

PROCEEDINGS

The 17th HKBU-CSD Postgraduate Research
Symposium

PG Day 2014



Department of Computer Science
Hong Kong Baptist University
June 3, 2014

The 17th HKBU-CSD Postgraduate Day Program

June 3 rd 2014, Tuesday			
Time	Sessions		
09:00-09:20	On-site Registration (T909)		
09:20-09:30	Welcome: Prof. P. C Yuen (<i>Head of Department of Computer Science, HKBU</i>) (T909)		
09:30-11:00	Session I (Chair: Mai Guangcan, T909)	Session II (Chair: Yi Peipei, T714)	Session III (Chair: Lou Jian, T716)
	<ul style="list-style-type: none"> ➤ Mai Guangcan (PC YMC LU TAM) ➤ Lan Xiangyuan (PC YMC CHU TAM) ➤ Li Jiawei (PC YMC TAM) 	<ul style="list-style-type: none"> ➤ Chen Qian (XU BC Haibo YW CHU) ➤ Chen Jingjing (XU BC Haibo RuiChen) ➤ Chen Lei (XU BC LG Haibo) 	<ul style="list-style-type: none"> ➤ This session starts at 10:00 ➤ Liang Fengfeng (LC WC JM XU) ➤ Wang Feng (LC Clement CHU BC)
	Tea Break		
11:00-11:15	Session IV (Chair: Mai Guangcan, T909)	Session V (Chair: Lou Jian, T714)	Session VI (Chair: Fan Zhe, T716)
	<ul style="list-style-type: none"> ➤ Mei Xinxin (CHU TAM LC HH) ➤ Liu Chengjian (CHU YW HL JNG) ➤ Yu Lu (YW HL CHU JNG XU) 	<ul style="list-style-type: none"> ➤ Liu Fei (BC CC BX Clement) ➤ Chan Sheung wai (Clement JM XU) ➤ Li Chen (WC BC JM LC) 	<ul style="list-style-type: none"> ➤ Bao Qing (WC JM YMC) ➤ Peng Qinmu (YMC LC CHU BC)
	Noon Break		
12:45-14:30			
14:30-16:00	Session VII (Chair: Yi Peipei, T909)		
	<ul style="list-style-type: none"> ➤ Fan Zhe (BC XU LC CHU) ➤ Yi Peipei (BC XU YMC CHU) ➤ Gu Fangqing (YMC LC CHU BC) 		
	Best Paper and Best Presentation Awards Announcement (T909)		
16:30-17:00	Closing		

Table of abstracts

Section I:	1
Discriminative and Minimally-correlated Biometric Representation via Feature-Level Binary Fusion for Multi-biometric Systems	1
Mai Guangcan	1
Multi-Cue Visual Tracking Using Robust Feature-Level Fusion Based on Joint Sparse Representation	1
Lan Xiangyuan	1
Domain Transfer Region Label Propagation Re-Identification	1
Li Jiawei	1
Section II:	2
Query Authentication on MultiSourced and Crowdsourced Data	2
Chen Qian	2
Ebanshu: An Interactivity-aware Blended Virtual Learning Environment	2
Chen Jingjing	2
Answering Why-not Questions on Spatial-Keyword Top-k Queries	3
Chen Lei	3
Section III:	4
Forecasting Disease Incidences Using State Space Model	4
Liang Fengfeng	4
Collective Review Mining for Aspect-Level Rating and Weight Estimation	4
Wang Feng	4
Section IV:	5
Benchmarking the Memory Hierarchy of Modern GPUs	5
Mei Xinxin	5
Reliable In-Memory Key-Value Store: Memcached with RAID	5
Liu Chengjian	5
Channel-Hopping Based on Available Channel Set for Rendezvous of Cognitive Radios	5
Yu Lu	5
Section V:	7
Is This Website Appealing to You? Explaining Electronic Commerce Consumers' Purchase Intention from an Appealing Perspective	7
Liu Fei	7
Knowledge Extraction and Mining in Biomedical Research Using Biomedical Rule Network	7
Chan Sheung Wai	7
Understanding Human Mobility using the Probabilistic Modelling Approach	8
Li Chen	8
Section VI:	9
Incorporating Structural Diversity of Neighbors in a Diffusion Model for Social Networks	9
Bao Qing	9
A new model for eye gaze tracking under natural light	9
Peng Qinmu	9
Section VII:	10
Structure-Preserving Subgraph Query Services	10
Fan Zhe	10
Privacy-Preserving Reachability Query Services for Sparse Graphs	10
Yi Peipei	10
Online Objective Reduction for Many-Objective Optimization Problems	10
Gu Fangqing	10

Section I:

Discriminative and Minimally-correlated Biometric Representation via Feature-Level Binary Fusion for Multi-biometric Systems

Mai Guangcan

Abstracts: Template security is a critical issue in multibiometric systems as it stores and processes information of multiple biometric traits per user. Feature-level fusion employed in multibiometric system better than other level fusion from security aspect. However, state of art binary feature level fusion is limited to simple concatenation, bitwise-operation-based fusion and bit selection, the discrimination power and entropy of the fused binary representation may not be optimal. A novel feature level binary fusion method consists of two stages, i.e., clustering of bit elements and within-cluster fusion is proposed in this paper. Binary feature fused by proposed method archive minimally correlation between different bits also with high discrimination power. Experimental result shows that proposed fusion method can not only give fused feature with high security, but also high matching performance.

Multi-Cue Visual Tracking Using Robust Feature-Level Fusion Based on Joint Sparse Representation

Lan Xiangyuan

Abstracts: The use of multiple features for tracking has been proved as an effective approach because limitation of each feature could be compensated. Since different types of variations such as illumination, occlusion and pose may happen in a video sequence, especially long sequence videos, how to dynamically select the appropriate features is one of the key problems in this approach. To address this issue in multicue visual tracking, this paper proposes a new joint sparse representation model for robust feature-level fusion. The proposed method dynamically removes unreliable features to be fused for tracking by using the advantages of sparse representation. As a result, robust tracking performance is obtained. Experimental results on publicly available videos show that the proposed method outperforms both existing sparse representation based and fusion-based trackers.

Domain Transfer Region Label Propagation Re-Identification

Li Jiawei

Abstracts: This paper addresses in domain transfer person re-identification problem with unlabeled data. For reliable label information in target domain, region label propagation is proposed. Graph label propagation algorithm is employed to learn the positive region with positive mean, pre-learnt classifier, target negative and unlabeled data. After that, a Rank support vector machine is introduced to learn classifier with positive region and negative samples.

Section II:

Query Authentication on MultiSourced and Crowdsourced Data

Chen Qian

Abstracts: Crowdsourcing has emerged as a promising paradigm for big data collection and analytics. However, as the data sources are distributed and complex, there is an increasing necessity to assure the integrity of the server who collects the data and provides query services to the clients. In this paper, we propose a novel authentication code called homomorphic secret sharing seal that can aggregate the input from individual owners faithfully even by an untrusted server for future query authentication. Based on this, we design two authenticated data structures and authentication schemes for spatial queries on distributed multi-dimensional data. We further study the freshness problem in query authentication and propose several advanced update strategies for the server. Analytical models and empirical results show our seal design and authentication schemes are efficient and robust under various system settings.

Ebanshu: An Interactivity-aware Blended Virtual Learning Environment

Chen Jingjing

Abstracts: Virtual learning environment (VLE), as a type of e-learning platform, has been widely used to serve the teaching and learning for education in many countries. However, most of the existing systems fail to seamlessly support and monitor the real-time interactivity and collaboration among the learners and instructors in the virtual learning environment. Moreover, instructors are unable to know the real-time learning statuses of learners in distance, which is critical to effective teaching and learning achievement. This paper presents a novel interactivity-aware blended VLE system for synchronous and asynchronous teaching and learning, called ebanshu [9]. It provides many real-time tools to support the interactivity and collaboration among the instructors and learners on both desktop PCs and mobile devices. By using the real-time visualized popularity dashboard, instructors can monitor real-time learning statuses of learners. Furthermore, all the teaching and learning activities in the virtual classroom environment will be automatically recorded as online videos and lectures for online open access. The ebanshu system had been deployed in more than 10 universities and institutions in China (including Hong Kong Baptist University, Peking University, Zhejiang University, and Jilin University), currently hosting more than 30 online courses and offering teaching/learning service for more than 10,000 users. The extensive impact analysis using the course MATH7090 in Hong Kong Baptist University as case study shows that ebanshu can effectively improve the efficiency of teaching and learning activities.

Answering Why-not Questions on Spatial-Keyword Top-k Queries

Chen Lei

Abstracts: The quality and the usability of database systems now have drawn more attention these years. The "Why-not" question, which is used to explain the reason why some objects are missed in the query result set, has been a very hot topic. On the other hand, spatial-keyword related problems have been studied for years and have many applications already, in which the spatial-keyword top-k query is a very important one. Thus, taking into consideration of both the two aspects above, we are motivated to study the problem answering why-not questions on spatial-keyword top-k queries. In this paper, we formulate this query, and propose a basic method which reduces the problem into a two-dimensional geometrical problem. To optimize the performance, we further propose an index-based ranking estimation algorithm, which prunes candidate results at early time. We also adapt the proposed algorithms to process the queries on multiple missing objects. Extensive experimental evaluation demonstrates the efficiency of our proposed query techniques in terms of both the computational cost and I/O cost.

Section III:

Forecasting Disease Incidences Using State Space Model

Liang Fengfeng

Abstracts: Disease surveillance systems are important towards effective disease control and intervention. These systems usually monitor disease incidences as time series data, and make predictions using statistical methods (e.g., autoregressive integrated moving average). However, most statistical methods are limited in revealing relationships between the dynamics of infection incidences and the changes of underlying causative factors. In this paper, we aim to (i) present a state space model to mathematically model the dynamics of disease transmission, and (ii) use an extended Kalman filter algorithm to fit the model to surveillance data. By doing so, we can not only predict the potential disease incidences, but also understand the transmission dynamics through parameter learning. By applying our model to the malaria transmission in Yunnan province, China, we evaluate the performance of our model by comparing with existing statistical methods, the results show that the forecasting results are more accurate if the underlying transmission model is employed.

Collective Review Mining for Aspect-Level Rating and Weight Estimation

Wang Feng

Abstracts: Fine-grained opinions are buried in the vast amount of review data that have been contributed by users. Generally, in a review, various opinions are expressed on multi-facet aspects of a product (e.g., service and ambiance of a hotel). Beyond such aspect-based opinions, the weights reviewers have placed on different aspects can also be reflected by the aspect opinions and the associated overall ratings. These aspect-related opinions and weights are valuable resources to know users' preferences for products and can hence be utilized by a wide range of online applications, such as aspect-based opinion summarization, personalized product rankings and recommendations. However, few of existing works have focused on the three inter-connected tasks (i.e., aspect identification, opinion rating inference, and weight estimation) simultaneously by considering their complementary effects.

Thus, in this paper, we propose a unified framework for addressing the three tasks. Specifically, they are 1) identifying the aspects mentioned in reviews of a product, 2) inferring the rating of each aspect based on the sentiments expressed in the review, and 3) estimating the aspect-level weights that reflect how much reviewers care about different aspects of a product. The relationship between these three tasks is not inherently independent in that the output of one task can influence the accuracy another task. We particularly develop a unified unsupervised model (termed as CARW), which can collectively perform all three tasks by enhancing each other mutually. Because the proposed CARW model is an unsupervised generative model, it can be easily applied to process the review data in different domains. We conduct extensive experiments on three real-life datasets to evaluate the CARW model as to its performance on conducting the three tasks. Experimental results reveal that the proposed model can achieve better results than the related methods.

Section IV:

Benchmarking the Memory Hierarchy of Modern GPUs

Mei Xinxin

Abstracts: Memory access efficiency is a key factor for fully exploiting the computational power of Graphics Processing Units (GPUs). However, many details of the GPU memory hierarchy are not released by the vendors. We propose a novel fine-grained benchmarking approach and apply it on two popular GPUs, namely Fermi and Kepler, to expose the previously unknown characteristics of their memory hierarchies. Specifically, we investigate the structures of different cache systems, such as data cache, texture cache, and the translation lookaside buffer (TLB). We also investigate the impact of shared memory bank conflict. Our benchmarking results offer a better understanding on the mysterious GPU memory hierarchy, which can help in the software optimization and modelling of GPU architectures. Our source code and experimental results are publicly available.

Reliable In-Memory Key-Value Store: Memcached with RAID

Liu Chengjian

Abstracts: Object storage applied to memory storage can introduce huge access performance compared with disk storage. Node failure in memory storage system can bring huge access performance penalty. RAID1 has been used in memory storage to obtain reliability. RAID5/6 can obtain higher storage efficiency than RAID1 in disk storage and widely adopted in the disk storage with acceptable access performance penalty. How RAID5/6 performs in memory storage system is worth to study. In this paper, we designed and implemented RAID1, RAID5 and RAID6 on top of the popular Memcached storage system. We further evaluated the access performance penalty of different levels of RAID.

Channel-Hopping Based on Available Channel Set for Rendezvous of Cognitive Radios

Yu Lu

Abstracts: Rendezvous is a necessary operation for cognitive users to establish communication links in cognitive radio networks (CRNs). To guarantee the rendezvous in finite time, all existing rendezvous algorithms generate CH (channel-hopping) sequences using the whole channel set and attempt rendezvous on each of the channels (i.e., both available channels and unavailable channels). In practice, the available channel set is usually a small portion of the whole channel set due to dynamics of channel availabilities and limited sensing capabilities of cognitive users. Thus, the CH sequences using the whole channel set may attempt unnecessary rendezvous in uncertain channels (e.g., unavailable channels or randomly-selected channels) which greatly degrades the performance. In this study, we propose a new rendezvous algorithm that generates channel-hopping sequences based on available channel set (CSAC) for more efficient rendezvous.

We prove that CSAC gives guaranteed rendezvous and derive its upper-bound on maximum time-to-rendezvous (MTTR) which is an expression of the number of available channels instead of the number of all potential channels. To the best of our knowledge, CSAC is the first one in the literature that exploits the only available channels in designing CH sequences while providing guaranteed rendezvous. Experimental results show that CSAC can significantly improve the MTTR compared to state-of-the-art.

Section V:

Is This Website Appealing to You? Explaining Electronic Commerce Consumers' Purchase Intention from an Appealing Perspective

Liu Fei

Abstracts: Business-to-consumer (B2C) electronic commerce (eCommerce) website is a technology mediated marketplace where consumers interact with only the website interface to evaluate and purchase various products. Due to the lack of direct access to products, an inherent problem of information asymmetry exists between consumers and sellers in eCommerce setting. In the present study, we adopt signaling theory as the theoretical foundation to investigate how an eCommerce website can address information asymmetry. Specifically, we incorporate the concept of appeal from advertising literature and delineate it into website appeal and product appeal as cognitive indicators of the alleviated information asymmetry. By conducting an online survey on members in a Dutch panel, we found that both website appeal and product appeal are effective determinants of online consumers' purchase intention and they can be positively influenced by sending signals related to quality of both service and products provided by a website. We also found that product appeal posits a strong positive effect on website appeal. Furthermore, trust was found to strengthen the positive relationship between website appeal and purchase intention while deterring the positive influence of product appeal on purchase intention. Both theoretical and practical implications of this study are discussed.

Knowledge Extraction and Mining in Biomedical Research Using Biomedical Rule Network

Chan Sheung Wai

Abstracts: Recent findings show that the quantity of published biomedical literature is increasing at a dramatic rate. Carrying out knowledge extraction from large amounts of research literature becomes a significant challenge. Here we introduce an automatic mechanism for processing such information and extracting meaningful medical knowledge from biomedical literature. Data mining and natural language processing (NLP) are applied in a novel model, called biomedical rule network. Using this model, information and relationships among herbal materials, diseases, as well as the chemical constituents of herbs can be extracted for effective chemical constituent discovering. Moreover, with the overlapping chemical constituents of herbs, alternative herbal materials can be discovered, and suggestions can be made to replace expensive treatment options with lower cost ones. Also the future improvement also is mentioned.

Understanding Human Mobility using the Probabilistic Modelling Approach

Li Chen

Abstracts: With the recent advent of pervasive computing and data mining technologies, there are a lot of recent research initiatives on developing algorithms to model and analyze human mobility data collected via sensors. Understanding human mobility patterns from sequences of sensor data is non-trivial due to the stochastic spatio-temporal variations in movement patterns even for the same type of activities. Also, it is often that human mobility data captured are caused by different activities which are happening in a concurrent and interleaved manner. These makes direct application of many existing machine learning algorithms unfeasible.

This thesis research investigates how the stochastic process modeling approach can be adopted for understanding indoor human mobility. In particular, we propose to use stochastic sequence models and graph clustering methods so as to infer process/flow models so that human mobility can be easier to be interpreted by providing a multi-level model representation. We have applied the proposed methodology to a publicly available smart house data set and illustrated how the inferred model can be used to gain deeper understanding of the human mobility as revealed via the data-set. Some preliminary experimental results are depicted in this prospective. Future research directions are also discussed.

Section VI:

Incorporating Structural Diversity of Neighbors in a Diffusion Model for Social Networks

Bao Qing

Abstracts: Diffusion is known to be an important process governing the behaviours observed in network environments like social networks, contact networks, etc. For modeling the diffusion process, the Independent Cascade Model (IC Model) is commonly adopted and algorithms have been proposed for recovering the hidden diffusion network based on observed cascades. However, the IC Model assumes the effects of multiple neighbors on a node to be independent and does not consider the structural diversity of nodes' neighbourhood. In this paper, we propose an extension of the IC Model with the community structure of node neighbours incorporated. We derive an expectation maximization (EM) algorithm to infer the model parameters. To evaluate the effectiveness and efficiency of the proposed method, we compared it with the IC model and its variants that do not consider the structural properties. Our empirical results based on the MemeTracker dataset, shows that after incorporating the structural diversity, there is a significant improvement in the modelling accuracy, with reasonable increase in run-time.

A new model for eye gaze tracking under natural light

Peng Qinmu

Abstracts: Eye gaze tracking is a promising technology that can be used in many applications. This paper addresses the eye gaze tracking problem using a single camera under natural light. We propose a combination of intensity energy and edge strength to locate the iris center and utilize multi-scale eye corner detector based on the curvature scale space and template match rechecking method to detect the eye corner accurately. The gaze vector formed by the iris center and eye corner is mapped on the screen to obtain the gaze point. Experimental results show the proposed method can achieve better accuracy in comparison to the existing counterparts.

Section VII:

Structure-Preserving Subgraph Query Services

Fan Zhe

Abstracts: A fundamental problem of graph databases is subgraph isomorphism query (a.k.a subgraph query): given a query graph Q and a graph database, it retrieves the graphs from the database that contain Q . Due to the cost of managing massive data coupled with the computational hardness of subgraph isomorphism testing, outsourcing the computations to a third-party provider is an appealing alternative. However, confidentiality has been a critical attribute of Quality of Service (QoS) in query services. To the best of our knowledge, subgraph query services with tunable preservation of privacy of structural information have never been addressed. In this paper, we present the first work on structure-preserving subgraph isomorphism (SPsubIso). A crucial step of our work is to transform subIso — the seminal subgraph isomorphism algorithm (the Ullmann’s algorithm) — into a series of matrix operations. We propose a novel cyclic group based encryption (CGBE) method for private matrix operations. We optimize SPsubIso by proposing static indexes and a protocol that involves the query client. We prove that the structural information of both Q and G are preserved under CGBE and derive the upper bounds of privacy preservation due to the optimizations. Our extensive experiments on both real and synthetic datasets verify that SPsubIso is efficient and the optimizations are effective.

Privacy-Preserving Reachability Query Services for Sparse Graphs

Yi Peipei

Abstracts: This paper studies privacy-preserving query services for reachability queries under the paradigm of data outsourcing. Specifically, graph data have been outsourced to a third-party service provider (SP), query clients submit their queries to the SP, and the SP returns the query answers. However, SP may not always be trustworthy. Therefore, this paper considers protecting the structural information of the graph data and the query answers from the SP. This paper proposes simple yet optimized privacy-preserving 2-hop labeling. In particular, this paper proposes that the encrypted intermediate results of encrypted query evaluation are indistinguishable. The proposed technique is secure under chosen plaintext attack. We perform an experimental study on the effectiveness of the proposed techniques on both real-world and synthetic datasets.

Online Objective Reduction for Many-Objective Optimization Problems

Gu Fangqing

Abstracts: For many-objective optimization problems, i.e. the number of objectives is greater than three, the performance of most of the existing Evolutionary Multi-objective Optimization algorithms

will deteriorate to a certain degree. It is therefore desirable to reduce many objectives to fewer essential objectives, if applicable. Currently, most of the existing objective reduction methods are based on objective selection, whose computational process is, however, laborious. In this paper, we will propose an online objective reduction method based on objective extraction for the many-objective optimization problems. It formulates the essential objective as a linear combination of the original objectives with the combination weights determined based on the correlations of each pair of the essential objectives. Subsequently, we will integrate it into NSGA-II. Numerical studies have shown the efficacy of the proposed approach.