Mengyan Zhang

(+61) 406 200 706
mengyan.zhang@anu.edu.au
mengyanz.github.io

EDUCATION

Ph.D. Candidate in Machine Learning
Computational Media Lab, The Australian National University, 2018.08- present
Machine Learning Research Group, Data61, CSIRO, 2018.08- present

Bachelor of Computer Science (first class honours)
The Australian National University and Shandong University, 2+2 joint degree, 2014-2018, GPA 6.938/7.0

RESEARCH INTEREST

My research interests are online experimental design in machine learning, including multi-armed bandits and active learning. I work on both theoretical and practical views of experimental design with two goals: (I) Designing robust algorithms to reflect the preference of agents and understanding the concentration of measures and adaptive strategies involved. (II) Designing the pipeline and recommendation strategies for synthetic biology experimental design applications.

RESEARCH & PROJECT

Quantile Bandits for Best Arms Identification
***Zhang, M.**, Ong, C. S. (2021). International Conference on Machine Learning.*
Motivated by risk-averse decision-making problems in fields like medicine, biology and finance, we consider the task of identifying a set of m arms with the highest τ-quantile values under a fixed budget. We propose Quantile Successive Accepts and Rejects algorithm (Q-SAR), and upper bound the probability of error by using the proposed two-sided asymmetric concentration inequalities for order statistics and quantiles.

Opportunities and Challenges in Designing Genomic Sequences
***Zhang, M.**, Ong, C. S. (2021). ICML Workshop on Computational Biology.*
We show results on using Bayesian optimization to maximise gene expression in bacteria, where machine learning enabled us to discover a strong regulatory element. Using the Design-Build-Test-Learn (DBTL) workflow as a case study of how to effectively use machine learning in genomic sequence design, we argue that machine learning has tremendous potential in this area. Based on our experience, we discuss several opportunities and challenges that we have identified, and conclude with a call to action for more collaborations.

Machine Learning guided Design of Ribosome Binding Sites
***Zhang, M.**, Holowko M. B., Hayman Zumpe H., Ong, C. S. (2021). under review.*
I work with collaborators in Synthetic Biology apartment in CSIRO and BioFoundary Lab. We aim to adaptively design the ribosome binding site (RBS) of E. coli to optimise the protein expression level with a fixed budget. We build a Gaussian Process regression model and an Upper Confidence Bound recommendation model to find the optimal choice with high protein expression.

Active Learning Library for Knowledge Graph
*Alger M., **Zhang M.**, Ong, C.S. (2018), Python software*
We build a Python active learning software, Acton, for scientific research usage. The design of Acton includes three parts: Labeller, Predictor and Recommender. I mainly contribute to the modular of active learning for knowledge graph. The main idea is to use active learning techniques for knowledge graph completion.

Classification of Historical Death and Occupation coding

Zhang, M. *(2018), Honours thesis. (Supervisors: Christen, P., Graham, T.)*

We consider the classification tasks for real historical death and occupation data sets. The objectives are to develop text classification techniques and evaluate these techniques on large real-world data collections. The project mainly uses Python as the programming language.

AWARDS

2019 Data61 Top-up Postgraduate Research Scholarship
2018 PhD Scholarship of ANU
2018 ANU HDR Fee Remission Merit Scholarship
2017 Paul Thistlewatte Memorial Honours Year Scholarship of ANU
2015-2016 National Scholarship (China)
2016 Lan Qiao Cup Programming Competition Shandong province 1st prize

PRESENTATIONS

Quantile Bandits for Best Arms Identification
Machine Learning Summer School 2020 (acceptance rate: 13.84%)
Max Planck Institute for Intelligent Systems, Tübingen, Germany

Optimized Experimental Design for Translation Initiation using Machine Learning
Collaborative Conference on Computational and Data Intensive Science (C3DIS) 2020
Canberra, Australia

TEACHING

Tutor for COMP8600 Statistical Machine Learning (S1 2019, S1 2020, S1 2021)
Tutor for COMP6670 Introduction to Machine Learning (S2 2020)

TECHNICAL SKILLS

Programming: Python (familiar with PyTorch), Java, C#, C++
Language: Chinese, English
Others: Git, LaTex, Photoshop

REFEREE

Dr. Cheng Soon Ong
Principal Research Scientist, Director of the ML and AI future science platform at CSIRO
chengsoon.ong@anu.edu.au

Prof. Lexing Xie
Professor, School of Computing, ANU; Director of Computation Media lab
lexing.xie@anu.edu.au