

SHUWEN CHAI

EDUCATION

Renmin University of China

Beijing, China

Major: Statistics, Bachelor of Science, School of Statistics

September 2018 – July 2022(expected)

- GPA: **3.79/4.00** (17/133)
- *Statistics disciplines of RUC ranked top 2 in the 4th national academic assessment by Chinese Ministry of Education*
- Course Highlights: Introductory Programming I II (98,93), Data Structure and Algorithm I (93), Introduction to Database Systems (91), Introduction of Machine Learning (96), Real Variable Function (95), Mathematical Analysis II (91), Probability Theory (88), Mathematical Statistics (91), Accounting (96), Statistical software (97), Big Data Analysis (95), Multivariate Data Analysis (93), Non-parametric Statistics (90)

The University of Hong Kong

Hong Kong, China

Faculty of Science (International Exchange Student)

January 2021 – June 2021

- GPA: **4.15/4.30**
- Course Highlights: Multivariate Data Analysis (4.3), Introductory Macroeconomics (4.3), Big Data Analysis (4.0), Time-series Analysis (4.0)

PUBLICATIONS & IMPORTANT PAPERS (*EQUAL CONTRIBUTION)

1. Pan Du, Suyun Zhao, Hui Chen*, **Shuwen Chai***, Hong Chen, Cuiping Li. Contrastive Coding for Active Learning under Class Distribution Mismatch. ICCV 2021.
2. **Shuwen Chai**, Hui Guo, Boyu Wang, Feng Liu, Grace Yi. Differentially Private Multi-Source Domain Adaptation. Under-review.
3. Pan Du, Hui Chen, Suyun Zhao, **Shuwen Chai**, Hong Chen, Cuiping Li. Contrastive Active Learning under Class Distribution Mismatch. Under-review.

SELECTED RESEARCHES & PROJECTS (CHRONOLOGICAL)

I College of Information Sciences and Technology, Penn State University

Pennsylvania, U.S. (virtual)

Advisor: Jinghui Chen, Assistant Professor

Sep 2021 – Dec 2021

Robust Membership Inference Attack & Backdoor

- Designed a novel membership inference attack model that has robustness against target model's adversarial attack.
- Designed a data free framework to unlearn the planted backdoor triggers, i.e., kind of perturbations that would make the machine learning model misclassify instances to a certain target label. Paper preparing.

II Department of Computer Science, University of Western Ontario

Ontario, Canada (virtual)

Advisor: Boyu Wang, Assistant Professor

Mar 2021 – Oct 2021

Differentially Private Multi-Source Domain Adaptation

- Proposed the DPMDA method, overcoming the limitations of previous differentially private domain adaptation framework when there are multiple sources available. Remarkably, our proposed approach learns domain relationship under consideration of both task similarity and noisy perturbations resulted from different privacy budgets from source domains.
- Developed a uniform upper bound of generalization error on the target domain for DPMDA, which reveals the impact of differential privacy guarantee on the target risk.

III School of Statistics, RUC

Beijing, China.

Advisor: Yang Li, Professor

Sep 2021– Graduation

Integrative Statistical model and Machine Learning approaches for GWAS Framework with multiple categorical traits traits

- Used Random forests, XGBoost and other ML methods to perform dimensionality reduction on the pre-selected locations.
- Introducing a penalized regression model to traditional least squares and maximum likelihood estimation.
- Applied analyzable functionalization processing to obtain an SNP curved penalized regression model.

Advisor: Yilin Wu, Professor

Feb 2020 – May 2021

Effects of Online Education on Fairness in Education

- Defined three aspects of fairness in education and proposed indicators of each aspect, through literature reading.
- Designed a questionnaire and would randomly sample 1,000 high school students in Beijing for investigation.
- Multiple methods of machine learning, econometric models, difference model will be used to analyze the data.

IV Laboratory of Data Warehouse and Business Intelligence, School of Information, RUC

Beijing, China

Email chaishuwen@ruc.edu.cn

Personal Website <https://shuwenchai.netlify.app/>

Contrastive Coding for Active Learning under Class Distribution Mismatch

- Designed several whole new frameworks that are expected effective in AL under this particular scenario.
- Implemented, tested, and documented different previous frameworks in PyTorch; perform tests on CIFAR10, CIFAR100, 4-to-1 mixed dataset under different mismatch rate.
- Our CCAL method reaches state-of-the-art performance on different mismatch ratios.

V Faculty of Science, HKU

Hong Kong, China

Lecturer: Guodong Li, Professor

Mar 2021– May 2021

Hong Kong Air Traffic: Explanation and Prediction based on Sparse Seasonal ARIMA Model

- Designed sparse seasonal ARIMA model to predict the air traffic from January to July in 2020.
- Extensively studied the effects from important social events on air traffic.

Lecturer: Michael Mingyi Zhang, Assistant Professor

Mar 2021– May 2021

Fake News Detection: Traditional Machine Learning and Deep Learning Approach

- Designed and built LSTM network using GloVe neural embedding and separate embeddings for headlines and bodies.

SELECTED AWARDS / HONORS

- | | |
|--|------|
| • Scholarship for outstanding exchange students of 2020-2021 Academic Year, RUC | 2021 |
| • MITACS Internship Award | 2021 |
| • The Third-Class Academic Scholarship of 2019-2020 Academic Year, RUC. | 2020 |
| • First Prize at provincial level (Beijing), Contemporary Undergraduate Mathematical Contest in Modeling | 2020 |
| • National Project of College Students' Innovative Entrepreneurial Program (Leader). | 2020 |
| • Honorable Mention (top 25%), Mathematical Contests in Modeling | 2020 |
| • The Third-Class Academic Scholarship of 2018-2019 Academic Year, RUC. | 2019 |
| • The Second-Class Prize, Data Visualization Competition, held by School of Statistic, RUC. | 2019 |

TEACHING

- **Teaching Assistant**, Renmin University of China: Introduction of Machine Learning (2021)

INTERNSHIP

- Research Internship at School of Information, **Renmin University of China** (Jul 2020 – Mar 2021)
- Research Internship at Department of Computer Science, **Western University** (Mar 2021 – Sep 2021)
- Research Internship at College of Information Sciences & Technology, **Penn State University** (Sep 2021 –)

ADDITIONAL INFORMATION

Languages

- TOEFL: 111 (R 30 L 29 S 25 W 27) | TOEFL Best Score 113 (R 30 L 29 S 27 W 27)
- GRE: 320 (V 151 Q 169 W 4.0)

Interests & Skills

- Piano and Classical Music, Golf, Basketball
- C/C++, Python, R, LaTeX, Markdown