PENG YU

(508)903-7819
peng.yu@wisc.edu

EDUCATION

University of Wisconsin-Madison Aug. 2018 - Present
PhD in Statistics
Advised by Michael A. Newton
MS in Computer Sciences 2019 - 2021(expected)
Peking University Sept. 2014 - June 2018
Bachelor of Science, Department of Mathematics.
Bachelor of Economics (Double Degree), China Center for Economic Research.

RELATED COURSES

Mathematical Statistics I, II | Nonlinear Optimization I, II | Machine Learning | Bioinformatics | Bayesian Statistics | Causal Inference | Survival Analysis | Data Science with Graphs | Probability Theory | Regression and Analysis of Variance I, II | Intro to Algorithms | Advanced Algorithms.

SKILLS

Programming Languages	Python, R, Matlab, C
Coursera	Deep Learning Specialization (*by deeplearning.ai*)
Tools	HTCondor, MySQL, git, LaTeX, Mathematica

RESEARCH INTEREST

Nonparametric Bayes; Clustering; Bayes Decision Theory; Adaptive design.

PUBLICATIONS

- Zhang, H., Ericksen, S.S., Lee, C., Ananiev, G.E., Wlodarchak, N., **Yu, P.**, Mitchell, J.C., Gitter, A., Wright, S.J., Hoffmann, F.M., Wildman, S.A., and Newton, M.A. (2019), "Predicting kinase inhibitors using bioactivity matrix derived informer sets," *PLoS Computational Biology*, 15(8): e1006813.
- **Yu, P.**, Ericksen, S.S., Gitter, A. and Newton, M.A., "Bayes Optimal Informer Sets for Early-Stage Drug Discovery", In revision. arXiv
- **Yu, P.** and Newton, M.A, "A Probabilistic Framework for Surrogate Selection", In manuscript.

WORK EXPERIENCE

Quantitative Researcher Intern - Scientech Research Capital 06/2021 - 08/2021

Project 1: Alpha research on supply chains

- Formalize a statistical problem from existing papers. Define proper research problem.
- Try to predict daily level returns with information from related stocks on the same supply chain. Try different methods to combine supply chain information.
- Diagnose models with proper plots and make necessary adjustments.
- Develop a trading strategy based on alpha values under certain portfolio restrictions.

Project 2: Alpha research on intraday momentum of US stock market

- Try to predict last half-hour returns with overnight and first half-hour information.
- Increase out-sample performance via different additional variables and stock partitions.
- Find best time to open and close positions with further analysis.
- Develop a slight different trading strategy with alpha values and same portfolio restrictions. Portfolio performances agree with statistical analysis.

RESEARCH EXPERIENCE

Bayes Optimal Informer Sets for Drug Discovery

- Develop Dirichlet process mixture model on binary data matrix. Implement DPMM clustering on binary data matrix.
- Formalize drug discovery problem with informer set as a two-stage decision problem. Come up with a general solution to the problem based on Bayes decision theory.
- Develop collapsed MCMC sampling and efficient informer set selecting algorithms. Reduce the running time from 60 hours to 5 hours on the same computer. Implemented with R/RStan and Python.
- Apply proposed methods on two drug-related data sets. Use parallel computing to reduce informer selection time from 10 days to 3 days. The platform is HTCondor.

PROJECTS

Portfolio Trading Simulator in Python
The project is designed to backtest trading strategies using minute level trade data and daily indicators. The designed simulator contains 3 Python classes combined to achieve: translate alpha values to desired positions under given restrictions; open, close and rebalance positions at given time within given time window; automatically adjust for CAC if positions are held overnight; calculate PnL, booksize, trading volume and other portfolio characteristics every day.

Survival analysis of age effect on COVID-19 mortality
The project uses Korea CDC COVID-19 data to asses the age difference in the case fatality rate. We factorized age variable and adopted Cox hazard model as well as χ^2 test to estimate the influence of age factor. Kaplan-Meier curve and log-rank test are also applied to illustrate the intuition of age factor influence.

SOFTWARE

BOISE: R package for Bayes optimal informer set selection.

informRset: R package for adaptive selection and coding selection for informer-based-ranking problem.

AWARDS

Excellent Academic Performance Award	*Sept. 2015 - Sep.2017*
May 4th Scholarship	*Sept. 2016 - Sep.2017*
The Yizheng Alumni Scholarship for Excellent Students	*Sept. 2014 - Sept. 2015*
2nd Prize in Beijing in 7th National Undergraduate Mathematical Contest	*Oct. 2015*
3rd Prize in 29th Chinese Mathematical Olympiad	*Dec. 2013*