大數據與小數據上的特徵學習
Representation Learning on Big and Small Data

時間：2018/03/10 (六) 上午09:30-11:00 (9:00報到)
演講人：張智威博士 / HTC健康醫療事業部總經理
　　　　Edward Y. Chang / President, HTC Research and Healthcare
地點：臺北科技大學宏裕科技研究大樓 B2國際會議廳

講者簡介 Biography

Edward Y. Chang currently serves as the President of Research and Healthcare (DeepQ) at HTC. Ed's most notable work is co-leading the DeepQ project (with Prof. CK Peng at Harvard), working with a team of physicians, scientists, and engineers to design and develop mobile wireless diagnostic instruments that can help consumers make their own reliable health diagnoses anywhere at anytime. The project entered the Tricorder XPRIZE competition in 2013 with 310 other entrants and was awarded second place in April 2017 with US$1M prize. DeepQ is powered by deep architecture to quest for cure. Similar deep architecture is also used to power Vivepaper, an AR product Ed's team launched in 2016 to support immersive augmented reality experience (for education, training, and entertainment).

Prior to his HTC post, Ed was a director of Google Research for 6.5 years, leading research and development in several areas including scalable machine learning, indoor localization, social networking and search integration, and Web search (spam fighting). His contributions in parallel machine learning algorithms and data-driven deep learning (US patents 8798375 and 9547914) are recognized through several keynote invitations and the developed open-source codes have been collectively downloaded over 30,000 times. His work on IMU calibration/fusion (US patents 8362949, 9135802, 9295027, 9383202, and 9625290) with project X was first deployed via Google Maps (see XINX paper and ASIST/ACM SIGIR/ICADL keynotes) and now widely used on mobile phones and VR/AR devices. Ed's team also developed the Google Q&A system (code-name Confucius), which was launched in over 60 countries.

Prior to Google, Ed was a full professor of Electrical Engineering at the University of California, Santa Barbara (UCSB). He joined UCSB in 1999 after receiving his PhD from Stanford University, and was tenured in 2003 and promoted to full professor in 2006. Ed has served on ACM (SIGMOD, KDD, MM, CIKM), VLDB, IEEE, WWW, and SIAM conference program committees, and co-chaired several conferences including MMM, ACM MM, ICDE, and WWW. He is a recipient of the NSF Career Award, IBM Faculty Partnership Award, and Google Innovation Award. He is a Fellow of IEEE for his contributions to scalable machine learning.

演講摘要 Abstract

The approaches in feature extraction can be divided into two categories: model-centric and data-driven. The model-centric approach relies on human heuristics to develop a computer model to extract features from data. These models were engineered by scientists and then validated via empirical studies. A major shortcoming of the model-centric approach is that unusual circumstances that a model does not take into consideration during its design can render the engineered features less effective. Contrast to the model-centric approach, which dictates representations independent of data, the data-driven approach learns representations from data. Example data-driven algorithms are multilayer perceptron (MLP) and convolutional neural network (CNN), which belong to the general category of neural network and deep learning. In this talk I will first explain why my team at Google in 2006 embarked on the data-drive approach. In 2010, we funded the ImageNet project at Stanford, and subsequently in 2011 filed two data-driven deep learning patents, one on feature extraction and the other on object recognition. We parallelized five widely used machine learning algorithms including SVMs, PFP, LDA, Spectral Clustering, and CNN, and open-sourced all these algorithms. I will present our latest work in accelerating CNN training using second order methods, and in reducing CNN model size. In the second half of this presentation, I will share our experience in the healthcare domain, where small data is the norm. I will discuss our experiences with transfer learning and GANs, both positive and negative ones. This talk concludes with a list of open research issues.

主辦單位：

台北科大
EMBA

國立臺北科技大學 管理學院EMBA專班
www.emba.ntut.edu.tw