2024

Workshop on Data Science and Statistics

2024.1.9-11

The Hong Kong Polytechnic University

Host

Research Centre for Mathematical Foundations of Generative AI

Department of Applied Mathematics

The Hong Kong Polytechnic University

Table of Contents

Registration: January 9, 2024
January 10, 2024
Workshop: January 10-11, 2024
Place: Polyu Campus, Building Y, Room Y301

Workshop on Data Science and Statistics

The Hong Kong Polytechnic University
January 9-11, 2024

The objectives of this workshop are to examine the fundamental mathematical and statistical principles that form the foundation of data science and artificial intelligence, to explore novel and theoretically sound methodologies for data science and AI, and to foster the exchange of ideas among researchers working in these related fields. Over 40 experts from the Mainland, China and overseas are invited to attend the Workshop.

In addition to research talks, graduate students in the Department of Applied Mathematics at The Hong Kong Polytechnic University have prepared posters to present their work advised by their supervisors.

This workshop is organized by the Research Centre for the Mathematical Foundations of Generative AI in the Department of Applied Mathematics at The Hong Kong Polytechnic University.

Workshop on Data Science and Statistics
Program

Date	Time	Activities
January 9	1:00 pm - 6:00 pm	Registration (TU724) (Registration can also be done on January 10 at the Workshop site Y301)
	6:00 pm	Dinner
January 10	08:30 am - 09:00 am	Registration
	09:00 am - 09:05 am	Opening remarks. Jian Huang, The Hong Kong Polytechnic University
	09:05 am - 09:10 am	Remarks, Defeng Sun, Head, Dept of Applied Math, The Hong Kong Polytechnic University
	Session 1. Chair: Jian Huang, The Hong Kong Polytechnic University	
	09:10 am - 09:35 am	Tony Cai, University of Pennsylvania *Title: The Cost of Privacy: Optimal Differentially Private Learning*
	09:35 am - 10:00 am	Kung-Sik Chan, University of Iowa *Title: Change Point Detection with Matrix-Variate Data: Leveraging and Adapting to Structured Mean Changes*
	10:00 am - 11:00 am	Group photo, tea break, poster viewing, and discussion
	Session 2. Chair: Xinyu Zhang, Chinese Academy of Sciences	
	11:00 am - 11:25 am	Xiaotong Shen, University of Minnesota *Title: Boosting Data Analytics with Synthetic Volume Expansion*
	11:25 am - 11:50 am	Hansheng Wang, Peking University *Title: Network Gradient Descent Algorithm for Decentralized Federated Learning*
	12:00 pm - 2:00 pm	Lunch
	Session 3. Chair: Zhao Chen, Fudan University	
	2:00 pm - 2:25 pm	Linglong Kong, University of Alberta *Title: Online Local Differential Private Quantile Inference via Self-normalization*
	2:25 pm – 2:50 pm	Juyong Zhang, University of Science and Technology of China *Title: Creating High-Fidelity Digital Avatar for Everyone*

Date	Time	Detail
January 10	2:50 pm - 4:00 pm	Tea break, poster viewing, and discussion
	Session 4. Chair: Yao Wang, Xi'an Jiaotong University	
	4:00 pm - 4:25 pm	Jinyuan Chang, Southwestern University of Finance and Economics and CAS *Title: Exploring Excellence: Bayesian Penalized Empirical Likelihood and MCMC Sampling*
	4:25 pm - 4:50 pm	Xueqin Wang, University of Science and Technology of China *Title: Green's Matching: An Efficient Approach to Parameter Estimation in Complex Dynamic Systems*
	4:50 pm - 6:00 pm	Discussion
	6:00 pm	Dinner
January 11	Session 5. Chair: Xingqiu Zhao, The Hong Kong Polytechnic University	
	9:00 am - 9:25 am	Qiwei Yao, London School of Economics *Title: Blind Source Separation over Space*
	9:25 am - 9:50 am	Ming Yuan, Columbia University *Title: On the Multiway PCA*
	9:50 am -10:15 am	Bin Nan, University of California, Irvine *Title: Survival Estimation with Time-Varying Covariates Using Neural Networks*
	10:15 am - 11: 00 am	Tea break, poster viewing, and discussion
	Session 6. Chair: Fang Fang, East China Normal University	
	11:00 am - 11:25 am	Johannes Schmidt-Hieber, Universiteit Twente *Title: Statistical Learning in Biological Neural Networks*
	11:25 am - 11:50 am	Lijian Yang, Tsinghua University *Title: Hypotheses Testing of Functional Principal Components*
	11:50 am - 12: 15 am	Zhou Yu, East China Normal University *Title: Nonlinear Sufficient Dimension Reduction*
	12:15 am - 2:00 pm	Lunch
	Session 7. Chair: Yixuan Qiu, Shanghai University of Finance and Economics	
	2:00 pm - 2:25 pm	Tengyuan Liang, University of Chicago *Title: Randomization Inference When N = 1*

Date	Time	Detail
January 11	2:25 pm - 2:50 pm	Yazhen Wang, University of Wisconsin, Madison *Title: Reinforcement Learning in a Continuous-Time Stochastic Setting*
	2:50 am - 3:15 am	Minge Xie, Rutgers University *Title: Inference for Discrete or Non-numerical Parameters and Unraveling Machine Learning Black-Box Models*
	3:15 pm - 4:15 pm	Tea break, poster viewing, and discussion
	Session 8. Chair: Binyan Jiang, The Hong Kong Polytechnic University	
	4:15 pm - 4:40 pm	Yancheng Yuan, The Hong Kong Polytechnic University *Title: DreamRec：Reshaping Sequential Recommendation Systems*
	4:40 pm - 5:05 pm	Jian Huang, The Hong Kong Polytechnic University *Title: A Bayesian Framework for Fine-Tuning Large Diffusion Models*
	5:05 pm - 6:00 pm	Discussion
	6:00 pm	Dinner

Workshop on Data Science and Statistics

January 9-11, 2024

Abstracts

The Cost of Privacy: Optimal Differentially Private Learning

Tony Cai, Department of Statistics & Data Science, The Wharton School, University of Pennsylvania

In today's data-driven world, the proliferation of personal data and technological advancements underscores the paramount importance of safeguarding privacy. Developing statistical methods with privacy guarantees is becoming increasingly important. However, privacy guarantees of statistical methods are often achieved at the expense of accuracy. In this presentation, we discuss the trade-off between statistical accuracy and differential privacy in several fundamental statistical problems, including Gaussian mean estimation and linear regression. We address both the traditional low-dimensional and the contemporary high-dimensional settings.

Change Point Detection with Matrix-Variate Data: Leveraging and Adapting to Structured Mean Changes

Kung-Sik Chan, Department of Statistics and Actuarial Science, University of Iowa

In high-dimensional time series, the component processes are often assembled into a matrix to display their interrelationship. We focus on detecting mean shifts with unknown change point locations in these matrix time series. Series that are activated by a change may cluster along certain rows (columns), which forms mode-specific change point alignment. Leveraging mode-specific change point alignments may substantially enhance the power for change point detection. We propose a powerful test to detect mode-specific change points, yet robust to non-mode-specific changes. We show the validity of using the multiplier bootstrap to compute the p-value of the proposed methods, and derive non-asymptotic bounds on the size and power of the tests. We also propose a parallel bootstrap — a computationally efficient variant that speeds up the double bootstrap for computing the p-value of the proposed adaptive test. In particular, we show the consistency of the proposed test, under mild regularity conditions. To obtain the theoretical results, we derive new, sharp bounds on Gaussian approximation and multiplier bootstrap approximation, which are of independent interest. We showcase the efficacy of the proposed method with numerical illustration. The talk is based on joint work with Dr. Xinyu Zhang.

Exploring Excellence: Bayesian Penalized Empirical Likelihood and MCMC Sampling

Jinyuan Chang, School of Statistics, Southwestern University of Finance and Economics and Chinese Academy of Sciences

In this study, we introduce a novel methodological framework known as Bayesian penalized empirical likelihood, designed to tackle the computational challenges associated with empirical likelihood methods. Our approach pursues two primary objectives: firstly, preserving the inherent flexibility of empirical likelihood to accommodate a wide range of model conditions, and secondly, providing convenient access to well-established Markov chain Monte Carlo (MCMC) sampling schemes. To achieve the first objective, we propose a penalized approach that effectively selects model conditions by regulating Lagrange multipliers, thereby reducing the dimensionality of the problem while leveraging a comprehensive set of model conditions. For the second objective, our approach overcomes the obstacles inherent in devising sampling schemes for Bayesian applications through efficient dimensionality reduction. Our Bayesian penalized empirical likelihood framework offers a flexible and efficient approach, enhancing the adaptability and practicality of empirical likelihood methods in statistical inference. Furthermore, our study illustrates the practical advantages of utilizing sampling techniques over optimization methods, as they exhibit rapid convergence to global optima of posterior distributions, ensuring robust parameter estimation. This framework provides a valuable tool for researchers and analysts grappling with complex problems.

A Bayesian Framework for Fine-Tuning Large Diffusion Models

Jian Huang, Department of Applied Mathematics,
The Hong Kong Polytechnic University

Diffusion-based generative models have achieved remarkable successes in learning complex probability measures for various types of data, including image, video, audio, and bioimaging data. Researchers have taken steps to fine tune pre-trained large-scale models with a significantly reduced amount of data, enabling them to generate samples that align with the dataset's support and achieve comparable quality. Additionally, these specific datasets include paired information, e.g., low resolution image, edge maps, segmentation maps, image space maps, cross attention maps etc., which are often treated as conditioning on the image generation process. The combination of learnable modules and large models has shown impressive generation capabilities. Therefore, it is useful to understand how fine-tuning transitions from "a large probability space" to "a small probability space" based on conditional controls. In this work, we formulate a Bayesian framework for fine-tuning in the context of large diffusion models. We clarify the meaning behind transitioning from a "large probability space" to a "small probability space" and explore the task of fine-tuning pre-trained models using learnable modules from a Bayesian perspective.

Demystifying Stable Diffusion

Yuling Jiao, School of Mathematics and Statistics, Wuhan University

Generative learning through diffusion models in latent space is the key to the success of Stable Diffusion, a product developed by Stability AI. In this presentation, we will provide an in-depth error analysis that encompasses the following aspects: (1) the error of a pre-trained encoder-decoder model under domain shift, (2) the error of score estimation using transformers, and (3) the sampling error of discretizing the stochastic differential equation. Additionally, we will explore an alternative approach using ordinary differential equation flows in latent space and present numerical results along with error analysis.

Online local differential private quantile inference via self-normalization

Linglong Kong, Department of Mathematical and Statistical Sciences, University of Alberta

Based on binary inquiries, we developed an algorithm to estimate population quantiles under Local Differential Privacy (LDP). By self-normalizing, our algorithm provides asymptotically normal estimation with valid inference, resulting in tight confidence intervals without the need for nuisance parameters to be estimated. Our proposed method can be conducted fully online, leading to high computational efficiency and minimal storage requirements with $\big{\mathcal{O}}{1}$ space. We also proved an optimality result by an elegant application of one central limit theorem of Gaussian Differential Privacy (GDP) when targeting the frequently encountered median estimation problem. With mathematical proof and extensive numerical testing, we demonstrate the validity of our algorithm both theoretically and experimentally.

Randomization Inference When N = 1

Tengyuan Liang, Booth School of Business, The University of Chicago

Neyman's seminal paper in 1923, which introduced the potential outcome framework and the analysis of randomized experiments, has arguably laid the foundation of causal inference for cross-sectional data. For time-series data, the framework of randomization inference is far less well-understood due to the interference: the potential outcomes at a particular time typically depend on treatments assigned before that time. Motivated by the literature of N-of-1 trials in clinical research and sequential AB testing in online marketing, in this talk, we study randomization experiments and causal inference when N = 1, borrowing insights from system identification and probability theory. The talk is based on joint work with Benjamin Recht (UC Berkeley).

Functional data analysis with covariate-dependent mean and covariance structures

Huazhen Lin, Center of Statistical Research and School of Statistics, Southwestern University of Finance and Economics

Functional data analysis has emerged as a powerful tool in response to the ever-increasing resources and efforts devoted to collecting information about response curves or anything varying over a continuum. However, limited progress has been made to link the covariance structure of response curves to external covariates, as most functional models assume a common covariance structure. We propose a new functional regression

model with covariate-dependent mean and covariance structures. Particularly, by allowing the variances of the random scores to be covariate-dependent, we identify eigenfunctions for each individual from the set of eigenfunctions which govern the patterns of variation across all individuals, resulting in high interpretability and prediction power. We further propose a new penalized quasi-likelihood procedure, which combines regularization and B-spline smoothing, for model selection and estimation, and establish the convergence rate and asymptotic normality for the proposed estimators. The utility of the method is demonstrated via simulations as well as an analysis of the Avon Longitudinal Study of Parents and Children on parental effects on the growth curves of their offspring, which yields biologically interesting results. Thai is joint work with Chenlin Zhang, Li Liu, Jin Liu and Yi Li.

Survival Estimation with Time-Varying Covariates Using Neural Networks

Bin Nan, Department of Statistics, University of California, Irvine

Most work in neural networks focuses on estimating the conditional mean of a continuous response variable given a set of covariates. In this talk, we consider estimating the conditional distribution function using neural networks for censored survival data. The algorithm is built upon the data structure particularly constructed for the Cox regression with time-dependent covariates. Without imposing any model assumption, we consider a loss function that is based on the full likelihood where the conditional hazard function is the only unknown parameter, for which unconstraint optimization methods can be applied. Through simulation studies, we show the proposed method possesses desirable performance, whereas the partial likelihood method yields biased estimates when model assumptions are violated.

Statistical Learning in Biological Neural Networks

Johannes Schmidt-Hieber, Department of Applied Mathematics, Universiteit Twente

Compared to artificial neural networks (ANNs), the brain learns faster, generalizes better to new situations and consumes much less energy. ANNs are motivated by the functioning of the brain but differ in several crucial aspects. For instance, ANNs are deterministic while biological neural networks (BNNs) are stochastic. Moreover, it is biologically implausible that the learning of the brain is based on gradient descent. In this talk we look at biological neural networks as a statistical method for supervised learning. We relate the local updating rule of the connection parameters in BNNs to a zero-order optimization method and derive some first statistical risk bounds.

Boosting Data Analytics with Synthetic Volume Expansion

Xiaotong Shen, School of Statistics, University of Minnesota

Synthetic data generation heralds a paradigm shift in data science, addressing the challenges of data scarcity and privacy and enabling unprecedented performance. As synthetic data gains prominence, questions arise regarding the accuracy of statistical methods compared to their application on raw data. Addressing this, we introduce the Synthetic Data Generation for Analytics framework, which applies statistical methods to high-fidelity synthetic data produced by advanced generative models like tabular diffusion models. These models, trained using raw data, are enriched with insights from relevant studies. A significant finding within this framework is the generational effect: the error of a

statistical method initially decreases with the integration of synthetic data but may subsequently increase. This phenomenon, rooted in the complexities of replicating raw data distributions, introduces the "reflection point," an optimal threshold of synthetic data defined by specific error metrics. Through three case studies--sentiment analysis, predictive modelling, and inference of tabular data, we demonstrate the effectiveness of this framework.

Network Gradient Descent Algorithm for Decentralized Federated Learning

Hansheng Wang, Guanghua School of Management, Peking University

We study a fully decentralized federated learning algorithm, which is a novel gradient descent algorithm executed on a communication-based network. For convenience, we refer to it as a network gradient descent (NGD) method. In the NGD method, only statistics (e.g., parameter estimates) need to be communicated, minimizing the risk of privacy. Meanwhile, different clients communicate with each other directly according to a carefully designed network structure without a central master. This greatly enhances the reliability of the entire algorithm. Those nice properties inspire us to carefully study the NGD method both theoretically and numerically. Theoretically, we start with a classical linear regression model. We find that both the learning rate and the network structure play significant roles in determining the NGD estimator's statistical efficiency. The resulting NGD estimator can be statistically as efficient as the global estimator, if the learning rate is sufficiently small and the network structure is well balanced, even if the data are distributed heterogeneously. Those interesting findings are then extended to general models and loss functions. Extensive numerical studies are presented to corroborate our theoretical findings. Classical deep learning models are also presented for illustration purposes.

On Generative Agents in Recommendation

Xiang Wang, School of Data Science/ Cyber Science and Technology/ Information Science and Technology, University of Science and Technology of China

Recommender systems are the cornerstone of today's information dissemination, yet a disconnect between offline metrics and online performance greatly hinders their development. Addressing this challenge, we envision a recommendation simulator, capitalizing on recent breakthroughs in human-level intelligence exhibited by Large Language Models (LLMs). We propose Agent4Rec, a novel movie recommendation simulator, leveraging LLM-empowered generative agents equipped with user profile, memory, and actions modules specifically tailored for the recommender system. In particular, these agents' profile modules are initialized using the MovieLens dataset, capturing users' unique tastes and social traits; memory modules log both factual and emotional memories and are integrated with an emotion-driven reflection mechanism; action modules support a wide variety of behaviors, spanning both taste-driven and emotion-driven actions. Each agent interacts with personalized movie recommendations in a page-by-page manner, relying on a pre-implemented collaborative filtering-based recommendation algorithm. We delve into both the capabilities and limitations of Agent4Rec, aiming to explore an essential research question: to what extent can LLM-empowered generative agents faithfully simulate the behavior of real, autonomous humans in recommender systems? Extensive and multi-faceted evaluations of Agent4Rec highlight both the alignment and deviation between agents and user-personalized preferences. Beyond mere performance comparison, we explore insightful experiments, such as

emulating the filter bubble effect and discovering the underlying causal relationships in recommendation tasks.

Green's matching: an efficient approach to parameter estimation in complex dynamic systems

Xueqin Wang, School of Management, University of Science and Technology of China

Parameters of differential equations are essential to characterize the intrinsic behaviors of dynamic systems. Many scientific challenges are hindered by a lack of computational and statistical efficiency in parameter estimation of dynamic systems, especially for complex systems with general-order differential operators, such as motion dynamics. Aiming at discovering these dynamic systems behind noisy data, we develop a computationally tractable and statistically efficient two-step method called Green's matching via estimating equations. Particularly, we avoid time-consuming numerical integration by the pre-smoothing of trajectories in the estimating equations, and the pre-smoothing of curve derivatives is generally not involved in the estimating equations due to the inversion of differential operators by Green's functions. These appealing features improve both computational and statistical efficiency for parameter estimation. We prove that Green's matching attains statistically optimal convergence for general-order systems. While for the other two widely used two-step methods, their estimation biases may dominate the estimation errors, resulting in poor convergence rates for high-order systems. We conduct extensive simulations to examine the estimation behaviors of two-step methods and other competitive approaches. Our results show that Green's matching outperforms other methods for parameter estimation, which also supports Green's matching in more complicated statistical inferences, such as equation discovery or causal network inference, for general-order dynamic systems.

Reinforcement Learning in a Continuous-Time Stochastic Setting

Yazhen Wang, Department of Statistics, University of Wisconsin-Madison

Reinforcement learning was developed mainly for discrete-time Markov decision processes. We consider solving reinforcement learning problems in a continuous-time setting with noisy data or jumps, where the true model to learn is governed by a jump-diffusion model with data sampled from the diffusion model, or the true model is an ordinary differential equation, and data samples are generated from a stochastic differential equation that is considered as a noisy version of the ordinary differential equation. Learning developed for deterministic or diffusion models is unstable and in fact diverges when applied to data generated from stochastic or jump-diffusion models. Furthermore, because there are measurement errors or jumps in the observed data, new reinforcement learning frameworks are needed to handle the learning problems with noisy data or jumps. We established an asymptotic theory for the proposed approaches and carried out numerical studies to check the finite sample performance of the proposed methods.

Inference for discrete or non-numerical parameters and unraveling machine learning black-box models

Minge Xie, Department of Statistics, Rutgers University

Rapid data science developments require us to have innovative frameworks to tackle frequently seen, but highly non-trivial "irregular inference problems" in machine learning models, e.g., those involving discrete or non-numerical parameters and those involving non-numerical data, etc. This talk presents an effective and wide-reaching framework, called repro samples method, to conduct statistical inference for the irregular problems and more. We develop theories to support our development and provide effective computing algorithms for problems in which explicit solutions are not available. The method is likelihood-free and is particularly effective for irregular inference problems. For commonly encountered irregular inference problems that involve discrete or nonnumerical parameters, we propose a three-step procedure to make inferences for all parameters and develop a unique matching scheme that turns the disadvantage of lacking theoretical tools to handle discrete/nonnumerical parameters into an advantage of improving computational efficiency. The effectiveness of the proposed method is illustrated through case studies by solving two highly nontrivial problems in statistics: a) how to quantify the uncertainty in the estimation of the unknown number of components and make inference for the associated parameters in a Gaussian mixture; b) how to quantify the uncertainty in model estimation and construct confidence sets for the unknown true model, the regression coefficients, or both true model and coefficients jointly in high dimensional regression models. The method also has extensions to complex machine learning models, e.g., (ensemble) tree models, deep neural networks, graphical models, etc. It provides a new toolset for unraveling the black box issues in machine learning models.

Hypotheses testing of functional principal components

Lijian Yang, Center for Statistical Science & Department of Industrial Engineering
Tsinghua University

We propose a test for the hypothesis that the standardized functional principal components (FPCsI of a functional data equal a given set of orthonormal basis (e.g., the Fourier basis). Using estimates of individual trajectories that satisfy certain approximation conditions, a chi-square type statistic is constructed and shown to be oracally efficient under the null hypothesis in the sense that its limiting distribution is the same as an infeasible statistic using all trajectories, known by "oracle". The null limiting distribution is an infinite Gaussian quadratic form, and a consistent estimator of its quantile is obtained. A test statistic based on the chi-square type statistic and approximate quantile of the Gaussian quadratic form is shown to be both of the nominal asymptotic significance level and asymptotically correct. It is further shown that B-spline trajectory estimates meet the required approximation conditions. Simulation studies illustrate superior finite sample performance of the proposed testing procedure. For the EEG (Electroencephalogram) data, the proposed procedure has confirmed an interesting discovery that the centered EEG data is generated from a small number of elements of the standard Fourier basis.

Blind source separation over space

Qiwei Yao, Department of Statistics, London School of Economics and Political Science

We propose a new estimation method for the blind source separation model of Bachoc et al. (2020). The new estimation is based on an eigen analysis of a positive definite matrix defined in terms of multiple spatial local covariance matrices, and, therefore, can handle moderately high-dimensional random fields. The consistency of the estimated mixing matrix is established with explicit error rates even when the eigen-gap decays to 0 slowly. The proposed method is illustrated via both simulation and a real data example.

Nonlinear Sufficient Dimension Reduction

Zhou Yu, Faculty of Economics and Management, East China Normal University

Linear sufficient dimension reduction, as exemplified by sliced inverse regression, has seen substantial development in the past thirty years. However, with the advent of more complex scenarios, nonlinear dimension reduction has become a more general topic that gains considerable interest recently. This article introduces a novel method for nonlinear sufficient dimension reduction, utilizing the generalized martingale difference divergence measure in conjunction with deep neural networks. The optimal solution of the objective function is shown to be unbiased at the general level of σ-fields. And two optimization schemes considered, based on the fascinating deep neural networks, exhibit higher efficiency and flexibility compared to the classical eigen decomposition of linear operators. Moreover, we systematically investigate the slow rate and fast rate for the estimation error based on advanced U-process theory. Remarkably, the fast rate is nearly minimax optimal. The effectiveness of the deep nonlinear sufficient dimension reduction methods is demonstrated through simulations and real data analysis.

On the multiway PCA

Ming Yuan, Department of Statistics, Columbia University

Multiway data are becoming more and more common. While there are many approaches to extending principal component analysis (PCA) from usual data matrices to multiway arrays, their conceptual differences from the usual PCA, and the methodological implications of such differences remain largely unknown. This work aims to specifically address these questions. In particular, we clarify the subtle difference between PCA and singular value decomposition (SVD) for multiway data, and show that multiway principal components (PCs) can be estimated reliably in absence of the eigengaps required by the usual PCA, and in general much more efficiently than the usual PCs. Furthermore, the sample multiway PCs are asymptotically independent and hence allow for separate and more accurate inferences about the population PCs. The practical merits of multiway PCA are further demonstrated through numerical, both simulated and real data, examples.

DreamRec：Reshaping Sequential Recommendation Systems

Yancheng Yuan, Department of Applied Mathematics
The Hong Kong Polytechnic University

Sequential recommendation aims to recommend the next item that matches a user's interest, based on the sequence of items he/she interacted with before. Scrutinizing previous studies, we can summarize a common learning-to-classify paradigm — given a positive item, a recommender model performs negative sampling to add negative items and learns to classify whether the user prefers them or not, based on his/her historical interaction sequence. Although effective, we reveal two inherent limitations: (1) it may differ from human behavior in that a user could imagine an oracle item in mind and select potential items matching the oracle; and (2) the classification is limited in the candidate pool with noisy or easy supervision from negative samples, which dilutes the preference signals towards the oracle item. Yet, generating the oracle item from the historical interaction sequence is mostly unexplored. To bridge the gap, we reshape sequential recommendation as a learning-to-generate paradigm, which is achieved via a guided diffusion model, termed DreamRec. To better understand the generated oracle items, we leverage the power of Large Language Models by designing a novel residual prompting learning mechanism. Numerical results on large recommendation datasets demonstrate the superior performance of our proposed generative sequential recommendation paradigm.

Creating High-Fidelity Digital Avatar for Everyone

Juyong Zhang, School of Mathematical Sciences,
University of Science and Technology of China

Traditional digital human modeling and animation methods rely on expensive acquisition equipment, complex production processes, and a large number of manual interactions by professional staff, which greatly limit its wide applications. The 3DV group of USTC has conducted research on the aspect of monocular camera based high-fidelity digital human modeling and animation toward the target of "digitalize everyone in the world". In this talk, I will share our research work about: high-fidelity 3D head modeling, audio-driven talking head, clothed human modeling and animation, and text prompt-based avatar editing.

Testing for independence and measuring the degree of nonlinear associations

Liping Zhu, Institute of Statistics and Big Data, Renmin University of China

Modern statistical sciences start with metrics and bloom from measuring nonlinear associations. Testing for independence and measuring the degree of nonlinear associations are two fundamental problems in statistics and the keys to measuring prediction power. Predictions lie in the hearts of both statistics and machine learning. We will briefly review many existing metrics in the literature and introduce some most recent progresses.

Posters

Yu Chen, Jian Huang, Guohao Shen, and Xingqiu Zhao:
Model-Free Variable Selection by Deep Differential Neural Networks.

Wenhai Cui, Wen Su, Xiaodong Yan, and Xingqiu Zhao:
Demographic Parity-Aware Individualized Treatment Rules.

Yuan Gao, Jian Huang, and Yuling Jiao:
Gaussian Interpolation Flows: Demystifying Gaussian Denoising for Generative Learning.

Bingyao Huang, **Fangran Miao**, Ting Li, and Jian Huang:
Fair Adaptive and Robust Federated Learning.

Ding Huang, Jian Huang, Ting Li, and Guohao Shen:
Conditional Stochastic Interpolation for Generative Learning.

Yuanhang Luo, Ruijian Han, and Jian Huang:
Online Inference in High-Dimensional Models.

Hoi Min Ng and Kin Yau Wong:
A Global Kernel Estimator for Partially Linear Varying Coefficient Additive Hazards Models.

Shengjie Niu, Lifan Lin, Chao Wang, and Jian Huang:
Confident Self-Labeling for Open-World Classification.

Chenyu Ren, Daihai He, and Jian Huang:
Hierarchical Functional Protein Generation with Latent Representation.

Maojun Sun:
LlamaCare: A Large Medical Language Model for Enhancing Healthcare Knowledge Sharing.

Xinyang Yu, Binyan Jiang, Chenlei Leng, Ting Yan, and Qiwei Yao:
A Two-Way Heterogeneity Model for Dynamic Networks.

Yangzi Zheng and Binyan Jiang:
A Semiparametric Model for Zero-Inflated Fraction Data.

Xueyu Zhou and Jian Huang:
Deep Sufficient and Invariant Representation Learning.

Transportation Guidance

来港交通

达到香港有三种路径:香港国际机场，香港西九龙高铁，深圳香港口岸。

- **香港国际机场入境**

  - **出租车:**香港有三种颜色的出租车(红色，绿色，蓝色)。请乘搭红色的出租车到达香港理工大学。全程大约30分钟，费用在300港币左右。搭乘出租车需要现金。
  - **机场巴士:**乘搭机场巴士A21，巴士的终点站在红磡地铁站，于A1或D1出口走行人天桥到校园，步行5分钟即可到达理工大学。乘搭巴士需要使用八达通卡，部分巴士也可以用支付宝。全程大概75分钟，费用在35港币左右。
  - **地铁：**首先从香港国际机场乘坐机场快线到青衣站。下一站于青衣站4号月台转乘东涌线往香港站，再于南昌站2号月台转乘屯马线往乌溪沙站。于红磡站下车，于A1或D1出口走行人天桥到校园。全程约需 60 分钟，费用为港币 65 元（使用八达通）/港币 81.5 元（不使用八达通）。

- **香港西九龙高铁站入境**

  从12306上可以购买从深圳北站/福田站出发到香港西九龙的高铁票，西九龙位于柯士甸站，距离理工大学所在的红磡站仅有俩站地铁。从西九龙高铁入境后，

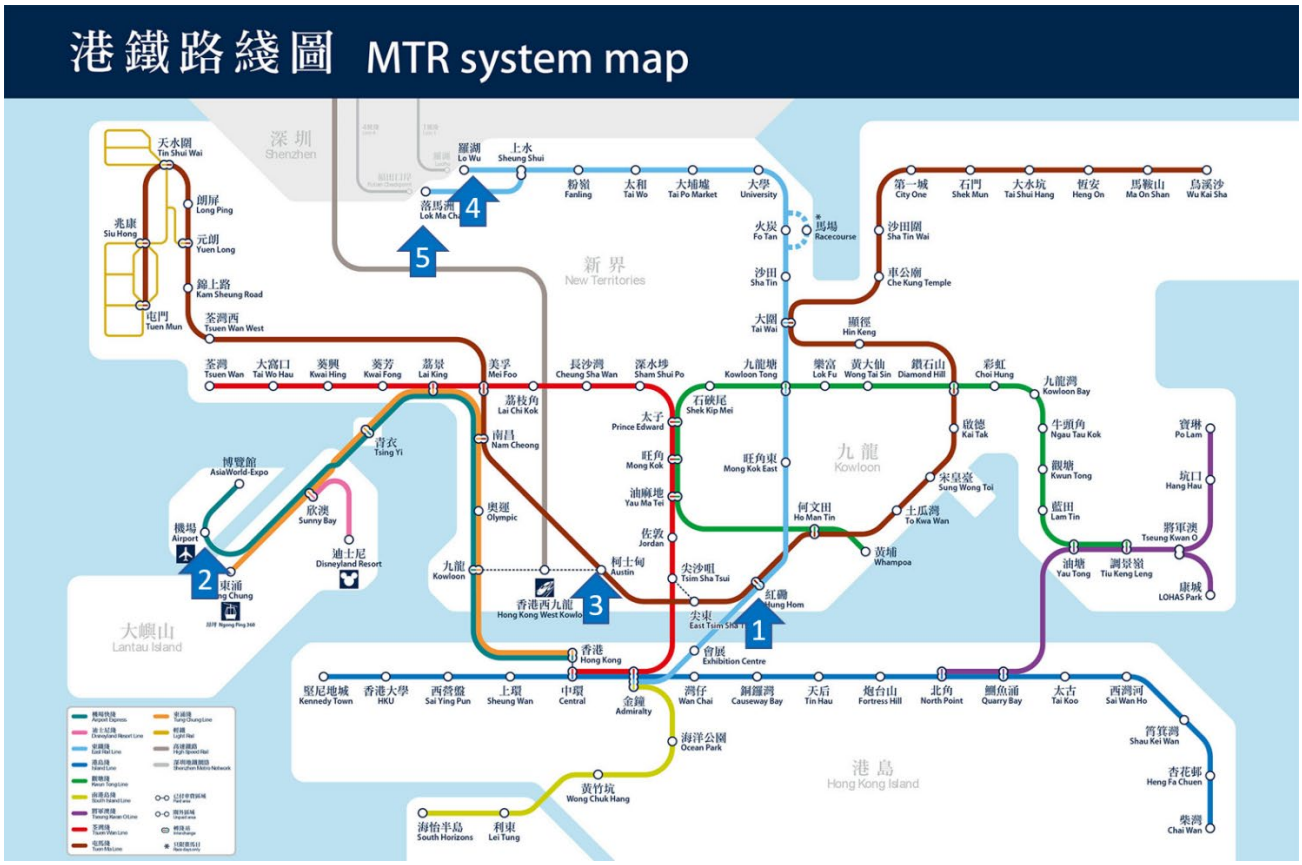
  - **出租车:**香港有三种颜色的出租车(红色，绿色，蓝色)。请乘搭红色的出租车到达香港理工大学。全程大约10分钟，费用在80港币左右。搭乘出租车需要现金。
  - **地铁:** 从柯士甸站出发，乘坐屯马线（乌溪沙方向）到达红磡地铁站，于A1或D1出口走行人天桥到校园。全程约需 20 分钟，费用约为港币 10 元。乘坐地铁需要八达通。

- **深圳/香港口岸入境**

  - **福田口岸/罗湖口岸：**乘坐东铁线（金钟方向）到达红磡地铁站，于A1或D1出口走行人天桥到校园。全程约需 60 分钟，费用约为港币 40 元。乘坐地铁需要八达通。
  - **深圳湾口岸：**
    - 出租车：请乘搭红色的出租车到达香港理工大学。全程大约30分钟，费用在300港币左右。搭乘出租车需要现金。

- **到达深圳机场再入境**

过境巴士：如果从深圳宝安机场到达，可以在机场购买过境巴士票，或者搜索微信小程序‘环岛中港通’，起点选在深圳宝安国际机场，终点选在香港尖沙咀海港城，每半个小时会有一班车（如下图）。乘坐过境巴士在中间会在深圳湾换乘一次。全程大概60分钟（不含等巴士的时间），费用为人民币 120 元。



1 ：红磡站，香港理工大学所在地

2 :香港国际机场

3 :柯士甸站，香港西九龙高铁站

4 : 罗湖口岸

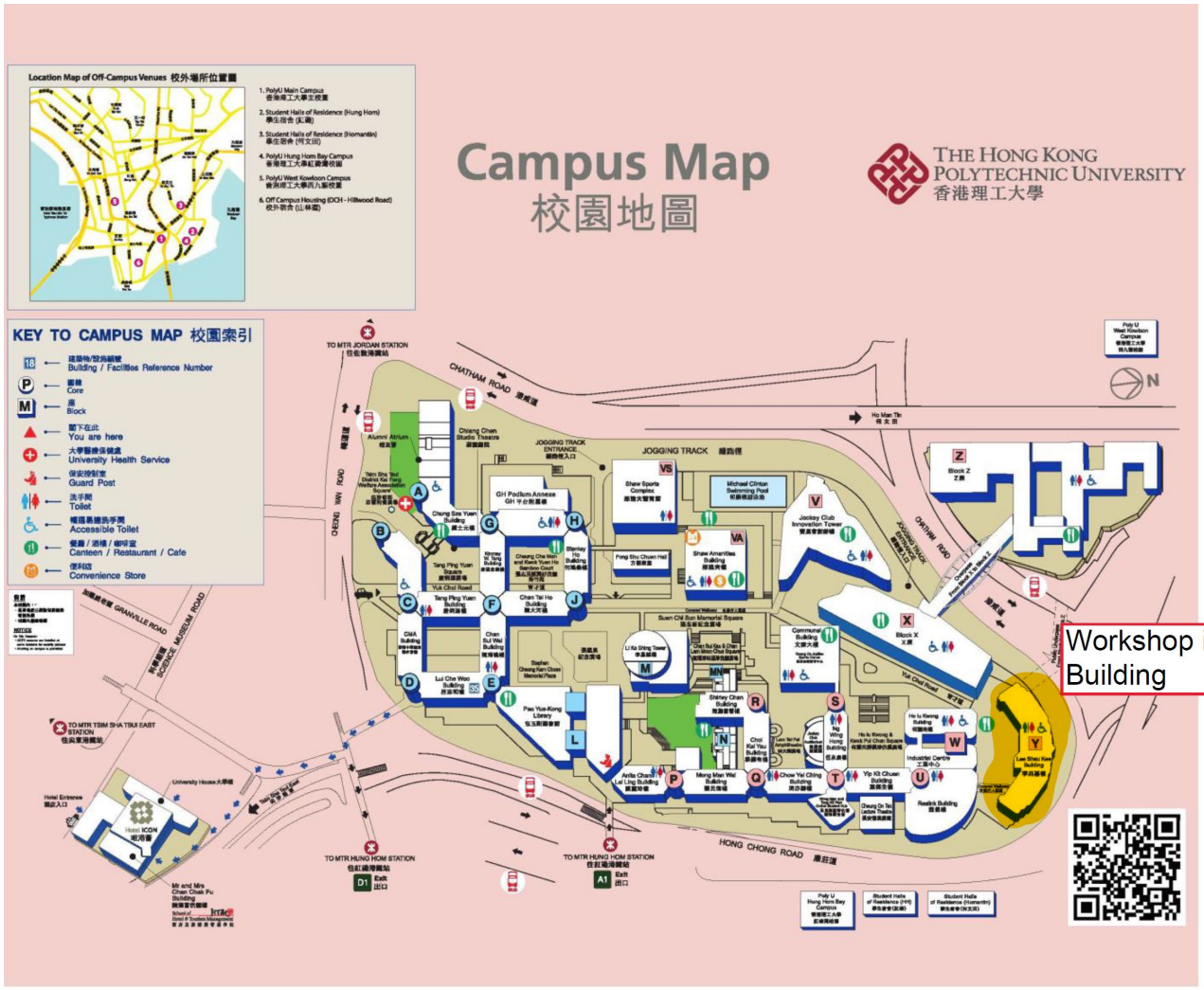
5 :福田口岸

紧急联系人

如果您在中途遇到了问题，请随时联系我们，可以通过微信或者直接拨打以下几位紧急联系人的电话：

韩睿渐博士 +852 60985400
李挺博士 +852 51259568
申国豪博士 +852 69106097
黄坚教授 +852 68718304

Campus Map


Campus Map
校園地圖
THE HONG KONG POLYTECHNIC UNIVERSITY 香港理工大學
Location Map of Off-Campus Venues 校外場所位置圖
KEY TO CAMPUS MAP 校園索引
Building / Facilities Reference Number
Core
Block
You are here
University Health Service
Guard Post
Toilet
Accessible Toilet
Canteen / Restaurant / Cafe
Convenience Store
Workshop in this Building