Lu Tang

(*last updated: November 16, 2016*)

University of Michigan
Department of Biostatistics
1415 Washington Heights
Ann Arbor, MI 48104

Email: 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 (Computer Science minor)	Charlottesville, VA
2008 - 2010	Sun Yat-sen University Information and Computational Science major	Guangzhou, China

Research Interests

Methodology: machine learning, data mining, high-dimensional regression, clustering, variable selection, statistical inference, statistical computing, optimization

Applications: data integration, text mining, information retrieval, quality control, epigenetics, bioinformatics, children's health

Experience

Professional Experience

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

- Dissertation research: developing machine learning methods to simultaneously cluster and estimate regression coefficients for generalized linear models when combining data sets from a large number of studies.
- Created and maintaining R package `metafuse` to provide estimation and visualization of parameter grouping structures during data integration.
- Collaborate with social science investigators to examine geographical clustering of associations between important demographic factors and health outcomes.

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

- Work as data manager of the center; provide support in data requests, storage, documentation, quality controls, and integration of databases for conduct of projects (using SAS and R).

- Provide statistical support to investigators to examine the long-term consequence of prenatal and early life environmental exposures on child development, and to understand the role of DNA methylation in this process.
- Developed a data pretreatment procedure for metabolomics array data to reduce measurement biases cause by batch effect; developed a network partitioning algorithm to identify metabolite clusters.

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

- Used text mining and natural language processing (NLP) techniques to help the Virginia Department of Transportation develop an internal data quality control system based on free-style text information from police-filed traffic accident reports (using Java and R).
- Built a multinomial classification model based on NLP features, such as unigrams, bigrams and synonym concepts; developed feature filtering procedures to remove non-informative features and achieved 84% classification accuracy.
- Provided suggestions to the Virginia Department of Transportation to improve highway safety based on the analysis of highway crash accident reports; used topic modeling and other text mining methods, such as removal of stop words, word stemming and part-of-speech tagging.

Other Experience and Activities

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

- Serve as representative of the Biological and Health Sciences division, member of Student Life Committee and Budgetary Committee.
- Organize 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 in India with Professor John Echeverri-Gent from the University of Virginia; the program was partially funded by the Rai Foundation.
- Visited local non-profit organization and participated in community service activities focused on promoting gender equality.

2009 - 2010 VJoin Volunteer Organization Guangzhou, China

- Organized the Volunteer Propagation Week activities in the Sun Yat-sen University to bring short term and long term volunteering opportunities to college students.
- Participated in Lenovo Youth Social Entrepreneurship Competition and won the fifth prize.

Honors and Awards

2016	Michigan Institute for Data Science (MIDAS) Annual Symposium Poster Award for Most Likely Transformative Scientific Impact
2015	The International Biometric Society Eastern North American Region (ENAR) Poster Award
2014, 2015	University of Michigan Rackham Conference Travel Grant
2012	University of Virginia Department of Statistics Outstanding Student Scholarship
2012	Member of Pi Mu Epsilon Mathematics Society (inducted 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

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*, doi:10.1007/s12561-016-9170-z.

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.

Presentations

1. 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 presentation)
   * Won the Poster Award for Most Likely Transformative Scientific Impact.

2. Fused Lasso Approach in Regression Coefficients Clustering. *Joint Statistical Meetings*, Chicago IL. August 2016. (Poster presentation)

3. Regression Coefficients Clustering in Data Integration – Learning Data Heterogeneity. *Sun Yat-sen University Precision Medicine Workshop*, Guangzhou, China. June 2016.

4. Learning Parameter Heterogeneity in Data Integration. *International Biometrics Society ENAR Spring Meeting*, Austin, TX. March 2016.

5. Regularized Lasso Approach for Parameter Fusion in Data Harmonization. *International Biometrics Society ENAR Spring Meeting*, Miami, FL. March 2015. (Poster presentation)
   * Won the Poster Award.

Technical Skills

Programming	R, Python, Java, SAS, Stata, C++
Presentation	LaTeX, PowerPoint, Prezi, Web page design (Html, CSS, Javascript)
Web design	HTML, CSS, JavaScript
Language	Native speaker of Mandarin and Cantonese, proficient in English

Software

R package `metafuse` (available on CRAN)

Referee Service

IEEE Transactions on Intelligent Transportation Systems

Professional Memberships

American Statistical Association
International Biometric Society, Eastern North American Region