# Xiaosu Tong

2101 Cumberland Ave. Apt. 3207, West Lafayette, IN, 47906

765-586-6016   tongx@purdue.edu

## OBJECTIVE

To secure a data scientist internship for summer 2015 with a focus on statistical analysis and machine learning for large and complex data.

## EDUCATION

**Purdue University, Department of Statistics, West Lafayette, IN, US**
**August 2012 - Present**
- ➢ **Ph.D. Candidate in Computational Statistics**
  Currently working in Visualization and modelling for large, complex data using the Divide & Recombine paradigm [tesseradata.org];
  Programming environments for analysis of large, complex data, including Hadoop, MapReduce, and RHIPE( R and Hadoop Integrated Programming Environment);
  Seasonal Trend Loess modeling for spatial-temporal large data.
- ➢ **Master of Science in Applied Statistics,  May 2012**

**Shenyang Jianzhu University, School of Science, SHENYANG, CHINA**
- ➢ **Bachelor of Science in Information and Calculation Science, July 2010**

## SKILLS

C, R, RHIPE/Hadoop/MapReduce, SAS, Matlab, SPSS, HTML, Linux operation systems;
Data visualization, statistical consulting, experimental design, machine learning, data mining;
Git, Vim, Emacs, version control (Concurrent Versions System and Apache Subversion), Latex.

## RESEARCH EXPERIENCE

**Research Assistant, Professor William Cleveland**
**May 2013 – Present**
- ➢ Visualized the critical attributes of Kiva(nonprofit financial organization) dataset. Discovered significant anomalies about default loans by using RHIPE/Hadoop in R. Conducted a hierarchical cluster analysis to detect repeated borrowers in Kiva by using Divide and Recombine method with MapReduce concept.
- ➢ Visualized the all data measurement status of NCAR spatial-temporal dataset about temperature and precipitation by using RHIPE in R, which speed up the visualizing procedure dramatically.
- ➢ Conducted multiple components decomposition for spatial-temporal dataset by using Loess (locally weighted regression) fitting and STL(seasonal trend loess) with the RHIPE/ Hadoop in R. Discovered the best prediction model after tunning smoothing parameters experiment.
- ➢ Optimized the distance calculation in the source code of Loess function by using great circle distance for Spatial smoothing, prediction error is decreased by 26%.
- ➢ In charge of writing multiple documentations of RHIPE/Hadoop package in R for Tessera (Open Source Environment for Deep Analysis of Large Complex Data) [http://xiaosutong.github.io/docs.Rhipe/tutorial/]

**Research Assistant, Professor Bowei Xi**
**August 2012 – April 2013**
- ➢ Applied differential privacy concepts to hypothesis and classification. Explored three different methods for modifying covariance matrix to solve non positive-definite matrix issue in LDA and QDA classification with sensitivity and differential privacy concepts.
- ➢ Formulated the distribution functions of test statistics with differential privacy to investigate the performance of type I and type II error of test, and then calculated and visualized the effect of differential privacy to the classification error.

## RELATED EXPPERIENCE

**Teaching Assistant, Department of Statistics, Purdue University**
**August 2013 – Present**
- 'Data visualization and R language', designed all homework assignments for the course, lectured part of the course which including lattice graphic package, plyr package, snow parallel computing package, and RHIPE package in R.
- 'R/Hadoop/RHIPE', designed all homework assignments, lectured part of the course which including base R, RHIPE package, and Hadoop basis.

**Senior Consultant, Statistical Consulting Service, Purdue University**
**May 2011 – August 2013**
- Abstract statistical models (linear regression, logistic regression, survival analysis, experiment design) from real life research problems in diverse areas to provide assistance to clients with statistics analysis or probability problems and running of a wide variety of statistical computing programs, including SAS, SPSS, and R. Worked with 33 clients total.

**Teaching Assistant, Department of Statistics, Purdue University**
**May 2011 – May 2012**
- Independently instructed the laboratory course 'Elementary Statistical Methods'.

## CONFERENCE POSTERS

Bowei Xi, Murat Kantarcıoglu, Xiaosu Tong, Ali Inan, "Hypothesis Tests and Classification with Gaussian Mixture Models under Differential Privacy", Southern Regional Council on Statistics Summer Research Conference, Burns, TN, June 2013 [www.louisville.edu/sphis/bb/srcos-2013]

## HONORS AND AWARDS

**National Scholarship** awarded by Department of Education, Shenyang, China (October 2008)
**2nd Prize** in National Mathematical Modeling Contest, Shenyang, China (October 2008)
**Top 2% of class** Scholarship of Shenyang Jianzhu University, Shenyang, China (October 2007)