Workshop on Data Science and Deep Learning (DSDL2018)

Sponsored by Department of Statistics and Actuarial Science
under the Kin Lam Development Fund for Data Analytics

April 7, 2018 (Saturday)

Theatre P4, Chong Yuet Ming Physics Building
The University of Hong Kong

Programme and Abstracts

ORGANIZING COMMITTEE

Fei Jiang	(The University of Hong Kong, Hong Kong)
Guosheng Yin	(The University of Hong Kong, Hong Kong)
Philip L.H. Yu	(The University of Hong Kong, Hong Kong)
Aijun Zhang	(The University of Hong Kong, Hong Kong)

INVITED SPEAKERS

Chunlin Ji	Shenzhen Kuang-Chi Institute of Advanced Technology
Rick Jin	Tencent, Shenzhen
Di Lin	Department of Computer Science, Shenzhen University
Michael Ng	Department of Mathematics, Hong Kong Baptist University, Hong Kong
Zhihong Xia	Department of Mathematics, Southern University of Science and Technology
Xin Yao	Department of Computer Science, Southern University of Science and Technology
Philip L.H. Yu	Department of Statistics and Actuarial Science, HKU
Yizhou Yu	Department of Computer Science, HKU
Aijun Zhang	Department of Statistics and Actuarial Science, HKU

ACKNOWLEDGEMENT

The Organising Committee would like to extend deep gratitude to Professor Kin Lam for pledging to establish the Kin Lam Development Fund for Data Analytics. The Fund is to support staff development, simulate research and collaborative initiatives in the area of data analysis. Professor Lam is currently an Honorary Professor and also a long-time patron of the Department of Statistics and Actuarial Science, The University of Hong Kong.

GENERAL INFORMATION

Registration & Enquiry Desk

The Registration & Enquiry Desk will operate at the following time and venue:

Date	Time	Venue
Saturday, 7 April	09:00 – 09:30	Theatre P4, Chong Yuet Ming Physics Building HKU

If participants require urgent assistance outside the operating times of the Registration & Enquiry Desk, they may visit the Department of Statistics and Actuarial Science, Room 303, 3/F, Run Run Shaw Building.

Opening Ceremony

Date: Saturday, 7 April, 2018
Time: 09:30 – 09:35
Venue: Theatre P4, Chong Yuet Ming Physics Building, HKU

Lunches & Dinner

Lunches and Dinner are by invitation only. You can find the details of canteens from the HKU food map.

Name Badges

Each participant will be issued a name badge upon registration. The badge will be the official pass to talks and coffee breaks. As access to events will be strictly controlled, it would be appreciated if you could wear your badge at all times. If any participants have lost their badges, they can contact the Registration & Enquiry Desk for a replacement.

Announcements and Notices

A message board will be placed near the Registration & Enquiry Desk for announcements and important notices.

Internet Facilities

HKU provides free WiFi service for visitors of the University to surf the Internet for research or administrative purposes on campus. Connection to Wi-Fi.HK via HKU is easy and no registration is required. However, Wi-Fi.HK via HKU has limited capacity and provides web browsing service only. The service uses unencrypted channel and hence, may not be secure. To connect PC/mobile device to Wi-Fi.HK via HKU service, please find the procedure here: http://www.its.hku.hk/documentation/guide/network/wifi/openwifi.

Car Parking

Complimentary parking is available on campus. Please take the parking ticket to the Registration & Enquiry Desk for arranging complimentary parking.

No Smoking Policy

Smoking is prohibited in all areas of the campus, including corridors and restrooms.

Safety and Security

Please do not leave your belongings unattended at any times inside or outside the venue.

Disclaimer of Liability

The Organizer of the DSDL2018 will not accept any liability for damages of any nature sustained by participants or their accompanying persons, or loss of or damages to their personal property during the workshop or any related events.

The Secretariat

For all enquiries, please contact the Secretariat of the Workshop on Data Science and Deep Learning:

The Secretariat, DSDL2018
Department of Statistics and Actuarial Science
Room 303, 3/F, Run Run Shaw Building
The University of Hong Kong
Pokfulam Road, Hong Kong
Telephone: (852) 3917-8312
Facsimile: (852) 2858-9041
E-mail: saas@hku.hk

Conference venue and HKU food map

Workshop on Data Science and Deep Learning (DSDL2018)

PROGRAMME

April 7, 2018 (Saturday)

Morning session	
09:00 – 09:30	Registration
09:30 – 09:35	Opening Address
09:35 – 10:05	***Big Data, a Golden Opportunity for Mathematics*** Zhihong Xia Department of Mathematics, Southern University of Science and Technology, China
10:05 – 10:35	***Role of Statistical Metric in Big Industrial Data and Artificial Intelligence*** Chunlin Ji Vice President, Kuang-Chi Institute of Advanced Technology, Shenzhen, China
10:35 – 11:10	Photo Taking and Coffee Break
11:10 – 11:40	***Tensor Models for Data Analysis*** Michael Ng Department of Mathematics, Hong Kong Baptist University, Hong Kong
11:40 – 12:10	***Deep Learning for Image Recognition*** Yizhou Yu Department of Computer Science, HKU

12:10 – 14:00 Lunch

April 7, 2018 (Saturday)

Afternoon session	
14:00 – 14:30	***Ensemble Approaches in Machine Learning*** Xin Yao Department of Computer Science, Southern University of Science and Technology, China
14:30 – 15:00	***Learning Deep Representation for Image Recognition*** Di Lin Department of Computer Science, Shenzhen University, China
15:00 – 15:30	***Forecasting High-Dimensional Covariance Matrices Using Deep Learning*** Philip L.H. Yu Department of Statistics and Actuarial Science, HKU
15:30 – 16:00	Coffee Break
16:00 – 16:30	***Profiling Users from Online Social Behaviors, with Applications in Tencent Social Ads*** Rick Jin Director of Ads Quality R&D Center, Social & Performance Ads Department, Tencent, Shenzhen, China
16:30 – 17:00	***AI-powered Medical Imaging Analytics via Deep Learning*** Aijun Zhang Department of Statistics and Actuarial Science, HKU
17:00 – 17:05	Closing Remarks

18:00 Workshop Dinner

Workshop on Data Science and Deep Learning (DSDL2018)

ABSTRACTS

Chunlin Ji

Vice President, Kuang-Chi Institute of Advanced Technology, Shenzhen, China

Role of Statistical Metric in Big Industrial Data and Artificial Intelligence

This talk discusses a statistical metric called energy distance that measures the distance between distributions of random vectors, as well as its applications in a) metamaterial design based on big industrial data, and b) video tracking in the field of intelligent security. By incorporating the energy distance in approximate Bayesian computation (ABC) framework, we show that the energy distance is more informative than traditional summary statistics in most cases, and the rejection rule based on equal distribution test is easier to implement than traditional ABC methods. We apply the proposed algorithm to the reversal design of metamaterial, and obtain the improved efficiency. We may also apply the energy distance to discriminant analysis in machine learning. Compared to linear discriminant analysis based on the cross-variance loss, the energy discriminant analysis has more robust performance in handling some complex data, thus it can be applied to practical surveillance system with complex scenarios.

Rick Jin

Director of Ads Quality R&D Center, Social & Performance Ads Department, Tencent, Shenzhen, China

Profiling Users from Online Social Behaviors, with Applications in Tencent Social Ads

QQ and Wechat are the two largest instant messaging/social networks in China. Tencent Social Ads is the advertising system for both Wechat and QQ, serving well over 10B page views per day, for hundred million daily users. We strive to understand as much as possible on our users' multiple aspects, so as to serve the best personalized ads for them. The rich user behaviors on Tencent's many products lay a solid foundation in user profiling. We develop audience targeting on many dimensions, including demographics, interests, intents, transactions, physical locations, and access environment, etc. In this presentation, we will share our experience in large-scale user data mining for audience targeting, and discuss the challenges we face and the solutions we have employed.

Di Lin

Department of Computer Science, Shenzhen University, China

Learning Deep Representation for Image Recognition

Image classification and semantic segmentation have been long standing topics in high-level computer vision. They emphasize understanding of scene context and object. Towards image classification, we address the problem regarding appropriate representation of the scene structure and differentiation among object categories. Compared to image classification, semantic segmentation task faces more challenges. Not only the global information of the image, but also the underlying relationship among pixels need to be captured by the image representation. Therefore large-scale data is vital to enable reliable learning of representation. We propose a weakly-supervised learning system for semantic segmentation, which can be used in practice to improve segmentation with cost-effective data.

Michael Ng

Department of Mathematics, Hong Kong Baptist University, Hong Kong

Tensor Models for Data Analysis

In this talk, we study the robust tensor completion problem which aims to learn a low-tubal-rank tensor from partial observations that are grossly corrupted. Numerical examples are presented to illustrate our theoretical results for the robust tensor completion problem.

Zhihong Xia

Department of Mathematics, Southern University of Science and Technology, China

Big Data, a Golden Opportunity for Mathematics

Recent development for data science and artificial intelligence brought tremendous challenges and opportunities for mathematical research. Mathematicians are slow to join in the exciting adventure. In this talk, we will discuss the data science from mathematical perspectives, and present an amateurish journey for a mathematician trying to understand machine learning and AI.

Xin Yao

Department of Computer Science, Southern University of Science and Technology, China

Ensemble Approaches in Machine Learning

Ensemble learning can be seen as an automatic approach towards divide-and-conquer in machine learning. Instead of trying to construct a large and complex monolithic learning system, it is sometimes more appropriate to construct a group of learning machines, each of which focuses on solving one aspect of a complex problem. This talk introduces my personal biased view towards ensemble learning, highlights its link with evolutionary computation (especially in multi-objective learning), and describes some recent ensemble approaches to online learning and class imbalance learning. Diversity in ensemble learning is singled out as one of the most important factors in determining the generalisation performance of ensemble learning.

Philip L.H. Yu

Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong

Forecasting High-Dimensional Covariance Matrices Using Deep Learning

Modeling and forecasting covariance matrices of asset returns play a crucial role in finance. The availability of high frequency intraday data enables the modeling of the realized covariance matrix directly. However, most models in the literature depend on strong structural assumptions and they also suffer from the curse of dimensionality. To solve the problem, we propose a deep learning model which treats each realized covariance matrix as an image. The network structure is designed with simplicity in mind, and yet provides superior accuracy compared with several advanced statistical methods. The model could handle both low-dimensional and high-dimensional realized covariance matrices.

Yizhou Yu

Department of Computer Science, The University of Hong Kong, Hong Kong

Deep Learning for Image Recognition

With recent rapid advances in deep learning, there have been major breakthroughs in computer vision, one of the core subareas of artificial intelligence. In this talk, I present representative computer vision works from my research group with a focus on object recognition and scene understanding. Specifically, I present deep learning algorithms for image classification, RGB-D scene labeling as well as salient object detection. For image classification, I introduce a deep transfer learning scheme, called selective joint fine-tuning, for boosting the performance of tasks with insufficient training data. In this scheme, a target learning task with insufficient training data is carried out simultaneously with another source learning task with abundant

training data. However, the source learning task only uses a subset of training data most relevant to the target learning task. Experiments demonstrate that our deep transfer learning scheme achieves state-of-the-art performance on multiple visual classification tasks with insufficient training data for deep learning.

Aijun Zhang

Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong

AI-powered Medical Imaging Analytics via Deep Learning

Medical imaging is facing a fundamental AI (artificial intelligence) transformation in the sense that traditional radiologists will be replaced by automated machine algorithms when examining diagnostic images. In this talk we discuss the deep learning approach to medical imaging analytics, in the aspects of image segmentation and radiomic prediction. A real data example is presented with analysis of rectal tumor MR images for therapeutic effect evaluation.