

YUJIA DENG

3701 Parkview Ln., Apt. 17C, Irvine, CA, 92612 | yujiad2@illinois.edu | (224)-704-7978

EDUCATION

PhD in Statistics, University of Illinois, Urbana Champaign, IL

Expected May 2021

- Research Interest: tensor learning, active learning, semi-supervised clustering

Bachelor of Science in Mathematics, Fudan University, Shanghai, China

Sep 2012 – June 2016

SKILLS

- Python, R/Rcpp, MATLAB, SAS, C++, SQL, Presto, Hive

PUBLICATIONS

- Yuan, Y, **Deng, Y**, Zhang, Y, Qu, A. (2020). Deep learning from a statistical perspective. *Stat.* 9:e294.
- Radaideh, M. I., Pigg, C., Kozlowski, T., **Deng, Y.**, & Qu, A. (2020). Neural-based time series forecasting of loss of coolant accidents in nuclear power plants. *Expert Systems with Applications*, 113699.
- **Deng, Y.**, Tang, X. and Qu, A. (2020+). Correlation Tensor Decomposition and Its Application in Spatial Imaging Data. *Journal of American Statistical Association*. *Under review*.
 - Incorporate pixel correlation in image classification, improving accuracy by 30% in the breast cancer imaging data when images are noisy and signals are weak
 - Enhance correlation estimation efficiency by 50% when the sample size is small
 - Design a fast tensor decomposition algorithm with Rcpp implementation
- **Deng, Y.**, Yuan, Y., Fu, H and Qu, A. (2020+). Query-augmented Active Learning for Large-scale Clustering. *NeurIPS 2020*. *Under review*. (Presented at ICML2020 workshop)
 - Clustering with users' preference and learn underlying features by augmented labels
 - Propose a novel query strategy to select data pairs that help clustering most to boost learning efficiency with less labeling cost

PROFESSIONAL EXPERIENCE

Software Engineer Intern, Facebook, Menlo Park, CA

June 2020 - Aug 2020

- Adapt normalization and regularization to deep NN, enhancing model stability and decreasing AUC loss by 0.5%
- Train a lifespan comment model and improve the user interaction on posts by 1%
- Improve team development tools by adding model recurring evaluation visualization function
- End-to-end model development from data warehouse, feature extraction, offline training and online experiments
- Work with big data using SparkSQL, Presto, Hive.

Statistical Consultant, Illinois Statistics Office, Champaign, IL

May 2018 – May 2019

- Provide consulting service to more than 20 individual researchers and mid-sized businesses
- Communicate efficiently with clients and present the analysis result in non-technical way
- Sample projects: HIV Intervention effect analysis with GEE model, survey data analysis with latent factor model, portfolio optimization, moderation effect analysis of shopping behavior

Programmer, Boehringer-Ingelheim Biometric Department, Shanghai, China

June 2014 - Aug 2014

- Collaborate on a team of 3 with biologists and statisticians to conduct Phase II clinical trial and analyze patient response towards a new therapy
- Perform data cleansing, survival analysis of pharmaceutical efficacy and safety using SAS

Instructor: STAT 200 Statistical Analysis, UIUC, IL

Spring 2018

- Fundamental statistic course for a class of 50 undergraduates

SELECTED PROJECTS

Convolutional Neural Networks for Text Categorization via Region Embedding *Sep 2017- Dec 2017*

- Apply CNN to document classification on 4 large datasets containing 2 million documents
- Use an extra neural network to extract the relationship between the word region and its neighboring regions

Special Time Series Prediction Based on Gaussian Process Regression *Jan 2016 – June 2016*

- Auto-detect the occurrence of the change point in time series and make prediction
- Predict the traffic flow with special events such as concerts and basketball games
- Reduce the normalized root mean square error of the prediction by 8%

Text Mining on Yelp Reviews via Non-negative Matrix Factorization (NMF) *Sep 2015 – Dec 2015*

- Perform natural language processing on 6 million Yelp reviews on restaurants in America
- Discover review topics and build a recommender system based on the hidden scores of restaurants implied by reviews

AWARDS

- Statistics in Imaging Section Student Paper Competition Runner-up, Denver *July 2019*
- Student Poster Award at Statistical Methods in Imaging Conference, Irvine *June 2019*