**上海交通大学-悉尼科技大学2014年度智能系统研讨会**

**时间：**2014年4月22日9:00-17:00

**地点：**电信楼群5号楼406大会议室

**主办单位：**电信学院图像通信与网络工程研究所

**会议日程：**

|  |
| --- |
| **上午，主持：杨小康（上海交通大学）** |
| **09:30-09:40，张文军（上海交通大学）**  ***致开幕辞*** |
| **09:40-10:10，Chengqi Zhang（悉尼科技大学）**  ***Overview of research in the Centre for Quantum Computation & Intelligent Systems*** |
| **10:10-10:40，Ivor W Tsang（悉尼科技大学）**  ***Feature Selection for Big Data with Trillion Dimensions*** |
| **10:40-11:00**  ***茶歇*** |
| **11:00-11:30，Ling Chen（悉尼科技大学）**  ***Frequent Pattern Mining Over Uncertain Data*** |
| **11:30-12:00，张娅（上海交通大学）**  ***用户购买行为的归因分析*** |
| **下午，主持：Chengqi Zhang（悉尼科技大学）** |
| **14:00-14:30，Ying Zhang（悉尼科技大学）**  ***Efficient Query Processing On massive spatial Data*** |
| **14:30-15:00，Lu Qin（悉尼科技大学）,**  ***Graph Processing in the Era of Big Data*** |
| **15:00-15:30，周憬宇（上海交通大学）**  ***Extracting URLs from JavaScript via Program Analysis*** |
| **15:30-15:50**  ***茶歇*** |
| **15:50-16:20，Dacheng Tao（悉尼科技大学）**  ***Learning from Multi-view Data: Practice and Theory*** |
| **16:20-16:50，杨小康（上海交通大学）**  ***海量视频中的人群行为分析*** |
| **16:50-17:00，Chengqi Zhang（悉尼科技大学）**  ***致闭幕辞*** |

**Speaker: Prof Chengqi Zhang**

**Title*: Overview of research in the Centre for Quantum Computation & Intelligent Systems***

**Abstract:**

The Centre for Quantum Computation &Intelligent Systems (QCIS) was established in April 2008 as a priority investment Centre of the University of Technology, Sydney. The Centre's research focus is to develop theoretical foundations, innovative technology and practical systems that will result in next-generation enterprise intelligent information systems. Over the last six years, QCIS staff have published around 400 high quality papers in prestigious journals and conference proceedings, achieved 37 Australian Research Council grants (around AU$12.5 million), and helped to raise the UTS-IT world-wide ranking to 100-150. This talk will summarise the achievements of QCIS over last six years, and discuss the reasons behind these achievements.

**Short bio:**

Chengqi Zhang has been a **Professor** of Information Technology at the University of Technology, Sydney (UTS) since December 2001. He has been the **Director** of the UTS Priority Investment Research Centre for Quantum Computation & Intelligent Systems (QCIS) since April 2008. He has been Alternative Dean of UTS Graduate Research School since April 2013. He has been **Chairman** of the Australian Computer Society’s National Committee for Artificial Intelligence since November 2005.

Prof Zhang obtained his Bachelor degree from Fudan University in 1982, his Master degree from Jilin University in 1985, his PhD degree from The University of Queensland in 1991, and he followed these with a Doctor of Science (DSc – Higher Doctorate) from Deakin University in 2002 – all in Computer Science.

Prof Zhang’s research interests mainly focus on Data Mining and its applications. He has published more than 200 research papers, including a number of papers in first-class international journals such as *Artificial Intelligence, IEEE* and *ACM Transactions*. He has published seven monographs and edited 16 books. He has delivered 14 keynote/invited speeches at international conferences. He has attracted 11Australian Research Council grants. Due to his outstanding research achievements, he was awarded the 2011 NSW Science and Engineering Award in the Engineering and ICT category.

Since his appointment as the Director of QCIS six years ago, Prof Zhang has led the Centre's researchers to publish around 400 high quality papers, double the national grant funding, and improve the grant ranking of UTS-IT to 100-150 world-wide, based on the Shanghai Jiao Tong University ranking. His Centre has achieved seven ARC Future Fellows in the last five ARC rounds, which is 18% of the national share from 38 Universities. Due to his leadership achievements, Prof Zhang was awarded the 2011 UTS Vice-Chancellor's research excellence awardsin the Leadership category.

Prof Zhang is a Fellowof the Australian Computer Society (ACS) and a Senior Member of the IEEE Computer Society (IEEE). He has been serving as an Associate Editor for three international journals, including IEEE *Transactions on Knowledge and Data Engineering* from 2005 to 2008; and he served as General Chair, PC Chair, or Organising Chair for five international Conferences, including ICDM 2010 and WI/IAT 2008. He is also General Co-Chair of KDD 2015 in Sydney and the Local Arrangements Chair of IJCAI-2017 in Melbourne (International Joint Conference on Artificial Intelligence).

**Speaker:Prof Dacheng Tao**

Title: *Learning from Multi-view Data: Practice and Theory*

**Abstract:**

In recent years, many methods have been proposed to learn from multi-view data by considering the diversity of different views. These views may be obtained from multiple sources or from different feature subsets. For example, a person can be identified by their face, fingerprint, signature or iris using information obtained from multiple sources, while color or texture features can be used to represent an image. In this talk, we will discuss the similarities and differences between the variety of multi-view learning approaches, highlight their limitations, and then introduce a new multi-view learning algorithm, LMIB (Large-margin Multi-view Information Bottleneck). LMIB formulates the multi-view learning problem as one of encoding a communication system with multiple senders, each of which represents one view of the data. Some notable theoretical results on the consensus and complementary principles in multi-view learning will be presented, and in doing so we will shed new light on the design of new multi-view learning algorithms.

**Short bio:**

Dacheng Tao is Professor of Computer Science with the Centre for Quantum Computation & Intelligent Systems and the Faculty of Engineering & Information Technology in the University of Technology Sydney (UTS). He mainly applies statistics and mathematics for data analysis problems in machine learning, data mining & engineering, computer vision, image processing, multimedia, video surveillance and neuroscience. His research results have expounded in one monograph and 300+ publications at prestigious journals and prominent conferences, such as IEEE T-PAMI, T-IP, T-SP, T-NNLS, JMLR, IJCV, NIPS, ICML, CVPR, ICCV, ECCV, AISTATS, ICDM; ACM KDD and Multimedia, with several best paper awards, such as the best theory/algorithm paper runner up award in IEEE ICDM’07 and the best student paper award in IEEE ICDM’13.He is/was a guest editor of 10+ special issues, an editor of 10 journals, including IEEE Trans. on Cybernetics (T-CYB), IEEE Trans. on Circuits and Systems for Video Technology (T-CSVT), IEEE Trans. on Knowledge and Data Engineering (T-KDE), Pattern Recognition (Elsevier), Information Sciences (Elsevier), Signal Processing (Elsevier), and Computational Statistics & Data Analysis (Elsevier).

**Speaker: A/Prof Ivor W Tsang**

**Title: *Feature Selection for Big Data with Trillion Dimensions***

**Abstract:**

The world continues to generate quintillion bytes of data daily, leading to pressing needs for new endeavours to deal with the grand challenges brought about by Big Data. Today, there is consensus between themachine learning and data mining communities that data volume presents an immediate challenge pertaining to scalability. However, when addressing volume in Big Data analytics, researchers have taken a one-sided viewof volume, which is the "big instance size" factor of the data. The flip side of volume, which is the dimensionality factor of Big Data, has received much less attention. In this talk, I will attempt to fill this gap and place special focus on therelatively under-explored topic of ultra-high dimensionality. Specifically, I will first reformulate the resultant non-convex problem as a convex semi-infinite programming (SIP) problem, and then present an efficient feature generation paradigm to solve it. The proposed feature generation willachieve lower feature selection bias compared to L1-regularized methods. In addition, trillions of correlations among millions of features can be efficiently handled by the proposed feature generation framework. Comprehensive experiments on a wide range of synthetic and real-world datasets with tens of millions of data points and O(10^14) dimensions demonstrate that the proposed method achieves superb performance when compared with state-of-the-art feature selection methods, in terms of generalisation performance and training efficiency.

**Short bio:**

Ivor W Tsang is an Australian Future Fellow and Associate Professor with the Centre for Quantum Computation & Intelligent Systems (QCIS), at the University of Technology, Sydney (UTS). Before joining UTS, he was the Deputy Director of the Centre for Computational Intelligence, Nanyang Technological University, Singapore. He received his PhD degree in computer science from the Hong Kong University of Science and Technology in 2007. His research focuses on kernel methods, transfer learning, feature selection, big data analytics for data with millions of dimensions, and their applications to computer vision and pattern recognition. He has more than 100 research papers published in refereed international journals and conference proceedings, including 4 JMLR, 8 T-PAMI, 18 T-NN, 12 ICML, NIPS, UAI, AISTATS, SIGKDD, IJCAI, AAAI, ICCV,CVPR, ECCV, ACL, etc.

Dr Tsang received hisprestigious Australian Research Council Future Fellowship in 2013, and had previous been awarded the 2008 Natural Science Award (Class II) by the Ministry of Education, China, and the *IEEE Transactions on Neural Networks* Outstanding 2004 Paper Award in 2006. His research also earned him the Best Student Paper Award at CVPR'10, the Best Paper Award at ICTAI'11, and the Best Poster Honorable Mention at ACML'12. He was also conferred with the Microsoft Fellowship in 2005.

**Speaker: Dr Ling Chen**

**Title: *Frequent Pattern Mining Over Uncertain Data***

**Abstract:**

Data uncertainty is inherent in various applications. It has also posed many unique challenges to nearly all types of data mining tasks, creating a need for uncertain data mining. This talkintroduces our recent research on mining frequent patterns over uncertain data. In particular, it focuses on the following two topics: 1) frequent serial episode mining over uncertain data, and 2) probabilistic frequent pattern summarisation.

For frequent serial episode mining, research is undertaken to devise data mining algorithms to discover independent frequent serial episodes and dependent frequent serial episodes. We have developed exact solutions and approximate solutions, as well as optimisation solutions for mining the two types of frequent patterns. Our work was published at EDBT2013 and ICDM2013.

For probabilistic frequent pattern summarisation, we have formally defined the problem of probabilistic representative frequent pattern mining, which aims to find the minimal set of patterns with sufficiently high probability to represent all other patterns. The developed exact solution and approximate solution were published at SDM2013 and KDD2013, respectively.

**Short bio:**

Dr Ling Chen is a Lecturer with the Faculty of Engineering and Information Technology (FEIT), University of Technology, Sydney. She received her PhD in 2008 in Computer Engineering, from Nanyang Technological University, Singapore. She was a Postdoctoral Research Fellow with the L3S Research Centre, Leibiniz University Hannover, Germany from 2007 to 2009. Ling's main research interests include data mining, machine learning and social media. Ling is currently a core member of the Centre for Quantum Computation & Intelligent Systems (QCIS) at UTS.

**Speaker:DrYing Zhang**

**Title*: Efficient Query Processing On massive spatial Data***

**Abstract:**

With the rapid development of positioning technologies and the boosting deployment of inexpensive location-aware sensors, a large volume of spatial data has been rapidly increasing. The value lockedup in the overwhelming amounts of spatial data presents extraordinary and unprecedented opportunities to discover and share new knowledge in many critical applications, such as location based service (LBS), national security and defence, marketing, traffic management, health care and environmental monitoring. This talk will introduce important research work on massive spatial data processing with special focus on uncertain spatial data processing and spatial keyword search, including some of our recent contributions in these fields. Iwill also discuss our recent research work on streaming enriched geo-spatial data.

**Short bio:**

Dr Ying Zhang is a senior lecturer and ARC DECRA Research Fellow at the University of Technology, Sydney (UTS). He received his BSc and MSc degrees in Computer Science from Peking University, and his PhD in Computer Science from the University of New South Wales. His research interests include query processing on data streams, uncertain data, spatial data and graphs. He has published 20+ papers in the most prestigious international journals and conference proceedings, such as TODS, VLDBJ, TKDE, SIGMOD, SIGIR, VLDB and ICDE. He also serves as a PC member formany international database conferences such as PVLDB and ICDE. He was awarded an Australian Research Council Australian Postdoctoral Fellowship (ARC APD, 2014-2016) and an Australian Research Council Discovery Early Career Researcher Award (ARC DECRA, 2011-2013).

**Speaker: Dr Lu Qin**

**Title: *Graph Processing in the Era of Big Data***

**Abstract:**

With the emergence and rapid proliferation of applications that deal with big graphs, such as web graphs (Google, Yahoo), social networks (Facebook, Twitter), e-commerce networks (Amazon, Ebay), and road networks, graph processing has become increasingly prevalent and important in recent years. However, in the era of big data, the explosion and profusion of available graph data in a wide range of application domains raises new challenges and provides new opportunities in graph processing.

In this talk, I will present my group’s recent progress in graph processing in terms of new graph query semantics and new computing paradigms. For new graph query semantics, I will introduce our work on structural keyword search, shortest path computation, graph matching, and community detection. For new computing paradigms, I will present our research on multi-core graph processing techniques, external graph processing techniques, and distributed graph processing techniques (MapReduce, BSP) in the cloud. Finally, I will discuss potential future research directions for graph processing.

**Short bio:**

In July of 2006, Dr Lu Qin received his bachelor degree from the Department of Computer Science and Technology in Renmin University, China (RUC). He had been the team leader of RUC for the ACM International Collegiate Programming Contest (ICPC). His team won three gold medals and two silver medals in the ACM/ICPC regional contests and advanced to the world finals twice. In August 2010, Dr Lu Qin received his PhD degree from the Department of Systems Engineering and Engineering Management (SEEM) in the Chinese University of Hong Kong (CUHK). He was a postdoctoral research fellow in CUHK from August 2010 to August 2013. Following this, Dr Qin joined the University of Technology, Sydney (UTS) as a core member in the Centre for Quantum Computation &Intelligent Systems (QCIS). He is also an adjunct lecturer at the University of New South Wales (UNSW), Australia.

Dr Qin’s research interests include algorithm design and analysis for new big data problems, big graph processing in the cloud, and big graph searching and mining. He has been very productive in databases. In the last five years, Dr Qin has published more than 30 top conference/journal papers, including four SIGMOD papers, five PVLDB papers, and five ICDE papers in the top-three database conferences, and five VLDB journal papers, one Algorithmica paper, and one TKDE paper in top-ranked database and algorithm journals. His book entitled “Keyword Search in Databases” is the first monograph on this field of research. Dr Qin has served as a program committee member for many top database and data mining conferences. He has received several research grants from the Australia government, the Hong Kong government, and from UTS.

Dr Qin has extensive experience in student supervision. Over the last four years, he has supervised/co-supervised five PhD students,two of whom have successfully received their PhD degrees, and three of whomhave continuously published their work in top-ranked conferences and journals under the supervision of Dr Qin.

**Speaker:** 周憬宇

**Title: *Extracting URLs from JavaScript via Program Analysis***

**Abstract:**

With the extensive use of client-side JavaScript in web applications, web contents are becoming more dynamic than ever before. This poses significant challenges for search engines, because more web URLs are now hidden inside JavaScript code and most web crawlers are script-agnostic, significantly reducing the coverage of search engines.We present a hybrid approach that combines static analysis with dynamic execution, overcoming the weakness of a purely static or dynamic approach that either lacks accuracy or suffers from huge execution cost. We apply program analysis techniquessuch as statement coverage and range analysis to improve the performance of URL mining. Our approach can handle modern web applications with DOM interactions, dynamic code generation and JavaScript libraries.Experiments on large sets of popular web pages show that our approach is effective and efficient in practice.

**Bio:**

Dr.Jingyu Zhou received the B.S. degree in Computer Science from [Zhejiang University](http://www.zju.edu.cn/), China, in 1999. He received the M.S. and Ph.D. degrees in Computer Science from [University of California at Santa Barbara](http://www.ucsb.edu/) in 2003 and 2006. He joined the [Software Institute](http://www.se.sjtu.edu.cn/) at Shanghai Jiao Tong University in 2006 and is currently affiliated with [Department of Computer Science and Engineering](http://www.cs.sjtu.edu.cn/index.action).

He is generally interested in systems, information retrieval, and security. He has published more than 30 papers at various conferences and journals, including WWW, INFOCOM, ICS, TPDS, DSN, FSE, CIKM, IS, and IPDPS. He has served as PC members for 18 international conferences, including RACS, FCST, GPC, and EUC. His current work focuses on distributed systems and information retrieval. His past work includes network and application security, pervasive computing, cluster-based storage systems and middleware systems.

**Speaker:** 杨小康

**Title: *Crowd Behaviour Analysis in Video***

**Abstract:**

In recent years, behavior analysis in crowded environments has drawn increasing attentions in computer vision and pattern recognition community because of its wide industrial applications in video surveillance and the important scientific values in physics and biology. This talk will introduce several models to solve this challenge in the context of multi-granularity of crowd behaviours. According to the availability of body features, individual features, group layer, and group interaction, crowd behaviours can be in the range of individual action, fine granularity, medium granularity and coarse granularity. For action recognition, a “Acton” representation scheme is given to capture mid-level action “part” and relevant object so as tobridge gap between low level features and high level action semantics. For fine granularity, a hierarchical random conditional field model is built to parse the relationship between individuals and groups. For medium granularity, a spatiotemporal viscous fluid field model is used to exploit the nature of the behavior from both the appearance and stimulus aspects of crowd. For coarse granularity(very crowded scene), a large-scale dataset of manually annotated 50,000 crowd segments for crowd understanding is built and an algorithmic benchmark is given.

**Bio:**

Xiaokang YANG received the B. S. degree from Xiamen University, Xiamen, China, in 1994, the M. S. degree from Chinese Academy ofSciences, Shanghai, China, in 1997, and the Ph.D. degree from ShanghaiJiao Tong University, Shanghai, China, in 2000.

He is currently a professor and Vice Dean, School of Electronic Information and Electrical Engineering, and the deputy director of the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai, China. From August 2007 to July 2008, he visitedthe Institute for Computer Science, University of Freiburg, Germany, as an Alexander von Humboldt Research Fellow. From September 2000 toMarch 2002, he worked as a Research Fellow in Centre for SignalProcessing, Nanyang Technological University, Singapore. From April 2002to October 2004, he was a Research Scientist in the Institute forInfocomm Research (I2R), Singapore. He has published over 150 refereed papers, and has filed 30 patents. His current research interests include visual signal processing and communication, media analysis and retrieval, and pattern recognition.

He received National Science Fund for Distinguished Young Scholars in 2010, Professorship Award of Shanghai Special Appointment (Eastern Scholar) in 2008, the Microsoft Young Professorship Award in 2006, the Best Young Investigator Paper Award at IS&T/SPIE International Conference on Video Communication and Image Processing (VCIP2003) and awards from A-STAR and Tan KahKee foundations. He is currently a member of Editorial Board ofIEEE Signal Processing Letters, Serias Editor of Springer CCIS, a member of APSIPA, a senior member of IEEE, a member of Visual Signal Processing and Communications (VSPC) Technical Committee of the IEEE Circuits and Systems Society. He was the special session chair of Perceptual Visual Processing of IEEE ICME2006. He was the technical program co-chair of IEEE SiPS2007 and the technical program co-chair of 3DTV workshop in junction with 2010 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting.

**Speaker: Ya Zhang**

**Title**:Multi-Touch Attribution Model in Online Advertising

**Abstract**:

Nowadays, an advertisement campaign generally involves multiple media channels, suchas search, display, social, mobile and video. As a result, one of the most important problems in digitaladvertising is multi-touch attribution, which is to assign credit to different advertisements for driving the userto the desirable actions such as clicking an advertisement link or making a purchase.For a long time, the most commonly used attribution strategy in the industry is to simply assign fullcredit to a single advertising channel, typically the last adsbefore a desired action (”last-touch” attribution).Ratherthan giving all the credit to the last ad a user sees, multi-touch attribution allows more than one ads to get the creditbased on their corresponding contributions. Using a using data-driven approach, we here propose a novel multi-touch attribution model called Additive Impact Model (AIM) based on survival analysis. We experiment with a real-world advertisement data set and show that the proposed AIM model is quite promising.

**Bio**:

Ya Zhang received her B.S. degree from Tsinghua University, China in 2000, and Ph.D. degree in Information Sciences and Technology from the Pennsylvania State University in 2005.

Ya Zhang is a Professor in the Department of Electronic Engineering at the Shanghai Jiao Tong University since 2010. Her research interest is mainly on machine learning applications, information retrieval, web science, and social network mining.Before joining Shanghai Jiao Tong University, she was a research manager at Yahoo Labs, where she leaded a R&D team of researchers with strong background in data mining and machine learning to improve the web search quality of Yahoo international markets. Prior to joining Yahoo, Dr. Zhang was a tenure-track assistant professor at University of Kansas. Dr. Zhang published over 40 refereed publications in prestigious international conferences and journals including SIGIR, ICDM, WWW, NIPS, ICDM and KDD. She was nominated for the best paper award of the American Medical Informatics Association Annual Symposium in 2002 and was a recipient of the best paper award of the IEEE CVPR 2005 Workshop on Computer Vision Methods for Bioinformatics in 2005. She currently holds 3 US patents and has a dozen ofChinese and USpatents pending in the areas of web search and data mining.