

Skills

- **Quantitative:** Data mining, machine learning, statistical modeling, linear / nonlinear optimization, Monte Carlo simulation, partial differential equation. Classification and regression trees (CART, GUIDE, QUEST, CRUISE, M5, C4.5), Support Vector Machine (SVM), Lasso/LARS, random forests, Spline methods (MARS), boosting, bagging.
- **Computer:** R/S-Plus, Matlab, SAS, Mathematica, C, C++, Delphi, FORTRAN, Windows, Linux.
- **Business:** Experience in Fraud Protection, Real Estate, Consumer Satisfaction Survey, interested in classification and prediction problems in Finance.

Research Interests

- Data mining, machine learning, statistical computing, Monte Carlo methods.
- Nonparametric analysis, parametric inference, Bayesian statistics.
- Functional data analysis, survival analysis, survey data analysis, microarray data analysis.

Education

Ph.D. Statistics, University of Wisconsin-Madison, 2007

Dissertation topic: Model-free variable selection algorithm for nonparametric regression and classification

B.S. Mathematics, University of Science and Technology of China, 2002

Working Experience

Senior Research Biostatistician

October 2007- present

Bristol-Myers Squibb Company, Wallingford, CT
 New drug discovery (Hepatitis C, HIV), Drug life circle management (Hepatitis B),
 Developing clinical trial methods

Research Assistant

Fall 2006 - Fall 2007

Dr. Kjell A. Doksum and Dr. Kam-Wah Tsui
 Developing a new variable selection algorithm/software for nonparametric classification problem. The new algorithm/software can improve the performance of other widely used data mining tools such as Support Vector Machine, CART, QUEST. Applications include improving the accuracy and sensitivity of detecting credit card fraud.

Consulting Statistician

NRS Corporation, Madison, WI

Summer 2006

Helping an independent consulting firm to build a standard approach of analyzing survey data collected from customers. Using regression trees (CART, GUIDE) and random

forests to identify important features that affect customer satisfaction; building tree-based models to describe customer behavior.

Research Assistant

Fall 2005

Dr. Kjell A. Doksum

Developing a new variable selection algorithm/software for nonparametric regression problem. The new algorithm/software can improve the performance of other widely used data mining tools such as MARS, CART, GUIDE. Applications include improving the accuracy of economic forecasting and income prediction.

Research Assistant

Fall 2003 - Spring 2004

Dr. Moo K. Chung, Keck Laboratory for Functional Brain Imaging and Behavior

Functional image segmentation and registration. 3D surface smoothing. Identifying corpus callosum boundary.

Consulting Experience

Dr. Michael Culbertson, Laboratory of Genetics and Molecular Biology, UW-Madison

Spring 2005 - Fall 2006

Experimental design, developing new methods to analyze time course microarray data.

Teaching Experience

- Teaching Assistant, Introduction to Bayesian Decision and Control, Fall 2006
- Instructor, Ph.D. Qualifying Exam Study Group, Summer 2006
- Teaching Assistant, Introduction to Statistical Methods, Spring 2003, Fall 2004, Spring 2006
- Teaching Assistant, Applied Multivariate Analysis, Spring 2005
- Teaching Assistant, Statistical Methods for Bioscience, Fall 2002

Articles

- Doksum, K., **Tang, S.**, Tsui, K-W. 2008. *Nonparametric Variable Selection: the EARTH Algorithm*, Journal of The American Statistical Association, 103 (484): 1609–1620.
 - Tsui, K-W., **Tang, S.** 2007 *Simultaneous Testing of Multiple Hypotheses Using Generalized p -values*, Statistics & Probability Letters, 77 (2007), 1362-1370.
 - Guan, Q.*, Zheng, W.*, **Tang, S.***, Liu, X., Zinkel, B., Tsui, K-W., Yandell, B., and Culbertson, M. 2006. *Impact of Nonsense-mediated mRNA Decay on the Global Expression Profile of Budding Yeast*, PLoS Genetics 2(11):e203.
* These authors contributed equally to this work.
 - **Tang, S.**, Tsui, K-W. 2007. *Distributional Properties for the Generalized p -value for the Behrens-Fisher Problem*, Statistics & Probability Letters, 77 (1):1-8.
 - Chung, M.K., **Tang, S.** 2004. *Nonparametric Estimation of Cortical Thickness*, 10th Annual Meeting of the Organization for Human Brain Mapping.
 - **Tang, S.**, Zhou, L., Wang, X. 2001. *Classification Models of DNA Sequences*, Mathematics in Practice and Theory, 31(1):19-26 (in Chinese).
-

Awards and Honors

- Honorable Mention, Interdisciplinary Contest in Modeling (ICM) in USA, organized by the Consortium for Mathematics and its Applications (COMAP), 2001
 - First Prize, Chinese Undergraduate Mathematical Contest in Modeling (CUMCM), 2000
-