

Yan Yin

Cornell University, Ithaca, NY 14853

yy855@cornell.edu • (607)319-9282

LinkedIn Profile: www.linkedin.com/in/yan-yin-588461168 • personal homepage: <https://yinyan5988.github.io/>

EDUCATION

Cornell University, Ithaca, NY

- MPS in Applied Statistics

Aug 2018- May 2019

Renmin University of China, Beijing, China

- B.S. in Statistics Minor in Economic Statistics
- Cumulative GPA: 3.78 / 4.0 Ranking: 4 / 27

Sep 2014- Jun 2018

RESEARCH EXPERIENCE

Deep Learning Collaborative Model on Implicit Feedback Datasets

Feb 2018- Jun 2018

- Combined matrix factorization and stacked denoising autoencoder algorithm on implicit feedback dataset aiming at cool start and sparse matrix and compared the model with PMF model

Split Questionnaire Design for Long Questionnaires

Jan 2018- Jun 2018

- Proposed a design dividing a lengthy questionnaire into subsets of survey items and then administered each subset to distinct subsamples to solve low response rate and poor quality of survey data
- Used predictive mean matching multiple imputation dealing with empty items

Textual Analysis Towards Spring Festival in Microblog

Jul 2017- May 2018

- Extracted 60GB data with SQL and cleaned it; did Exploratory Data Analysis with ggplot2 to maximize insight into our data; used light-LDA and association rules to explore topics and relations in shows and customers
- Explored online information transmission rules by characterizing the behavior of follower and being followed respectively, and used a Bayesian approach to capture repost dynamics

Geriatric Data Quality Assessment for China National Committee

Apr 2017- Nov 2017

- Used Capture-recapture method to assess quality of sampled data concerning the elderly
- Updated weights of the elderly population in 468 counties in China using weighting adjustment techniques based on auxiliary information from census data

Bike Demand prediction for Bike Sharing System

Dec 2016- Jan 2017

- Predicted occupancy of stations in long term and short term using stochastic network model and Principle Component Regression with R and Matlab; used the maximal covering location model to locate stations

Investigation of Land Rights Reform Process

Oct 2016- Feb 2017

- Conducted a national survey in 152 cities; found factors influencing satisfaction towards land rights reform in different dimensions using SPSS
- Conducted cross-sectional and longitudinal studies to analyze the status of women in rural areas with regression model, data visualization techniques and hypothesis testing

Relationship between the Value of Children(VOC) and Desired Fertility

May 2016- Apr 2018

- Designed a questionnaire and surveyed about 400 people under the two-child policy background
- Extracted three latent dimensions with factor analysis; explored the influence of VOC on desired fertility in three dimensions with hypothesis testing

INTERNSHIP

Data Analyst in Schlumberger: DDCE team

Oct 2017– Mar 2018

- Discussed with software engineers and product managers about features of products and what they need
- Wrote reproducible python scripts to extract, clean and convert 3.5 million data from database
- Characterized trajectory structures using feature engineering and partition-and-group clustering
- Predicted users' preference in trajectory design with modified Random Forest

AI Researcher in Baidu: Brand Marketing Department

Aug 2017– Oct 2017

- Collected and tracked daily international Artificial Intelligence reports by building a web crawling
- Designed a market research towards truck drivers before drowsiness detection devices released

CAMPUS ACTIVITIES

Statistical Research Association of RUC Minister of Human Resource Department in the Association

- Organized research projects such as annual Campus Satisfaction Survey
- Responsible for staffing in activities held by the association

SKILLS

R, Python, MATLAB, C++, SPSS, LaTeX, SQL, SAS, Photoshop