

Ningsheng Zhao

Tel: (514) 813-2388

Email: ningsheng.zhao@mail.concordia.ca

Education

Concordia University

2019 - , Montreal, QC, Canada

Information and Systems Engineering, PHD

Research Interest: Machine Learning, Deep Learning, Reinforcement Learning, Graphical models, Data Science, and Statistical Modeling.

University of Waterloo

2017-2018, Waterloo, ON, Canada

Statistics, Master of Mathematics

Courses: Statistical Learning - Classification and Function Estimation, Computational Inference, Machine Learning, and Predictive Analysis in Actuarial Science.

Concordia University

2014 - 2017, Montreal, QC, Canada

Honor Statistics, Bachelor of Arts

Courses: Advanced Probability Theory, Stochastic Differential Equations, Mathematical Statistics, Time Series and Object-Oriented Programming.

Jiangsu University

2003 - 2007, Jiangsu, China

Statistics, Bachelor of Economics

Courses: Statistical Analysis, Probability Theory, Economics and Econometrics, and Risk Management.

Relevant Projects

Applications of Boosting and Deep Models to Reinforcement Learning

April 2019, Project, Individual

A part of *Reinforcement Learning* course.

Description: An upgraded tree Boosting Monte Carlo algorithm is proposed to estimate the state value function with a given policy in reinforcement learning tasks. Two processes are designed in the proposed algorithm: one is the boosting process to learn a sequence of tree structures, the other one is the upgrading process to update the weights using stochastic gradient descent. In addition, A deep Q-learning algorithm and a deep SARSA algorithm are respectively designed on the experiment of Mountain Car Task, both of them show excellent performance. Language: R.

Gradient Tree Boosting Cluster-Weighted Model with Kernel Density

August 2018, Master Graduation Essay, Individual

Description: A novel model that incorporates Cluster-Weighted Model, Gradient Tree Boosting Model, and Kernel density estimator to solve density estimation, clustering and classification problems. A Boosting-EM algorithm, which combines the Boosting algorithm and EM algorithm together, is used to train a mixture of k boosted regression trees together to update the clusters of data. The results of some experiments on artificial data and real data strongly support the excellent performance of this model on both clustering and classification tasks.

Beijing PM2.5 Analysis and Prediction Using Additive Regression Models
April 2018, Project, Individual

A part of *Statistical Learning - Function Estimation* course.

Description: An analysis of the Beijing PM2.5 level dataset. Different regression techniques such as XGBoost and Random Forest are fitted to compare results. Especially, a novel model of Spline Boosting is also written, tested and showed comparable results with existing regression models in terms of accuracy and running time. Language: R.

R Implementation of High-dimensional Vector Autoregressive models

April 2018, Project, Team

A part of *Computational Inference* course.

Description: A package called `varngc` is written based on the combination of several results on Vector Autoregression Models. This package accommodates two loss functions (a) penalized least square and (b) penalized log-likelihood; and two penalty schemes of (i) lasso and (ii) group lasso. The curse of dimensionality in VAR model is addressed by implementing parallelizable algorithms proposed in the two above-mentioned papers. The written R package is thoroughly tested using a simulation study and a financial application on stock price. Responsible for group lasso models.

Irony Detection on Twitter

December 2017, Project, Team

A part of *Statistical Learning - Classification* course.

Description: An detection of irony in English tweets on Twitter. The data preprocessing is done to replace emojis, usernames, URL, numbers, stopwords, etc. using stemming and tokenization. The features are selected using n-grams, sentiment polarity lexicons, bags of words, and pre-trained word embeddings. The fitted models are XGBoost, Support Vector Machine, Logistic Regression and Majority Voting. The combination of feature selection techniques and model selections are tested and compared. Responsible for running and testing models. Language: Python.

The application of Dirichlet and Hierarchical Dirichlet Process

December 2016, Honor Undergraduate Project, Individual

Description: A study about the statistical application of Dirichlet distributions, Dirichlet processes and Hierarchical Dirichlet Process. Topics include estimation of the distribution function and the mean; the Stick-Breaking Construction, the Chinese Restaurant Model, the Chinese Franchise model and Dirichlet (Hierarchical) Finite Mixture Model.

Scholarships & Awards **Concordia University Graduate Doctoral Fellowship**

Concordia University, 2019.01-2021.12

Ontario Graduate Scholarship

University of Waterloo, 2017-2018

President's Graduate Scholarship

University of Waterloo, 2017-2018

Dean's List

Concordia University, 2014-2016

Personal skills

Programming languages: Python, R, & Java

Packages: `tensorflow`, `keras`, `scikit-learn`, `nltk`, etc. in Python and `glm`, `grplasso`, `xgboost`, `catboost`, etc. in R
Tools & Softwares: , Mathematica, SAS, Microsoft Excel, L^AT_EX
Languages: Mandarin (native), English and French (beginner).

Work Experience

Peoples Bank of China, Zhenjiang Central Sub-Branch, China

2010-2012, Credit Manager

Analyzed the risk of chattel mortgage business to prevent financial risk, managed database system of enterprise and individual credit information, promoted the tripartite corporation among banks, insurance companies and enterprises.

Peoples Bank of China, Zhenjiang Central Sub-Branch, China

2007-2010, Financial Statistician

Collected, examined and summarized financial data into the database system and produced statistical reports. Researched the abnormal fluctuation of statistical, financial, economic data by means of sampling surveys and questionnaires. Used mathematical and statistical methods to model the regular patterns and the correlation among financial variables for inference and prediction.