

BINGYU WANG  
(617) 750-6151 • rainicy@ccs.neu.edu  
69 Park Dr, Apt 6, Boston, MA, 02215

## Education

---

**Northeastern University** *Candidate for PhD of Computer Science, GPA: 3.85/4.00* 2015-present  
—**Core Courses:** Algorithms · Machine Learning · Data Mining · Information Theory  
—**Teaching Assistant:** Machine Learning · Information Retrieval  
**Northeastern University** *Master of Computer Science, GPA: 3.85/4.00* Sep 2012-Dec 2014

## Research Experience

---

**Big Data in Relationship between Air Pollution and Mortality Risk**(*Ongoing*) 2016-present  
• Developed models such as Cox and Poisson Regression, to handle over 45 billion observations.  
• Conducted big data study to identify patterns and trends in air pollution associated-mortality risk

**Extreme Multi-Label Classification(XCBM)**(*Ongoing*) 2017-present  
• Developed a sparse CBM(XCBM) by exploring feature sparsity, label sparsity and label imbalance.  
• Derived and implemented a Weighted Dual Coordinate Descent method to speed up training.  
• XCBM achieved a comparable performance comparing with other extreme classifiers.

**Regularizing Model and Label Structure for Multi-Label Classification**(*Ongoing*) 2016-2017  
• Regularized Multi-Label classifiers by ElasticNet to avoid overfitting and shrink model size.  
• Combined General F-Measure Maximizer(GFM) with Support Inferences to obtain optimal instance-F1 prediction.  
• Achieved better instance-F1 comparing with existing Multi-Label methods.

**Conditional Bernoulli Mixtures(CBM) for Multi-label Classification** 2015-2016  
• Derived and implemented a new Multi-label classifier using Mixtures of Bernoulli.  
• Developed an efficient inference to make joint prediction by dynamic programming.  
• CBM outperformed other state-of-the-art Multi-Label methods.

**Topic-Factorized Ideal Point Estimation Model for Legislative Voting Network** 2013-2014  
• Crawled Roll Call Votes data and built the dictionaries for Bill Text, Voting records and legislators.  
• Implemented the topic models on bill texts, like Probabilistic latent semantic analysis (PLSA), latent Dirichlet allocation (LDA) for the baseline.  
• Visualized the voters' ideological positions on website, using D3js.

## Professional Experience

---

**MassMutual Financial Group, Boston, MA** Jan-June 2014  
*Data Analyst (Python)*  
• Recognized the pattern and performed analysis and predictions on the web log data of Oppenheimer Website using the Aster Express Tool from Teradata  
• Analyzed the MassMutual HR data and produced the predictions on the Quality of Hire  
• Performed twitter analysis for GeoAnalytics project to collect data from twitter using sentimental keywords and find out the areas where MassMutual can promote the sales

**Federal Home Loan Bank, Boston, MA** June-Aug 2013  
*Information Technology Intern (Java)*  
• Developed a Test Automation Framework that can easily be used to test different web based projects using various technologies, such as Selenium, Open2Test and Eclipse  
• Delivered documentations, including test script based on SharePoint, test results covering test suite execution and screenshot, and user manual for non-computer staff

## Publications

---

Gu, Chen, Sun, **Wang**. "Ideology Detection for Twitter Users via Link Analysis" *SBP-BRiMS* 2017  
Li, **Wang**, Pavlu, Aslam. "Conditional Bernoulli Mixtures for Multi-Label Classification" *ICML* 2016  
Li, **Wang**, Pavlu, Aslam. "An Empirical Study of Skip-gram Features and Regularization for Learning on Sentiment Analysis" *ECIR* 2016  
Gu, Sun, Jiang, **Wang**, Chen. "Topic-Factorized Ideal Point Estimation Model for Legislative Voting Network" *KDD* 2014

## Computer Knowledge

---

Machine Learning, Data Mining, Java, Python, R, MATLAB, LaTeX, D3js