Yi Liu

5121 South Kimbark Ave, Unit 2E, Chicago, IL 60615

+1 9093255896 | yil@uchicago.edu | yiliu9090.github.io | yiliu9090 | yi-liu-uchicago | 0000-0001-7865-5638

Research Interests

Bayesian statistics, variable selection, machine learning, scalable for massive data, deep Neural Network, systems biology, and computational biology.

Education

University of Chicago

Chicago, IL, USA

PH.D. IN STATISTICS (GPA:3.97) *Oct. 2017 - PRESENT*

- Advisors: Veronika Ročková & John Reinitz
- Research Interest: Bayesian Statistics, Monte Carlo Methods, Approximate Bayesian Computation, Deep Learning on Genomics, Uncertainty Quantification
- Relevant Courses: Deep Learning, Fundamental of Computational Biology, An Introduction to the Theory of Machine Learning

Stanford University

Palo Alto, CA, USA

MS IN STATISTICS(GPA:4.00) *Oct. 2015 - Jun. 2017*

- Relevant Courses: Statistical Inference, Machine Learning, Data Mining and Analysis, Probabilistic Graphical Model, Empirical Bayesian Methods, Modern Applied Statistics and Machine Learning (I & II), Applied Statistics, Probability Theory

Imperial College London

London, England, UK

B.S. IN MATHEMATICS AND STATISTICS FOR FINANCE (FIRST CLASS HONORS) *Oct. 2012 - Aug. 2015*

- Top performance in Statistics
- Awarded LBG Prize for top performance in Statistics, First Year Project Prize best first year project
- Relevant Courses: Statistical Modelling; Applied Probability; Games, Risks and Decisions; Time Series; Credit Scoring; Statistical Pattern Recognition; Scientific Computation (in C)

Publications

Liu Y, Ročková V **Variable Selection via Thompson Sampling[J]** In Revision for Journal of American Statistical Association (JASA) (2020) URL: https://arxiv.org/abs/2007.00187

Liu Y, Barr K, Reinitz J. **Fully Interpretable Deep Learning Model of Transcriptional Control[J].** ISMB 2020 published on Bioinformatics (2020) URL: https://doi.org/10.1093/bioinformatics/btaa506

Liu Y, Ročková V, Wang Y. **Variable Selection with ABC Bayesian forests[J]** Accepted for Journal of Royal Statistical Society Series B (JRSSB) (2020) URL: https://arxiv.org/abs/1806.02304

Javangula, P., Modarre, K., Shenoy, P., Liu, Y., Nayebi, A **Efficient Hybrid Algorithms for Computing Clusters Overlap[J]** Procedia Computer Science, (2017) URL: https://www.sciencedirect.com/science/article/pii/S1877050917308050

Modarresi, K., Radu, I., Menguy, C., Muthiyil, J. V., Liu, Y., Qiang, S., Nayebi, A. **Segment Extension Based on Lookalike Selection[P]** U.S. Patent Application 15/700,343 (2019) URL: https://patents.google.com/patent/US20190080352A1/en

Modarresi K, Liu Y, Shenoy P P, et al. **User Data Overlap Determination in a Digital Medium Environment[P]** U.S. Patent Application 15/610,033 (2018) URL: https://patents.google.com/patent/US20180349933A1/en

Liu Y, Veturi K K, Modarresi K **Security Breach Detection in a Digital Medium Environment[P]** U.S. Patent Application 15/406,494 (2018) URL: https://patents.google.com/patent/US20180205752A1/en

Invited Talks

28TH CONFERENCE ON INTELLIGENT SYSTEMS FOR MOLECULAR BIOLOGY,MONTREAL, CANADA (ONLINE) *Jul. 2020*

Fully Interpretable Deep Learning Model of Transcriptional Control

JOINT STATISTICAL MEETING, SEATTLE, USA *Aug. 2021*

Variable Selection via Thompson Sampling

Honors & Awards

2015 **LBG Prize,** Imperial College London *London, UK*

2021 **Winner,** American Statistical Association, Section on Bayesian Statistical Science, Student Paper Competition *Seattle, USA*

Industry Experience

Wayfair *Boston ,MA , USA*

Ph.D. Data Science Intern *Jun. 2019 - Sep. 2019*

- Design dynamic pricing algorithms for highly sparse data sets under a non-parametric demand function structure
- Implement Real data based simulations using dynamic algorithms to improve profit performance
- Design hierarchical models to deal with highly zero-inflated demand.

Adobe Inc *San Jose,CA, USA*

Data Science Intern *Jun. 2016 - Sep. 2016*

- Design efficient algorithms for security breach detection through user behavior
- Design algorithms for efficiently counting the number of elements in the intersection of multiple sets
- Using Empirical Bayesian Methods to detect differentiating features between different market segments

Royal Bank of Scotland *London, England, UK*

Coporate Banking Intern *Jun. 2014 - Sep. 2014*

- Draft contracts for asset finance deal up to one million pound
- Design a standard operating procedure to keep clients informed throughout the documentation process

Teaching Experience

2018 **STAT 24400 Statistical Theory and Method 1,** Teaching Assistant at University of Chicago
2018 **STAT 22400 Applied Linear Regression,** Teaching Assistant at University of Chicago
2018 **STAT 20000 Elementary Statistics,** Teaching Assistant at University of Chicago
2019 **STAT 24500 Statistical Theory and Method 2,** Teaching Assistant at University of Chicago
2020 **STAT 34520 Stochastic Processes in Gene Regulation,** Teaching Assistant at University of Chicago

Professional Service

2019 - Present **Reviewer**, **Aging** ISSN: 1945-4589, **Bernoulli** ISSN: 1350-7265, **IEEE Assess** ISSN: 2169-3536,**Statistical Methods in Medical Research**,ISSN: 1350-7265, **BMC Bioinformatics**,ISSN: 1471-2105

Research Projects

Deep Learning for Systems Biology *University of Chicago*

Working under supervision of Professor John Reinitz *Jan. 2018 - PRESENT*

- Mathematically showing that Thermodynamic Models are Deep Neural Networks to derive a DNN that is full interpretable
- Develop and implement of specific form of Convolutional and Recurrent Neural Network in Tensorflow and Keras for transcriptional regulation models

ABC Methods for Variable Selections *University of Chicago*

Working under supervision of Professor Veronika Ročková *Oct. 2017 - PRESENT*

- Design and implement variable selection algorithm using Approximate Bayesian Computation algorithms with application in Bayesian Forest
- Design and implement variable selection with Thompson Sampling algorithms with application in Bayesian Forest

Rational infectious disease surveillance through a Bayesian value of information modeling framework *Stanford University*

Working under supervision of Professor Jason Andrews *June. 2016 - Dec. 2018*

- Design Novel framework for using prior data to inform efficient targeting of health interventions against infectious diseases
- Creating a new Bayesian framework to evaluate choices and cost of treatment

Skills

Programming Python, C, R, Matlab, LaTeX, Shell Scripts, Docker
Data Mining R ggplot2, R data.table, Python Pandas, Python Numpy, Python Matplotlib, SQL, Hadoop Hive, Hadoop Hue, PySpark
Machine Learning Tensorflow, Keras, Scipy
Languages English, Chinese