

# Tong He

(+86)13051031037  
hetong007@gmail.com

## EDUCATION

### Sun Yat-Sen(Zhongshan) University

B.S. in Statistics, Sep. 2009 - Jun. 2013

- ▶ Overall GPA: 3.6/4.0
- ▶ Last two years GPA: 3.9/4.0

## PUBLICATION

### Augmenting LASSO Regression with Decision Tree for Identifying the Correlation of Genetic Polymorphism and Adverse Events

- ▶ Accepted as regular paper by BIBM 2013

### Introductory Statistics with R in Chinese

- ▶ One of the translators of the Chinese version, originated by Peter Dalgaard

## RESEARCH PROJECT

### Single feature polymorphisms gene detection based on Bayes model

- ▶ The basic idea for this project is dealing with small samples, modelling with hierarchical Bayes model, and computing result via EM algorithm
- ▶ Undergraduate thesis, rated as the top level.

### Microarray Data with shrinked MANOVA statistic, analysed based on FDR

- ▶ Combining MANOVA statistics and shrinked statistic to accurately estimate false discovery rate
- ▶ Borrow techniques from *Large Scale Inference* by Bradley Efron

## WORK EXPERIENCE

### Douban

Algorithm Engineer Intern, Beijing, China, 2013

- ▶ Douban.com is a famous SNS website in China with the largest database of Music, Books, Movies, etc.
- ▶ Predict user gender and income level via preference data on books, movies and groups, with the final precision higher than 95%
- ▶ Clean and structure folksonomy data of Movies, the procedure becomes the standard in Douban
- ▶ Natural language processing involving Chinese word splitting, and construct an internal word dictionary of Douban

### Giordano

Business Analyst Intern, Guangzhou, China, 2013

- ▶ Mine influential factors associated with sales trends
- ▶ For instance, how is season, temperature, city, weather and holiday influence the performance of stores in an area.

## **Industrial and Commercial Bank of China**

Data Analyst Intern, Guangzhou, China, 2012

- Organize and formalize financial data for enterprises

## **At Home Academy**

AP test Teacher, Guangzhou, China, 2011

- Teach AP Computer and AP Statistics to high school students, explain basic topics in statistics and java
- Assign homeworks and projects related to real world problems, help them prepare for the AP test
- All students got at least 4/5 in the final test

## **INVITED TALKS**

### **Cleansing and analysis of folksonomy in Douban**

- A presentation on my analysis procedure, which was developed as an intern in douban, of folksonomy data
- Read it Online : Sixth useR China conference in Shanghai, Nov. 2013

### **Ensemble model in machine learning**

- An introductory presentation on Caruana et al. *Ensemble Selection from Libraries of Models* and its variation
- First R Salon in Guangzhou, Jun. 2013

## **AWARD**

### **KDD CUP, 2013**

- Top 20% in track 1

### **ACM/ICPC, 2012**

- Third prize in Guangdong province
- Second prize in SYSU campus

### **Individual Study Scholarship, 2012**

- Dedicate to those excellent in research and study

### **National Mathematical Modeling Competition, 2011**

- Third prize in Guangdong province

## **PROGRAMMING LANGUAGE**

### **R**

- Seasoned R user
- A package rLTP as the interface between R and the ltp cloud service which is a tool for Natural Language Processing in Chinese.
- Having a project *R for Codejam* aiming to implement classical algorithms in R

### **C/C++**

- Experience with ACM/ICPC in college and Olympiad in Informatics in high school

## **ENGLISH PROFICIENCY**

**TOEFL : 102**

**GRE : 321**

## **EXTRACURRICULAR ACTIVITIES**

**Volley ball team of Mathematical School**

- ▶ Team Leader

**Chorus of Sun Yat-Sen University**

- ▶ Tenor section leader