

# Yuantong Li

---

250 University St, West Lafayette, IN 47907      Phone: (919)480-3652      Email: liyuantong93@gmail.com

RESEARCH INTERESTS      high-dimensional statistics, Bayesian methods, data mining, machine learning, natural language processing, and reinforcement learning

EDUCATION      Aug 2023    Ph.D. in Statistics, Purdue University,    Advisor: Dr. Guang Cheng

May 2018    M.S. in Statistics, North Carolina State University

July 2016    B.S. in Mathematics (Honors), Chu Kochen College (Shing-Tung Yau Mathematical Talent Class), Zhejiang University

PUBLICATION    **Y. Li**, Q. Ma, and S. Ghosh. Determining the Number of Mixture Components of Heavy-Tailed Densities, *The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining 2020. (KDD)*

S. Zhao, Y. Huang, C. Su, **Y. Li** and F. Wang. Interactive Attention Networks for Semantic Text Matching, *2020 IEEE International Conference on Data Mining. (ICDM)*

MANUSCRIPT    **Y. Li**, C. Wang, and G. Cheng. (2020+). Online Forgetting Process for Linear Regression Models. *submitted*.

**Y. Li**, F. Yang, H. Rao, and R. Feng. (2020+). Effective Peel Learning for Small Data with Structured Features, *submitted*.

J. Duan, **Y. Li**, J. Guo, and G. Cheng. (2020+). Ranking with Tail-Attention Regression in Stock Cross-Sectional Selection, *working paper*.

WORKING EXPERIENCE    Feb 2018 - Aug 2019      **Research Assistant (Full Time)**    Advisor: Dr. Fei Wang  
*Cornell University*

- (Teamwork) With given query about patient's cancer, genes, and demographic information, provided useful precision medicine-related papers from PubMed to clinicians treating cancer patients.
- Did query expansion including Knowledge Integration and Normalization.
- Implemented the Precision Medicine Classifier including logistic regression, Navie Bayes, SVM, and RNN with Bert pretraining.
- Using three methods: Rule based, CNN and BERT to match query and scientific articles.
- Built a precise biomedical literature retrieval engine with deep learning and external knowledge.

Sep 2018 - Feb 2019      **Research Associate (Full Time)**    Advisor: Dr. Huan Sun  
*Ohio State University*

- Implemented the neural network topic model combined with variational Bayesian inference method to get the posterior distribution of topic words of texts.
- Added regularization to that generative model to improve text topic words coherence.
- Exploited attention based Seq2Seq model to generative long texts of product review.

Oct 2015 - May 2016      **Quantitative Analyst (Intern)**  
Department of Investment, Hangzhou CIEC International Co., Ltd, China.

RESEARCH EXPERIENCE    Jan 2018 - Aug 2018      **Research Assistant**    Advisor: Dr. Sujit Ghosh  
*North Carolina State University*

- Developed a Cauchy mixture model for heavy tailed data.
- Developed two algorithms "iterative-quantile change point" and "non iterative-quantile change point".
- Defined two new statistics for the evaluation of components overlapping.
- Compared models with posterior distribution of deviance and "Rousseau and Mengersen" method.
- Greatly improved the computational time by 500 fold and consistency.
- Proved the convergence of estimated components to true components.
- Applied the model to Standard & Poor's 500 index daily return.
- Code is available MMt V1.0.

May 2017 - Aug 2017      **Research Assistant (Intern)**    Advisor: Dr. Rui Feng  
*University of Pennsylvania*

- Developed a deep learning algorithm that incorporates structure relationship among predictors.
- Developed a back-propagation algorithm and conducted simulations under different settings.
- Applied the proposed method, graphical neural network and feed forward deep learning method to understand fMRI on response time.
- Improved the upper bound of the generalization error of deep learning with structure information.
- Code is available PL v1.0.

TEACHING ASSISTANT	Aug 2019 - Dec 2019	STAT 301 (Lab), Elementary Statistical Methods
	Aug 2019 - Dec 2019	STAT 517, Statistical Inference
	Jan 2019 - May 2020	STAT 511, Statistical Methods
	Jan 2020 - May 2020	STAT 513, Statistical Quality Control
	Aug 2020 - Dec 2020	STAT 190, Data Mining

COURSEWORK    Statistical Inference I, II, Linear Model, Generalized Linear Model, Probability, Stochastic Process, Statistical Computing, Bayesian Inference, Optimization, Machine Learning

SKILLS            Programming: R, PYTHON, C++, Perl, L<sup>A</sup>T<sub>E</sub>X  
 Statistical Software: R, SAS  
 Operating System: Linux, Unix, Windows

PROFESSIONAL SERVICES    **Journal Reviewer:** IEEE TNNLS.

WEBSITE        <https://liyuantong93.com/home/>