YANBO XU

Georgia Institute of Technology
CODA 15F, 756 W Peachtree St NW, Atlanta, GA 30308

E-mail: asnoppy@gmail.com Phone: 218-355-8432 Webpage: https://yanboxu.github.io

ABOUT ME

I am a PhD student working with Prof. Jimeng Sun in Machine Learning at Georgia Institute of Technology. My research focus on developing cutting-edge ML/DL algorithms learning from large scale multi-modal clinical time series data and generating precise predictions at individual levels, which enables **healthcare** professionals to make data-driven decisions. My research topics include but are not limited to **Time Series Analysis, Deep Modeling, (Non-parametric) Bayesian Modeling** and **Causal Inference**.

EDUCATION

2017 - Present Ph.D. Candidate in Machine Learning
Georgia Institute of Technology

2015 - 2016 Ph.D. student in Computer Science
Johns Hopkins University

2010 - 2013 Ph.D. student in Language Technology
Carnegie Mellon University

2010 M.S. in Applied and Computational Mathematics
University of Minnesota Duluth

2006 M.S. and B.S. in Computer Science
Harbin Institute of Technology

PROFESSIONAL EXPERIENCE

2020 Summer Data Science Intern at Home Deport, Atlanta
Mentored by Walid Shalaby, Xiquan Cui
Make Subsequent purchase predictions for email recommendations.

2019 Summer Research Intern at Microsoft Research AI, Redmond
Mentored by Emre Kiciman.
Identify candidate users causally for product recommendations.

2014 - 2015 Visiting Scholar at University of California Berkeley
Mentored by Prof. Zach Pardos.
Scale up student models to MOOC platform edX.

2006 - 2008 Software Engineer at Lenovo Group Ltd., China

SELECTED AWARDS

2018 Google Summer of Code (GSoC)

2012 Best Student Paper Award
5th International Conference on Educational Data Mining

2010 Outstanding Graduates
Department of Math & Statistics at University of Minnesota Duluth

2010 Data Science Summer Institute Fellowship
University of Illinois Urbana Champaign

2009 NAACL Scholarship Award
Johns Hopkins HLT Summer School

Programming Skills

- Skilled at Python/Pytorch, C/C++ under Unix/Linux, Perl
- Capable of Matlab, R, SQL, BUGs/STAN, Java

Preprints

- **Personalized Thresholds for Optimal Disease Detection in Health Monitoring**
  Xu, Y., Sun, J.
  Submitted to The Web Conference (**WWW**), 2020
- **Prediction of Step-down Transfer in Post-Operative Norwood Children Using Convolutional Neural Networks on Continuous ECG Waveforms**
  Xu, Y., Aljiffry, A., Hong, S., Long, J. B., Sun, J., & Maher, K.
  Submitted to the Lancet Journal (**Lancet**), 2020
- **DeepRite: Deep Recurrent Inverse TreatmEnt Weighting for Adjusting Time-varying Confounding in Modern Longitudinal Observational Data**
  Xu, Y., Xiao, C., & Sun, J. [**arXiv**]
- **Bayesian Estimation of Individualized Treatment-Response Curves in Populations with Heterogeneous Treatment Effects**
  Xu, Y., Xu, Y., & Saria, S.
  Tentatively accepted with minor revision by Journal of Machine Learning Research (**JMLR**)

Publications

- **Split-Treatments Analysis to Rank Heterogenous Causal Effects for Novel Treatments**
  Xu, Y., Mahajan, D., Manrao, L., Sharma, A., & Kiciman, E.
  *International Conf. on Web Search and Data Mining (**WSDM**)*, 2020
- **HOLMES: Health OnLine Model Ensemble Serving for Deep Learning Models in ICUs**
  Hong, S.*, **Xu, Y.***, Khare, A.*, Priambada, S.*, Maher K., Aljiffry, A,Sun, J., & Tumanov, A.
  *ACM SIGKDD Conf. on Knowledge Discovery and Data Mining (**KDD**)* [***Equal contribution**], 2020
- **RAIM: Recurrent Attentive and Intensive Modeling of Multimodal Continuous Patient Monitoring Data**
  Xu, Y., Biswal, S., Deshpande, S., Maher, K., & Sun, J.
  *ACM SIGKDD Conf. on Knowledge Discovery and Data Mining (**KDD**)*, 2018
- **Predicting Changes in Pediatric Medical Complexity using Large Longitudinal Health Records**
  Xu, Y., Bahador, M. T., Searles, E., Thompson, M., Tejedor-Sojo, J., & Sun, J.
  *Anual Symposium of American Medical Informatics Assocaition (**AMIA**)*, 2017
- **A Non-parametric Bayesian Approach for Estimating Treatment-Response Curves from Sparse Time Series**
  Xu, Y., Xu, Y., & Saria, S.
  *Machine Learning for Healthcare (**MLHC**)*, 2016
- **Comparison of methods to trace multiple subskills: Is LR-DBN best?**
  Xu, Y., & Mostow, J.
  *International Conference on Educational Data Mining (**EDM**)* [**Best Student Paper Award**], 2012
- **Using Logistic Regression to Trace Multiple Sub-skills in a Dynamic Bayes Net.**
  Xu, Y., & Mostow, J.
  *International Conference on Educational Data Mining (**EDM**)*, 2011