Tao Yang

CONTACT INFORMATION

Nankai University
School of Computer Science
Weijin Road, Tianjin, 300071
Email: stayt@mail.nankai.edu.cn
GitHub: https://github.com/stayt1

CITIZENSHIP

Chinese (The People's Republic of China)

EDUCATION

Nankai University

B.S. in Statistics, August 2018-June 2022.
Major: Data Science and Big Data Technologies
GPA: 87.48/100

M.S. in Computer Science, August 2022-Present.
Major: Computer Science

❑**Core Courses:**

Real Analysis(97)
Distributed Storage and Computing in Statistics (94)
Computer Network (90)
Doing Data Science(95)
Machine Learning(97)
Algorithms (92)
Mathematical Analysis I (91)
Mathematical Analysis II(94)
Mathematical Statistics I(88)
Mathematical Statistics II(87)
Advanced Algebra and Analytic Geometry (92)
Predictive Analysis(88)
Field Theory and Infinite Series (93)
Multivariable Calculus(99)

EXCHANGE INTERNS

Visiting Student at King Abdullah University of Science and Technology (**KAUST**) August 2021- August 2022.

Research Intern at HONG KONG University of Science and Technology (**HKUST**) August 2022- October 2022.

RESEARCH INTERESTS

My current research interests focus on

Data Science:

Asymptotic Statistics, Machine Learning, Bayesian Methods, Statistical Estimation, Variational Inference, Learning Theory

Security& Privacy:

Differential Privacy, Federated Learning , Cryptography

Research Experience

04/2020-03/2021 Distributed Radiation Detection with Bayesian Method
Advisor: Fan Yang, Nankai University, School of Physics
Instead of using $p(\mathbf{x}^* \mid y^*, \theta)$ and $p(y^* \mid \pi)$ to derive the posterior of a new label, we use $p(\mathbf{x}^* \mid y^*, \mathcal{D})$ and $p(y^* \mid \mathcal{D})$ by integrating the parameters out. The benefit of integrating parameters out using posteriors is to make full use of the prior distribution of the data and get more accurate results. Lastly, we use the EM algorithm to solve the parameters.

04/2021-07/2021 Anomaly Detection: Car Collision Detection using LGBM
Advisor: Qiaozhen Zhang, Nankai University, School of Statistics
The problem is predicting statement and motion information, which faced data imbalance. We use data under-sampling and resampling to solve it. Using cross-information correction to improve the correlation between variables and labels. The Light-GBM speeds up the model training. Finally, we use 5-fold cross-validation and Bayesian tuning to get the model.

10/2021-06/2022 Improved Private Word Embedding via Truncated Laplacian Mechanism (under Review in EACL2023)
Advisor: Di Wang, KAUST, School of Computer Science
We propose a novel private embedding method namely the high dimensional truncated Laplacian mechanism. Theoretically, we show that our method has lower variance compared with the previous private word embedding methods. Moreover, we conduct experiments on private embedding and downstream tasks over several datasets and show even in the high privacy regime our approach only results in a slight drop in utility compared with the non-private case.

01/2022-06/2022 Ensemble Learning for Password Guessing (Coming Soon)
Advisor: Jun Xu, Nankai University, School of Statistics
It is a very significant research topic whether machine learning-based guessing algorithms can reduce computational overhead and improve guessing efficiency while increasing the hit rate of attacks. Existing studies show that ensemble learning is significantly effective in improving weak learners. We analyze the theoretical and practical possibilities of various ensemble learning algorithms (such as Bagging, Boosting, and Stacking) for password guessing. The new password-guessing framework is designed for real attack scenarios.

Honors and Awards

Nankai University

2022	Nankai Admission Scholarship
2022	Nankai Research Scholarship
2021	KAUST Exchange Scholarship
2019, 2020	National Motivation Scholarship
2021	H Prize in COMAP Modeling
2020	First Prize in The Chinese Mathematics Competitions
2018	Provincial Excellent Student Cadre

Service

Reviewer for **EMNLP 2022, TBD 2022, WASA 2022, EACL 2023**

Skills

TOEFL:106/120 and **IBM** Z Xplore - **Advanced**

Languages:

C++ (Advanced), Python (Advanced), R (Advanced), HTML (Advanced)
MySQL (Advanced), PySpark (Medium), LaTeX (Advanced)

Libraries:

PyTorch, Tensorflow, Keras