAS, Stata, C++, HTML, CSS, JavaScript.
Language	Native speaker of Mandarin and Cantonese; proficient in English.

Software

`metafuse`: R package (available on CRAN).

`modac`: Python MapReduce implementation of GLM for Hadoop clusters (available on GitHub).

Referee Service

IEEE Transactions on Intelligent Transportation Systems.

Journal of Multivariate Analysis.

SCIENCE CHINA Mathematics.

Professional Memberships

American Statistical Association (ASA), International Biometric Society Eastern North American Region (ENAR).