

出國報告（出國類別：研習）

**參加 ACM SIGKDD 2016
「第 22 屆知識發現和資料探勘技術國際研習會」**

服務機關：行政院環保署

姓名職稱：魏文娟分析師

派赴國家：美國舊金山

出國期間：105 年 8 月 11 日~8 月 23 日

報告日期：105 年 10 月 17 日

目 錄

壹、會議源起與目的.....	2
貳、會議過程	5
8 月 13 日	5
8 月 14 日	6
8 月 15 日	6
8 月 16 日	6
8 月 17 日	7
參、心得及建議	8
肆、活動照片	13
附件一 會議議程	
附件二 贊助廠商	
附件三 最佳論文	
附件四 最佳學生論文	
附件五 最經時間考驗論文	
附件六 參加ACM SIGKDD 2016「第22屆知識發現和資料探勘 技術國際研習會」出國報告簡報檔	

參加ACM SIGKDD 2016

「第22屆知識發現和資料探勘技術國際研習會」

壹、會議源起與目的

「第 22 屆知識發現和資料探勘技術國際研習會」ACM SIGKDD 2016 乃由 「計算機協會知識發現和資料探勘專業組織」(ACM SIGKDD) 主辦，該組織核心成員主要為資料科學領域之學術、業界之領導人士，會議於 105 年 8 月 13 日至 8 月 17 日假舊金山聯合廣場希爾頓飯店舉行，本屆參與人數統計為 2726 人，多為各國資料科學研究人員、實作家及相關業者，其中更有 Microsoft、google、facebook、Linkedin、IBM、YAHOO、ebay、ADOBE、NEC、NVIDIA 等國際知名軟、硬體及社群網站公司，以及中國大陸騰訊、滴滴等通訊及電子商務公司之重量級人物出席及贊助。會議議程及贊助廠商詳如附件一、二。

茲列出本次會議網路相關資訊連結：

一、本次會議網站網址如下：

<http://www.kdd.org/kdd2016/>

二、精彩的 Keynote Speech (Youtube 影音連結)：

1. Graphons and Machine Learning : Modeling and Estimation of Sparse Massive Networks (Jennifer Chayes, Microsoft Research)
2. People, Computers, and The Hot Mess of Real Data (Joseph M. Hellerstein, Department of Electrical Engineering and Computer Sciences, UC Berkeley)
3. A VC View of Investing in ML (Greg Papadopoulos, NEA)
4. The Evolving Meaning of Information Security (Whitfield Diffie, Center for International Security and Cooperation, Stanford University) 2015 年圖靈獎得主

5. Learning to learn and compositionality with deep recurrent neural networks (Nando de Freitas, Department of Computer Science, University of Oxford)

三、座談 Plenary Panel : Is Deep Learning the New 42? (Youtube 影音連結) : moderator : Andrei Broder, Yahoo! Research
panelist : Pedro Domingos, Dept. of Computer Science & Engineering, University of Washington

四、本屆論文連結 :

<http://www.kdd.org/kdd2016/program/accepted-papers>

五、會議中的專題研討連結 :

<http://www.kdd.org/kdd2016/Workshops>

六、微軟學術推廣連結 :

<https://academic.microsoft.com/>

七、Youtube Channel : [kdd 2016 video](#)

八、KDD 2016 Opening Ceremony (Organizers : Shipeng Yu, LinkedIn Corporation; Mohak Shah, Bosch Research and Technology Center North America)

ACM SIGKDD「計算機協會知識發現和資料探勘專業組織」是全球性的資料科學社群，其宗旨在於鼓勵支持知識發掘基礎研究發展、應用實作、教育推廣，及協助相關標準之建立，以全面性研究資料探勘及機器學習核心問題、重視多元訓練及主導資料科學未來發展走向為己任，相信藉由人工智慧將改變世界樣貌，而人工智慧之提升乃資料探勘及從資料中學習所得之成果，深信當此人工智慧轉型時期，吾輩自當全力參與。

本屆 KDD 之特色包括：

- 安排 4 位 Keynote 及 1 位圖靈獎得主演講(Whitfield Diffie)

- 邀請 12 位資料應用科學家演講
- 安排 1 個 plenary panel 及 2 個 invited panels
- 錄取 142 篇研究論文(research track papers), 包括 70 篇口頭報告與 72 篇海報展示 (附件三收錄最佳論文、附件四收錄最佳學生論文、附件五收錄最經時間考驗論文 best test of time award)
- 錄取 66 篇應用論文(applied data science track paper)
- 舉辦 17 個研討會 workshops
- 舉辦 10 場教育訓練 tutorials
- 舉辦 8 場操作型教育訓練 hands-on tutorials
- 安排 3 場 social networking sessions
- 舉辦 KDD Cup 競賽

本屆大會於 Youtube 建立 KDD2016 video 頻道，陸續上傳資訊包括：

- 68 部口頭報告影片
- 30 部應用報告影片
- 20 部操作型教育訓練影片
- 27 部教育訓練影片
- 3 部開幕儀式及頒獎
- 13 部資料應用科學家邀請演講影片
- 6 部 keynotes 主題演講
- 162 部約 2~3 分鐘之短片介紹本屆錄取論文

經過現場近全程的錄影及後製，並動員所有論文作者製作介紹短片(本屆研究論文共 784 篇投稿，審核論文的專家約 675 位)，本屆 KDD 的成果幾近完整保存，提供給各界研究人員持續參考學習之用，該組織竭力實現學術開放及分享之理念由此體現，此會議之規劃與實踐成果至為豐碩，此乃人類智慧之高度呈現。

KDD KDD2016 video 影片 播放清單 頻道 討論 簡介



觀看次數：101 1 週前

Organizers:
Shipeng Yu, LinkedIn Corporation
Mohak Shah, Bosch Research and Technology Center North America
Balaji Krishnapuram, IBM

More on <http://www.kdd.org/kdd2016/>

KDD2016 Conference is published on <http://videlectures.net/>
[閱讀詳情](#)

Why KDD?

- Bridge Theory & Practice, in order to:
 - Maximize the global impact of data science over the next 10+ years
 - Help you be more successful in your career

已建立的播放清單

<p>Research Sessions 6 天前更新</p> <p>Robust Influence Maximization 20:59 Joint Community and Structural Hole Spann... 18:14 查看完整播放清單 (68 部影片)</p> <p>Hands-On Tutorials 6 天前更新</p> <p>Scalable R on Spark (Part 1) 1:17:32 Scalable R on Spark (Part 2) 1:28:46 查看完整播放清單 (20 部影片)</p> <p>KDD 2016 Opening & Award Ceremony</p> <p>KDD 2016 Award Ceremony 1:07:06 A Big Data Perspective (ACM SIGKDD 2016 ... 24:34 KDD 2016 Opening Ceremony 16:14 查看完整播放清單 (3 部影片)</p> <p>KDD 2016 Keynote Talks</p> <p>Graphons and Machine Learning: Modelin... 1:12:17 Learning to learn and compositionality wit... 1:23:45 The Evolving Meaning of Information Sec... 1:29:27 查看完整播放清單 (6 部影片)</p>	<p>Applied Data Science Sessions 6 天前更新</p> <p>From Online Behaviors to Offline Retailing 19:17 Online dual decomposition for performance... 18:54 查看完整播放清單 (30 部影片)</p> <p>KDD 2016 Tutorials</p> <p>Algorithmic Bias: From Discrimination Disc... 35:12 Algorithmic Bias: From Discrimination Dis... 1:14:46 Algorithmic Bias: From Discrimination Dis... 1:21:53 查看完整播放清單 (27 部影片)</p> <p>KDD 2016 Applied Data Science Invited Talks</p> <p>How Machine Learning has Finally Solved W... 24:40 Learning Sparse Models at Scale 37:05 The Wisdom of Crowds: Best Practices for ... 38:21 查看完整播放清單 (13 部影片)</p> <p>KDD 2016 Short Video Presentations</p> <p>KDD2016 paper 913 3:27 KDD2016 paper 839 2:22 KDD2016 paper 1036 2:10 查看完整播放清單 (162 部影片)</p>
---	---

KDD 於 Youtube 建立的 KDD2016 video 頻道相關內容

貳、會議過程：

以下按日分別紀錄現場聆聽、參觀之會議活動項目：

8 月 13 日：參加 ACM SIGKDD Tutorial Day

- 一、上午參加教育訓練：Enabling the Discovery of Reliable Information from Passively and Actively Crowdsourced Data (發掘被動及積極眾包數據可靠性之方法研究)
- 二、下午參加教育訓練：Healthcare Data Mining with Matrix Models (利用矩陣模型進行醫療資料探勘)

8 月 14 日：參加 ACM SIGKDD Workshop Day 及開幕儀式

- 一、上午參加 2016 Workshop on Large-Scale Deep Learning for Data Mining 大規模深度學習資料探勘
- 二、下午參加 KDD Workshop on Causal Discovery (以電腦技術進行因果關係研究)
- 三、晚上參加開幕儀式、大會介紹、頒獎及參觀 Poster Reception: Research (oral only) 研究論文海報展示

8 月 15 日：參加 ACM SIGKDD Conference

- 一、上午 8:00AM 聆聽 Keynote 演講：Jennifer Chayes, Graphons and Machine Learning: Modeling and Estimation of Sparse Massive Networks (Graphons 和機器學習：稀疏大規模網絡的建模與預測)
- 二、上午 9:40AM 安排參觀 Exhibit Hall 展覽廳
- 三、上午出席 Research Session R2 section：
Large Scale Data Mining: XGBoost: A Scalable Tree Boosting System;
Communication Efficient Distributed Kernel Principal Component Analysis
- 四、中午參加 Keynote 演講 - Joe Hellerstein, People, Computers, and The Hot Mess of Real Data (人，計算機和大量真實數據)
- 五、下午聆聽 Plenary Panel: Is Deep Learning the New 42?
- 六、晚間 7:00~10:00PM 安排參觀 Poster Reception: Applied Data Science, 應用論文海報展示。

8 月 16 日：參加 ACM SIGKDD Conference

- 一、上午 8:00AM 聆聽 Keynote 演講：Greg Papadopoulos, NEA, A VC View of Investing in ML (創投對投資機器學習的前瞻思考)。
- 二、上午聆聽 Applied Data Science Invited Talks 3, 應用數據科學邀請演講：
 1. Accelerating the Race to Autonomous Cars 加速進入無人駕駛汽車競賽,
Danny Shapiro: Senior Director of Automotive / NVIDIA

2. Profiling Users from Online Social Behaviors 分析用戶之在線社交行為, with applications in Tencent Social Ads, Ching Law : GM Social Ads / Tencent
3. Learning Sparse Models at Scale 學習大規模之稀疏模型, Ralf Herbrich : Director of Machine Learning / Amazon

三、下午參加 Research Session R10 場次：

1. Recommender Systems, When Social Influence Meets Item Inference 社群推薦與項目推薦研究, HuiJu Hung, Pennsylvania State University.
2. Goal Directed Inductive Matrix Completion 目標導向歸納式完全矩陣研究

四、下午安排參觀 Exhibit Hall 展覽廳

五、晚間 7：00~10：00PM 安排參觀 Poster Reception： Research (poster only) 研究論文海報展示。

8 月 17 日：參加 ACM SIGKDD Conference

- 一、上午 8：00AM 聆聽 Keynote 演講：Nando de Freitas, Professor / Oxford University / Google DeepMind, Learning to learn and compositionality with deep recurrent neural networks (專題演講：建構深遞歸神經網絡之學習與組合性)。
- 二、下午參加 Research Session R13 場次：
User Behavior Modeling 用戶行為建模

參、心得與建議

一、 本會議由ACM SIGKDD 組織主辦，該組織核心成員主要為資料科學領域之學術、業界之領導人士，會議於8月13日至8月17日假舊金山聯合廣場希爾頓飯店舉行，本屆參與人數高達2700多人，多為各國資料科學研究人員、實作家及相關業者，其中更有Microsoft、google、facebook、Linkedin、IBM、YAHOO、ebay、ADOBE、NEC、NVIDIA等國際知名軟、硬體及社群網站公司，以及中國大陸騰訊、滴滴等通訊及電子商務公司之重量級人物出席，並獲得許多廠商贊助，其目的在於分享最新之研究成果，行程成果如下：

1. 隨著大數據(BigData)時代的來臨，誰能在龐雜的訊息中掌握關鍵技術，誰就能取得解決問題的先機，ACM SIGKDD 為資料探勘知識發掘之旗艦級會議，其所發表的論文均為研究領域的翹楚，大數據資料探勘若類比於挖礦尋寶，KDD所提供各項資訊就是最好的藏寶圖，其中的主題演講(Keynotes)、論文(Accepted Papers)、實作教程(Hands on Tutorials)、工作坊(WorkShops)之內容極具參考價值。
2. 本會議規模之大為歷年罕見，除參與人數快速成長，相關論文品質、數量均為一時之選，論文依技術分類表列如下：

- (1) Clustering 集群
- (2) Frequent Pattern Mining 頻率模式挖掘
- (3) Outlier and Anomaly Detection 離群值和異常檢測
- (4) Dimensionality Reduction 降維
- (5) Recommender Systems 推薦系統
- (6) Graph Mining and Social Networks 圖挖掘和社交網絡
- (7) Deep Learning 深度學習
- (8) Big Data 大數據
- (9) Time Series and Stream Mining 時間序列和流挖掘
- (10) Mining Rich Data Types 挖掘豐富的數據類型
- (11) Privacy-Preserving Data Mining 隱私保護數據挖掘
- (12) Optimization Techniques 最佳化技術
- (13) Semi-Supervised Learning 半監督學習
- (14) Data Reliability and Truthfulness, 數據的可靠性和真實性
- (15) Large Scale Machine Learning Systems, 大規模機器學習系統
- (16) Classification 分類

3. KDD2016的論文分研究(Research Track)及應用(Applied Data Science Track)2個類別，研究類共784篇投稿，選出70篇上台報告及72篇海報展示，應用類共331篇投稿，選出40篇上台報告及26篇海報展示。在網頁Program項下的Accepted papers中用臺灣Taiwan作關鍵字可以找到2篇文章，一篇為臺大資工系(Wei-Lin Chiang、李穆竺及林智仁教授)研究團隊，另一篇為賓州州立大學(洪慧儒、李旺謙教授)、中研院(洪慧儒、De-Nian Yang、Liang-Hao Huang、陳銘憲教授)及加拿大Simon Fraser University的Jian Pei聯合發表之成果。
4. KDD2016除主會議外，同時舉行十幾個以研討會規格辦理之工作坊(workshops)，分別聚焦於不同應用領域，包括利用電腦圖論進行探勘與機器學習、依時間序列之數據挖掘與學習研究、以及將大數據技術應用於運動、食品能源和水、城市地區計算、交互式數據探索和分析、臨界定義、檢測、點播分析，因果關係之研究、資訊視覺化應用等主題，網頁連結<http://www.kdd.org/kdd2016/Workshops>。
5. 資料科學的領域又深又廣，除深奧的數學、統計、計算機軟、硬體應用外，許多領域都有研究20年以上的技術累積，並且隔行如隔山，目前有所突破的Machine Learning領域中，利用類神經網路的深度學習Deep Learning是一個快速增長的領域，為目前對數據分析最有前途的方法之一，其雖在最近短短的2、3年內大放異彩，其實已經是第3次捲土重來，研究工作的歷史相當長遠。深度學習已成功地應用於廣泛的領域中，如語音識別，電腦視覺，自然語言處理和分析大型企業的數據。眾多領域均迫切需要充分利用大數據的空前規模，擴充發展深度學習，但許多研究人員都表示要能成功的建模、讓系統跑出成效並非易事，許多知名的教授、廠商都看好此領域的後續發展，亟欲網羅相關人才。

二、心得與建議：

1. Big Data 大數據應用透過機器學習(Machine Learning)及深度學習(Deep Learning)技術於過去4年來的突飛猛進，主要基於下列3個因素：
 - (1) 首先GPU(graphics processing unit) 加速運算的出現，其以圖形處理單元(GPU) 搭配CPU，大幅提升電腦計算能力。

- (2) 網際網路的蓬勃發展累積了大量資訊，不論在社群網路或是電子商務應用上，均提供機器學習與深度學習可加運用之訓練資料及測試資料。
- (3) 各式演算法及程式開放源碼的建置(如Caffe、CNTK、TensorFlow、theano)，讓數學、統計學所發展的各项模型演算程式迅速累積，透過程式開放平台(如github平台)提供給大眾使用。

再加上雲端基礎建設供應商如AMZ、Google、Microsoft等，將機器學習所需技術元件拆解成各項獨立服務元件，讓使用者得以快速組合利用，並提供以GPU為基礎的運算能力，將此3因素中的2項整合起來，造成資料科學家如雨後春筍般由各領域中冒出，運用不同領域的資料進行機器學習試驗，產生多項成果，如AlphaGo破解圍棋，Computer Vision用於自動車駕駛及NLP自然語言處理產生的語音軟體及相關應用等。

2. 對資料探勘技術而言，資料就是礦脈，除網際網路上廣大免費的資料外，各國政府架構的開放資料、電信服務業者、教育機構、醫療機構、圖書館、出版商、汽車監理所、社群網站等等，凡是透過電子服務大量收集資訊的機構，乃是擁有不同資料原礦的所在，不論提供免費或收費的存取服務，都將直接或間接促進資料科技領域之演化，因其屬於上述3項基本因素之一，其價值已不可同日而語。也許這就是今年6月中Microsoft宣布欲以巨額之262億美金併購Linkedin(是一家商業客戶導向的社交網絡服務網站，成立於2002年12月並於2003年啟動。2011年1月，LinkedIn有超過9000萬的註冊用戶，到2012年1月，LinkedIn已經超過1.5億的註冊用戶。此網站的目的是讓註冊用戶維護他們在商業交往中認識並信任的聯繫人，這些人被稱為“人脈”。用戶可以邀請他認識的人成為人脈。)的因素之一。本署配合國家政策亦收集多項開放資料，並建構「環境即時通APP」及「iENV愛環境」網站，展示資料加值應用，期能拋磚引玉，擴大資料價值。
3. 透過創投基金(Venture Capital) 主題演講中窺得投資創投的利基及市場未來趨勢，透過NEA創投基金合夥人Greg Papadopoulos 以投資者的立場，談論創投基金對機器學習領域的展望，相關重點如下：

- (1) 創投營收目標按比例只要在100家公司中，有1家能爆紅(全壘打型新創

公司平均收益500倍以上), 5家能上市, 其餘的就算不賺錢, 不獲利, 也算收益可觀, 此一觀念像是將投資組合比喻為買保險, 不成功的新創公司由全壘打公司來彌補。

(2) 當創投提出資料科學究竟是產品還是技術的問題時, 就隱約感覺創投風向的轉變, 果不其然, 其對未來前景預測是持平延伸後走勢向下的, 但其原因並不是針對Data Science的研究, 而是預測資料科學不久將變成大眾化的工具, 甚至是免費的工具。創投尋找的是更貼近市場需求, 因此新創公司須提升vertical value、能發展出推廣全域的執行力, 甚或是Industrial IoT與機器人相關產業。

(3) 學術界研究多重視理論, 新創公司需能將技術融入業界實務需求中, 以下列出現場資深創投對新創公司提出的3大問題, 可以提供給新創業者做為進一步的省思:

- i. What's the differentiation? (產品的差異性為何)
- ii. What's the opportunity? (產品賣出的機會為何)
- iii. Do you like the people you work with?(你是否喜歡和創業的夥伴們一同工作打拼?人的因素很重要。)

4. 機器學習(Machine Learning)及深度學習(Deep Learning)突破性的發展, 對全球相關領域人才產生極大衝擊, 今年美國大學畢業生, 以資訊系中資訊安全(information Security)、大數據(Big Data)、雲端運算(Cloud computer)3個專業領域的畢業生獲得最多工作機會, 薪水也水漲船高。本人在會議上隨機遇上某位於美國任教的華裔大學教授, 曾向我打聽1位臺灣的大學畢業生, 因其在相關領域曾有傑出表現, 其名教授欲提供獎學金給該名畢業生, 請其考慮進入該校就讀。

5. 資料科學、機器學習、深度學習等研究之突飛猛進與未來第四次工業革命密不可分。第四次工業革命核心為智能化(AI)與信息化(Information exchange), 包括「智慧整合感控系統」CPS, Cyber-Physical Systems, 物聯網(IOT)和雲端運算(Cloud Computing)及互聯網(Internet)等等, 最近「遠見雜誌」報導8位國際上不同領域之大師預言之未來十大趨勢

(http://www.gvm.com.tw/Boardcontent_31621.html)，內容與本次會議中眾多學者專家所見不謀而合，其中機器人、物聯網、無人駕駛車、共享經濟(如UBER、Airbnb)、生物結合資訊發展等，預計對人類未來的生活將產生重大的衝擊，台灣如能趕得上這一波變革，捉住機會就有發展，跟得上是機會，跟不上就是挑戰，錯失了機會，國家將面臨衰退危機。

ACM SIGKDD 2016 「第 22 屆知識發現和資料探勘技術國際研習會」活動照片



KDD2016 大會開幕，參加人數超過 2700 人。



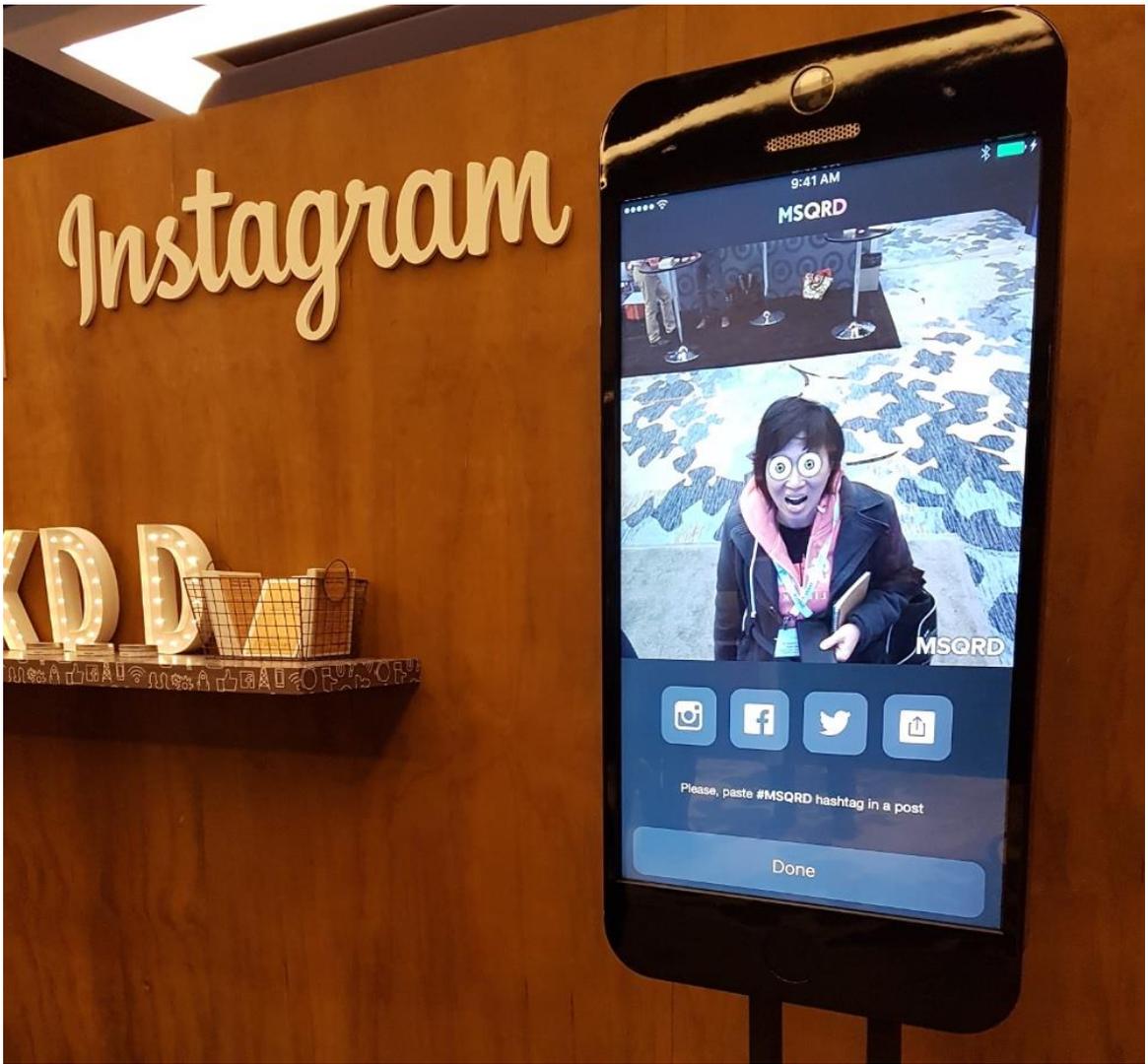
出國代表魏文娟分析師(中)與 Microsoft 資深科學家們合影



出國代表魏文娟分析師(右)與我國 Accepted poster 台大資工所教授林智仁合影



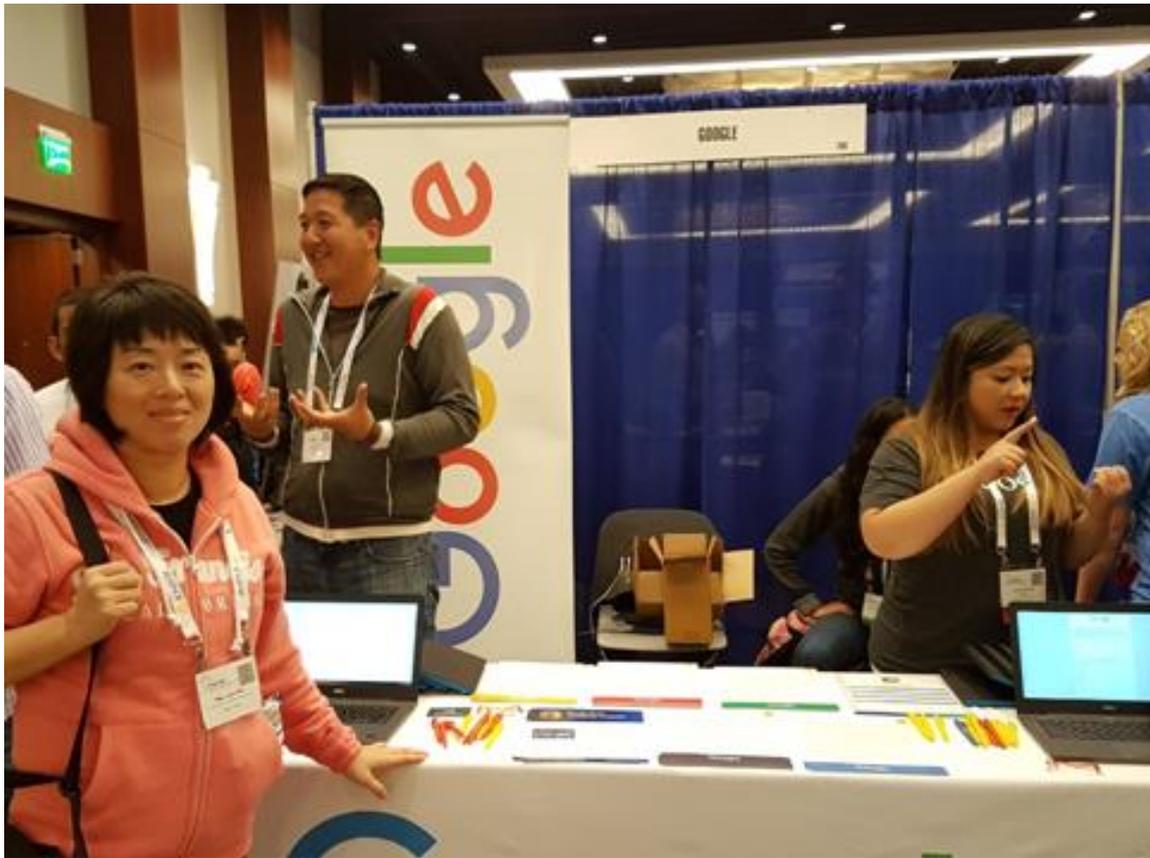
KDD2016 廠商展覽會場，facebook 及 Instagram (左)，DiDi(右)。



KDD2016 廠商展覽會場 Instagram 現場影像 APP 動態展示



KDD2016 廠商展覽會場，Microsoft (左)，ebay (右)。



KDD2016 廠商展覽會場，google 攤位。



KDD2016 廠商展覽會場，NEC 攤位。



Plenary Panel section 現場爆滿，討論主題： Is Deep Learning the New 42?



Ralf Herbrich, Director of Machine Learning, Amazon, surrounded by audiences after his fascinating talks ‘Learning Sparse Models at Scale’.



魏文娟(左)、'Parallel Dual Coordinate Descent Method for Large-scale Linear Classification in Multi-core Environm'論文發表者 Mu-Chu Lee 李穆竺(中)及 Microsoft 工程師合影。



本屆 KDD Sponsorship Chair Yu Zheng(左一)，The Pennsylvania State University associated professor Wang-Chien Lee 李旺謙(左二)，出國代表魏文娟分析師(右二)與及 When Social Influence Meets Item Inference 論文發表者 Hui-Ju Hung (右一)合影。