WESTERN SYDNEY
UNIVERSITY

School of Computer, Data and Mathematical Sciences

HDR Seminar 7
Program
24 March 2021

4:00pm - 4:10pm	**Briefing** **Topic: WSU Co-funding Scholarships & Candidature Support Fund policy** A/Prof Dongmo Zhang
4:10pm - 4:40pm	**Invited speech** **Topic:** Online machine learning for dynamic data stream **Speaker:** A/Prof Alan Wee-Chung Liew, Institute for Integrated and Intelligent Systems, Griffith University
4:40pm - 5:00pm	**Conference paper presentation** **Topic:** Quranic Education and Technology: Reinforcement learning System for Non-Native Arabic Children **Speaker:** Bayan Alsharbi (PhD Candidate -18516887) **Supervisory panel:** Dr Omar Mubin and Dr Mauricio Novoa
5:00pm - 5:20pm	**CoC Practice** **Topic:** *Analogising Style Transfer for Author Document Attribution* **Speaker:** Stuart Fitzpatrick (MPhil Candidate 17718699) **Supervisory panel:** A/Prof Laurence Park and A/Prof Oliver Obst
5:20pm	**Closing**

Venue: Online Zoom
Zoom ID: 886 7872 4041

Online machine learning for dynamic data stream
Speaker: A/Prof Alan Wee-Chung Liew, Institute for Integrated and Intelligent Systems, Griffith University

Abstract: Online learning presents unique challenges to machine learning algorithms. Offline or batch learning algorithms, which assume that all training data are available at once, are not suited to dealing with data streams. Storing large volumes of data in the main memory is often not feasible, and traditional batch learning methods where learning is performed on the entire training set becomes impractical. Moreover, batch learning algorithms are not applicable in real-time learning scenarios where an infinite stream of data is continuously arriving, and predictions must be made on the fly. They are also incapable of adapting to dynamic environment.

A setting unique for online learning is the 1-by-1 prequential learning (i.e., test-then-train on the fly) over data stream in the following way: At time t, the sample $\mathbf{x}_t$ arrives. A classifier h examines the sample $\mathbf{x}_t$ and predicts a label $\hat{y}_t$ to the sample. The true label y_t of the sample is then revealed. If the classifier h gets the prediction wrong, learning is initiated on h using the sample and its true label $(\mathbf{x}_t, y_t)$ to improve its accuracy. The sample $\mathbf{x}_t$ is discarded afterward before the next sample arrive.

Some major challenges under this learning setting are:

Class imbalance: In data stream, the assumption that the data are equally distributed over all classes is often invalid. Also, the degree of the class imbalance is not known in advance and the imbalance classes can change with time in streaming data.

Concept Drift: Streaming data are often non-stationary, and it cannot be assumed that future data follow the same distribution as past and current data.

Class evolution: In class evolution, novel classes appear and grow, and the outdated classes fade away and disappear. Therefore, the underlying set of classes is not fixed beforehand and evolve with time.

Dirty data: Streaming data can be very noisy due to sensor failure or dropout.

Our work in online learning addresses some of the above challenges. In this seminar I will talk about some of our recent results in this area.

Biography: Alan Wee-Chung LIEW is currently the Deputy Head of School (Research) of the School of ICT, and the Duty Director of the Institute for Integrated and Intelligent Systems at Griffith University. Prior to joining Griffith University in 2007, he worked as a Senior Research Fellow at the City University of Hong Kong, and later an Assistant Professor with the Chinese University of Hong Kong. His research interest is in the field of AI and machine learning, medical imaging, computer vision, and bioinformatics. He has published extensively in these areas and is the author of 250+ book chapters, journal and conference papers, and holds two international patents. He is a senior member of IEEE and serves as associate editor on several journals, including the IEEE Transactions on Fuzzy Systems, one of the top journals on computational intelligence.

Quranic Education and Technology: Reinforcement learning System for Non-Native Arabic Children
Speaker: Bayan Alsharbi (PhD Candidate -18516887)

Abstract: We built a simulator based on reinforcement learning to improve teaching experience in Quranic and Islamic education for nonnative Arabic speakers to evaluate their strength and weaknesses and allow the system to help improving the child in one hand, and provide an accurate actual report for each child on the other hand.

Analogising Style Transfer for Author Document Attribution
Speaker: Stuart Fitzpatrick (MPhil Candidate 17718699)

Abstract: Since Lorenzo Valla wrote his `On the Donation of Constantine' in the 15th Century, the attribution and analysis of authorship has been an ever increasing and important problem.

By looking through the lens of statistical features first pioneered by Wincenty Lutoslawski and used by many others in the _elds of Linguistics and Natural Language Processing to this day, this project seeks to; develop a definition of literary style to that end, show how it leads to author attribution and finally how it can be used to transform a document's stylistic features, so that it appears another author wrote it.

In pursuit of these aims, this project will explore generative models, Latent Dirichlet Allocation, and Metric Learning in a boostrapping style manner across all three aims, so that advancements in the definition may also advance the various methods and vice versa.

As a result of this project, the field of Natural Language Processing will benefit from a consistent definition and new baseline results to compare to, other areas of Machine Learning will be able to analogise the techniques in

this project to advance their own work and outside Machine Learning the gap of understanding between the three groups that study literary style (linguists, philologists and data scientists) will be reduced.