

CAI, Zeyuan

+86 15868295486 | zcaiz75@gatech.edu | <http://caizeyuan.github.io>

Education Background

Undergraduate: Lanzhou University

09/2015-07/2019

Major: Mathematics and Applied Mathematics

GPA: 3.3/4.0 **Rank:** 5%

Core Courses: *Mathematical Analysis, Advanced Algebra, Probability Theory, Differential geometry, Mathematical Statistics, Stochastic Process, Numerical Analysis, Real Analysis, Functional Analysis, etc.*

Summer Exchange: University of California, Los Angeles

07/2018-09/2018

Major: Statistics **GPA:** 4.0/4.0

Graduate: Georgia Tech

08/2019-12/2020

Major: Statistics in ISyE **GPA:** 3.9/4.0

Core Courses: *Theoretical Statistics, Computational Statistics, Bayesian Statistics, Design of Experiment, Testing Statistical Hypothesis, Multivariate Data Analysis, Data Mining&Statistical Machine Learning, etc.*

Research/Project Experiences

Detection of Abnormalities in Mammogram Based on Deep Learning

02/2020-04/2020

- Preprocessed the raw data and transform the original dataset into usable one, i.e. convert .LJPEG file to .PNG file, downsampled the images and extracted regions of interest to satisfy the training process.
- Based on transfer learning, trained the model on VGG-16 and VGG-19 model. Modified the initial CNN architectures, more specifically, changed the size of input image to 299*299 and deleted some layers to improve the generalization performance.

A Streamline Model for Heart Disease Diagnosis

03/2020-05/2020

- Extracted raw heart disease data from UCI Machine Learning Repository, performed preliminary data analysis and variable extraction. Based on the problem that the cost of the whole set of heart disease detection is too high, a "streamline model" is proposed: The whole set of heart disease detection is divided into two stages, the features with low detection cost are classified into the first stage, and the rest are classified into the second.
- For the data of stage one, the logic regression training is carried out by using the selected characteristics. If the test sample is positive, it will be moved to state two. Using random forest model in state two, the false positive samples in state one can be screened to improve the detection accuracy. Compared with the non-streamline model, the accuracy of the streamline model is only 3% lower than that model.

Sequential Batch Learning in Linear Contextual Bandits with Partially Observed Offline Data 10/2020-Now

- As a special case of reinforcement learning, multi-armed bandits can be used to simulate how to find the optimal strategy in recommendation business. In this study, the concepts of contexts, actions and rewards are extended to features, recommendations and results respectively. Under this framework, the selection of optimal strategy and action is analyzed.
- Different from the basic online/offline learning, we introduced the framework of batch learning to make the intuition more general. We analyzed the difference of batch size in the case of stochastic contexts and adverse contexts. Using partially observable data, we studied how to use this part of data for online control by converting partially observed data to linear side information, then got an optimization problem under linear constraints. We tried to prove that the regret bound in this case can be improved compare to the general batch learning problem.

Competence

Main Awards

Outstanding Graduates of Lanzhou University
First Class Scholarship of Lanzhou University

2019

2017-2018

Computer

C++; Matlab; R; Python; LaTeX