THE CHINESE UNIVERSITY OF HONG KONG

Department of Computer Science and Engineering

MSc in Computer Science – Reminders for Students

1. Graduate School Website

- Please go to http://www.gs.cuhk.edu.hk/ for all policies related to postgraduate students especially:
  - ✓ Postgraduate Student Handbook
    https://www.gs.cuhk.edu.hk/page/PostgraduateStudentHandbook
  - ✓ Code of Practice (Taught Postgraduate Studies)
    https://www.gs.cuhk.edu.hk/page/CodeofPracticeTPg
  - ✓ General Announcement
    https://www.gs.cuhk.edu.hk/general_announcements
  - ✓ Preparation for the new academic year
    https://www.gs.cuhk.edu.hk/page/PreparationforNewAcademicYear

2. Orientation Programme for New Postgraduate Students 2017-18

- Students who are interested in the *"Orientation Camp for New Postgraduate Students"* please send register at: http://cupsa.hk/en/ocamp
- Students who are interested please send register at: http://www.cuhk.edu.hk/osa/iss/pgso.html
- Read Compass for Incoming Student at:
  http://www.cuhk.edu.hk/osa/iss/PublicationP/Compass_Eng/Compass_Sthandbook_Eng.html
- Read A Guide for New Students Journey@CUHK at:
  https://www2.osa.cuhk.edu.hk/scds/images/content/swf/Chinese_Version/book.swf
- Read Survive & Thrive at CUHK
  http://www.iso.cuhk.edu.hk/ebook/index.html#lang=zh-tw&page=1&ui=zh-tw&issue_id=1137

3. Orientation Day for MSc in Computer Science Student

Date: 2 Sept, 2017 (Sat)

Time: 10:30am - 12:00nn

Venue: To be advised - *Details will be sent out later*

4. Email & Lab Access Accounts

- Students should complete the 'User Account Application Form' with a digital passport size photo and return it to Ms. Ivy Chau at Room 1028, Ho Sin-hang Engineering Building.
- Facilities that can be accessed with the account:
  - ✓ Computing Laboratory (Rm924) at 9/F Ho Sin Hang Engineering Building
  - ✓ Faculty Common Computing Laboratory (Rm123 & Rm904) at 1 & 9/F Ho Sin Hang Engineering Building
  - ✓ Printing quota: 500 pages per month

5. Improving Postgraduate Learning

- Students are encouraged to take "*Improving Postgraduate Learning*" courses which can be found at the Website of Centre for Learning Enhancement and Research (CLEAR): http://www.cuhk.edu.hk/clear/prodev/ipl.html

6. Guidelines to Academic Honesty

- The University places very high importance on honesty in academic work submitted by students, and adopts a policy of zero tolerance on cheating and plagiarism. Any related offence will lead to disciplinary action including termination of studies at the University. Students should read through the University guidelines: http://www.cuhk.edu.hk/policy/academichonesty/

7. Study Scheme

- Coursework Requirement**:**
  Students are required to take at least 8 courses and complete a minimum of 24 credit units for graduation.

  Students can select below coded courses:

  - CMSC
  - CSCI 5000-level or 6000-level (except CSCI6500)
  - CENG 5000-level or 6000-level
  - ENGG5101, 5103, 5104, 5105, 5106, 5107, 5189

  Subject to the approval of Programme Directors concerned, students may also select at most 2 courses from below coded courses offered by other M.Sc. programmes from Divisions within the Faculty of Engineering or courses offered by M.Sc. programme in Information Technology Management:

  - BMEG5000-level
  - ECLT5000-level
  - ELEG5000-level
  - IEMS5000-level
  - LSCM5000-level
  - MAEG5000-level
  - SEEM5000-level
  - DSME5000-level or 6000-level (except DSME6790)

- Other Requirements:
  - Minimum cumulative GPA of 2.0
  - A student who obtains a cumulative grade point average (GPA) below 2.0 in the preceding term or receives a failure grade in thesis monitoring courses (for Research Postgraduate Programmes) will be put on academic probation. For details, please refer to Clause 14.0 "Unsatisfactory Performance and Discontinuation of Studies" of the General Regulations Governing Postgraduate Studies which can be accessed from the Graduate School Homepage: http://www.cuhk.edu.hk/gss.

8. Course Registration

- Period: 8 - 11 August 2017. (select courses for both term 1 and term 2)
- Online Guide on Course Selection: https://www.gs.cuhk.edu.hk/page/CourseSelectionAddDrop
- Venue of Evening Courses: classroom 119 / 126 of Productivity Training Institute (HKPC Building, 78 Tat Chee Avenue, Kowloon) except CMSC5707, 5718, 5726, 5736 (refer to CUSIS for details)
- Venue of Daytime Courses: CUHK Campus (refer to CUSIS for the most updated version)
- Check the website of our programme for the most updated timetable at http://msc.cse.cuhk.edu.hk/en/corner/timetable.html
- If you wish to register course(s) with prerequisites or course(s) outside your prescribed study scheme i.e. not appear on the course list under 'My Requirement', please complete Form CS-1 and obtain approval from the parties concerned during the course registration period.

Term 1 (Evening Courses)

Course Code	Course	Lecturer	Weekday	Venue
CMSC5724	Data Mining and Knowledge Discovery	Prof. TAO Yufei	Mon	HKPC (Rm126)
CMSC5728	Decision Analysis and Game Theory	Prof. LEUNG Ho Fung	Tue	HKPC (Rm126)
CMSC5713	IT Project Management	Dr. HARTANTO Felix	Wed	HKPC (Rm119)
CMSC5741	Big Data Technology and Applications	Prof. LYU Rung Tsong	Wed	HKPC (Rm126)
CMSC5736	Mobile Apps Design and Implementation	Dr. OR Shiu Hang	Thu	CUHK (ELB 205)
CMSC5707	Advanced Topics in Artificial Intelligence	Prof. LEUNG Kwong Sak Prof. WONG Kin Hong	Fri	CUHK (WMY 407)
CMSC5720	Project I (must be paired with CMSC5721)	Prof. CHENG Sheung Chak Prof. HENG Pheng Ann Prof. LEE Pak Ching Prof. LEUNG Kwong Sak Prof. SUN Hanqiu Prof. WONG Kin Hong Prof. ZHANG Shengyu Dr. LAM Tak Kei	/	/

Term 2 (Evening Courses)

Course Code	Course	Lecturer	Weekday	Venue
CMSC5716	Web Based Graphics & Virtual Reality Systems	Dr. PANG Wai Man	Mon	HKPC (Rm119)
CMSC5735	Advanced Topics in Cloud Computing	Prof. LUI Chi Shing	Tue	HKPC (Rm119)
CMSC5702	Advanced Topics in Parallel / Distributed Systems	Dr. HARTANTO Felix	Wed	HKPC (Rm119)
CMSC5711	Image Processing and Computer Vision	Prof. WONG Kin Hong	Wed	HKPC (Rm126)
CMSC5718	Introduction to Computational Finance	Dr. CHAU Ka Lok	Thu	CUHK (WMY 508)
CMSC5726	Computer and Network Security	Dr. CHEONG Chi Hong	Fri	CUHK (WMY 306)
CMSC5721	Project II (must be paired with CMSC5720)	Prof. CHENG Sheung Chak Prof. HENG Pheng Ann Prof. LEE Pak Ching Prof. LEUNG Kwong Sak Prof. SUN Hanqiu Prof. WONG Kin Hong Prof. ZHANG Shengyu Dr. LAM Tak Kei	/	/

Term 1 (Day-time Course)

Course Code	Course	Lecturer	Weekday	
CSCI5120	Advanced Topics in Database Systems	Prof. Ada Fu	(L) Wed 12:30pm-1:15pm (L) Thu 2:30pm-4:15pm (T) Wed 1:30pm-2:15pm	ERB 408 ERB 703 ERB 703
CSCI5350	Advanced Topics in Game Theory	Prof. LEUNG Ho Fung	(L) Tue 10:30am-12:15pm (L) Wed 10:30am-11:15am (T) Wed 11:30am-12:15pm	TBA ERB 405 LHC G04
CSCI5460	Virtual Reality	Prof. SUN Hanqiu	(L) Thu 4:30pm-5:15pm (L) Fri 10:30am-12:15pm (T) Fri 1:30pm-2:15pm	ERB 408 ERB 706 ERB 402
CSCI5570	Large Scale Data Processing Systems	Prof James CHENG	(L) Tue 1:30pm-2:15pm (L) Wed 2:30pm-4:15pm (T) Tue 2:30pm-3:15pm	ERB 703 LSK 203 SHB 904
CSCI5580	Online Algorithms for Machine Learning and Optimizations	Prof. ZHANG Shengyu	(L) Mon 4:30pm-6:15pm (L) Tue 4:30pm-5:15pm	ERB 408 ERB 402
ENGG5103	Techniques for Data Mining	Prof. CHAN Lai Wan	(L) Mon 2:30pm-4:15pm (L) Thu 1:30pm-2:15pm (T) Thu 5:30pm-6:15pm	HTB B6 ERB 803 LSB LT4

Term 2 (Day-time Course)

Course Code	Course	Lecturer	Weekday	
CENG5270	EDA for Physical Design of Digital Systems	Prof. YOUNG Fung Yu	(L) Mon 1:30pm-2:15pm (L) Tue 2:30pm-4:15pm (T) Mon 2:30pm-3:15pm	ERB 804 TBA ERB 804
CSCI5010	Practical Computational Geometry Algorithms	Prof. TAO Yufei	(L) Mon 9:30am-10:15am (L) Tue 9:30am-11:15am (T) Mon 12:30pm-1:15pm	ERB 402 FYB UG02 (ERB 713
CSCI5030	Machine Learning Theory	Prof. XU Lei	(L) Tue 2:30pm-4:15pm (L) Wed 2:30pm-3:15pm (T) Wed 5:30pm-6:15pm	TBA LHC G06 ERB 703
CSCI5050	Bioinformatics and Computational Biology	Prof. YIP Yuk Lap Kevin	(L) Tue 12:30pm-2:15pm (L) Thu 9:30am-10:15am (T) Thu 10:30am-11:15am	ERB 408 LHC 106 LHC 106
CSCI5150	Machine Learning Algorithms and Applications	Prof. XU Lei	(L) Tue 4:30pm-5:15pm (L) Wed 3:30pm-5:15pm (T) Tue 3:30pm-4:15pm	MMW 707 SWC LT MMW 703
CSCI5160	Advanced Algorithms	Prof. CHAN Siu On	(L) Mon 4:30pm-5:15pm (L) Thu 10:30am-12:15pm (T) Wed 11:30am-12:15pm	LSK 212 LSB LT3 LSK 304
CSCI5210	Advanced Computer Graphics and Visualization	Prof. FU Chi Wing Philip Prof. HENG Pheng Ann	(L) Wed 11:30am-12:15pm (L) Thu 12:30pm-2:15pm (T) Wed 12:30pm-1:15pm	LSB LT1 ERB 404 LSB LT1
CSCI5240	Combinatorial Search and Optimization with Constraints	Prof. LEE Ho Man Jimmy	(L) Mon 10:30am-11:15am (L) Wed 9:30am-11:15am (T) Mon 11:30am-12:15pm	ERB 401 ERB 713 ERB 706
CSCI5320	Topics in Graph Algorithms	Prof. CAI Leizhen	(L) Tue 12:30pm-1:15pm (L) Wed 12:30pm-2:15pm	TBA ERB 804
CSCI5440	Theory of Cryptography	Prof. BOGDANOV Andrej	(L) Mon 2:30pm-4:15pm (L) Tue 5:30pm-6:15pm (T) Tue 4:30pm-5:15pm	FYB 603 LSB C3 LSB C3
CSCI5550	Advanced File and Storage Systems	Prof. LEE Pak Ching Patrick	(L) Mon 5:30pm-6:15pm (L) Thu 2:30pm-4:15pm (T) Mon 4:30pm-5:15pm	ERB 401 MMW 707 ERB 401
ENGG5189	Adv Topic in Artificial Intelligence	Prof.LEUNG Kwong Sak	(L) Thu 4:30pm-5:15pm (L) Fri 2:30pm-4:15pm (T) Fri 4:30pm-5:15pm	LSB C3 ERB 401 ERB 401

**(L) = Lecture, (T) = tutorial*

9. Inquiry

- CSE General Office:
  Ms Ivy Chau. Tel. 3943 8485; E-mail ivy@cse.cuhk.edu.hk
- MSc Programme Director:
  Prof. Heng Pheng Ann. E-mail pheng@cse.cuhk.edu.hk

Course List (Evening Courses)

Appendix I

CMSC5702 Advanced Topics in Parallel / Distributed Systems

This course covers selected topics in parallel/distributed systems. The detailed contents may be changed from year to year depending on the current development and the teacher specialty.

CMSC5707 Advanced Topics in Artificial Intelligence

This course will cover selected topics from: advanced pattern recognition, neural networks, expert systems and fuzzy systems, evolutionary computing, learning theory, constraint processing, logic programming, probabilistic reasoning, computer vision, speech processing, and natural language processing. (Not for students who have taken CSCI6200 or ENGG5189.)

CMSC5711 Image Processing and Computer Vision

This first part of the course includes fundamental topics in image processing, e.g., image enhancement techniques, color image processing, image segmentation, and image compression. The second part of the course focuses on topics concerning methodologies of recovering 3D information from 2D images. Specifically, techniques for camera calibration, stereoposis, motion analysis, pose estimation and structure from motion will be discussed. These techniques will have practical applications to virtual reality, model reconstruction and graphics. (Not for students who have taken CSCI5280 or ENGG5104.)

CMSC5713 IT Project Management

This course covers the key elements of the project management framework related to information technology. Topics include the identification of elements and processes of project management, processes involved in project integration management, project scope management, various tools and techniques used to develop project schedules and resource planning, processes of project quality management, project communications management, risk management, human resources management, and project procurement management. Experts from industry will also be invited to share their experience in the topics. (Not for students who have taken DSME6730)

CMSC5716 Web Based Graphics & Virtual Reality Systems

This course aims to provide students the platform to learn about interactive media and virtual reality with emphasis on their internet applications. Up-to-date interactive media techniques, including modeling, rendering, illumination, texture mapping, animation, and visualization, will be introduced in the first-half of the course. The second-half course will focus on the Web-based VR interfaces which build up the fundamental basis for testing new ideas and alternative solution for the latest VR research, including VR scene modeling, dynamic objects, interactive navigation and sensors, real-time rendering, and diversifed web-based VR applications. (Not for students who have taken CSCI5460.)

CMSC5718 Introduction to Computational Finance

This course introduces some basic concepts in computational finance. Topics include risk and return,

modern portfolio theory, calculating the efficient frontier, multiple factor models, various models for portfolio optimization, utility functions and evaluation of portfolio performance.

CMSC5720/5721 Project I and II

The project provides a challenge for students to apply their computing knowledge and expertise to carry out independent research and development work in any area of Computer Science. A project report has to be written under the supervision of the lecturing staff. *(Students must take both CMSC5720 and CMSC5721 in order to have the credits counted towards graduation.)*

CMSC5724 Data Mining and Knowledge Discovery

This course introduces the techniques used in data mining. Topics include clustering, classification, estimation, forecasting, statistical analysis and visualization. Data Mining provides useful tools for the analysis and visualization. Data Mining provides useful tools for the analysis, understanding and extraction of useful information from huge databases. Applications range from business, finance, medicine and engineering. (Not for students who have taken CSCI5180 or ENGG5103.)

CMSC5726 Computer and Network Security

Issues of computer and network security. Weaknesses of network protocols. Security protocols. Firewalls. Computer viruses. System security threats. Applications of Cryptography. (Not for students who have taken CSCI5470 or ENGG5105 or CENG5240.)

CMSC5728 Decision Analysis and Game Theory

This course introduces decision theory and game theory used in computer science, in particular, artificial intelligence and multiagent systems. Topics include utility theory, decision under risk, decisions under uncertainty, social choices, strategic games and Nash equilibrium, extensive games and subgame perfect equilibrium, repeated games and folk theorems, and applications in computer science. (Not for students who have taken CSCI5350.)

CMSC5735 Advanced Topics in Cloud Computing

This course covers advanced topics in cloud computing. Topics will include new problems proposed in each year. Some topics to be discussed include: cloud computing models (e.g., SaaS, PaaS, IaaS); distributed and parallel data processing (e.g. MapReduce, Hadoop); data storage (cloud storage architectures, data centers, data deduplication); case studies of real-world cloud services (e.g. Amazon EC2, Windows Azure).

CMSC5736 Mobile Apps Design and Implementation

This course aims at introducing principles involved in mobile application design and implementation. Topics include: development and deployment process for mobile applications, graphical user interface design, mobile file systems, networking and energy saving consideration. Location & map based application will be studied specifically. Finally, the marketing plan as well as various revenue models in mobile application will also be considered.

CMSC5741 Big Data Technology and Applications

This course aims at teaching students the state-of-the-art big data technology, including techniques, software, applications, and perspectives with massive data. The class will cover, but not be limited to, the following topics: advanced techniques in distributed file systems such as Google File System, Hadoop Distributed File System, and map-reduce technology; similarity search techniques for big data such as minhash, locality-sensitive hashing; specialized processing and algorithms for data streams; big data search and query technology; recommendation systems for Web applications. The applications may involve business applications such as online marketing, computational advertising, location-based services, social networks, recommender systems, healthcare services.

Course List (Day-time Courses)

Appendix II

CENG5270 EDA for Physical Design of Digital Systems

This course aims to present the fundamental concepts and algorithms applied in Design Automation (CAD) of VLSI circuits. The scope will include various areas in Physical Design of digital systems, including circuit partitioning, FPGA technology mapping, floorplanning, placement, routing, compaction and interconnect optimization.

Prerequisite: CSCI2100 or ENGG2020 or ESTR2102 or ESTR2104

CSCI5010 Practical Computational Geometry Algorithms

This course will discuss data structures and algorithms for solving fundamental problems in computational geometry with good theoretical guarantees. Topics covered include line-segment intersection, polygon triangulation, convex hull, linear programming, orthogonal range searching, point location, voronoi diagram, delaunay triangulation, and so on.

Prerequisite: CSCI2100 or CSCI2520 or equivalent

CSCI5030 Machine Learning Theory

This course first introduces fundamentals of machine learning with a large size of samples, including basic principles (maximum likelihood vs least redundancy) and typical structures (linear systems of hidden factors, mixture of local structures, and Markov temporal models), The second part of the course covers learning theories towards small sample size challenge, including major topics (model selection, learning regularization, two stage implementation, sparse learning, and automatic model selection) and three streams of efforts, namely generalization error estimation (CV, AIC, VC theory), shortest coding length (MML vs MDL) or similarly various Bayes (BIC, MAP, Laplace, marginal, and variational) , and BYY learning (BYY system, best harmony theory, Ying-Yang alternation updating, and five action circling implementation).

CSCI5050 Bioinformatics and Computational Biology

This course introduces several core topics in bioinformatics and computational biology. Each topic will be discussed from three aspects: 1) motivation and concepts, 2) computational problems and methods, and 3) available tools and data. The topics include basics in molecular biology, high-throughput experiments and data preprocessing, sequencing and alignment, motifs and domains, ontology and functional enrichment, biological networks and data mining, secondary and tertiary structures, and other latest developments in this research area.

CSCI5120 Advanced Topics in Database Systems

This course will introduce to students advanced topics in database systems including advanced data structures, concurrency control, deadlock resolutions, recovery schemes, distributed database systems, multimedia database indexing techniques, data mining, data on the web and network data analysis.

Prerequisite: CSCI3170

Exclusion: CSCI5705 and SEEM5010

CSCI5150 Machine Learning Algorithms and Applications

This course introduces a dozen of machine learning algorithms and typical applications in computational finance, bioinformatics, and other big data analyses, including six topics that consist of (1) unsupervised learning algorithms for clustering analysis, local subspaces, manifold learning, and their applications in image analysis and bioinformatics; (2) arbitrage pricing theory (APT) and temporal factor analysis for finance market modelling and stream data analysis；(3) supervised learning algorithms(decision tree and deep learning) for pattern recognition, (4) learning biomedical case-control analyses and machine fault detection from an integrative view of description, classification, and hypothesis test；(5) algorithms of graph analysis and nonnegative matrix decomposition for learning biology networks and social computing; (6) brief introductions of other learning algorithms such as transfer learning, recommendation systems, etc .

CSCI5160 Advanced Algorithms

This course will study the design and analysis of exact and approximation algorithms using advanced techniques such as combinatorial methods, probabilistic methods, linear programming, semidefinite programming, and spectral methods.

CSCI5210 Advanced Computer Graphics and Visualization

This course provides in-depth treatment of the following advanced computer graphics and visualization topics: radiosity rendering and global illumination, procedure texturing and modelling, image-based rendering, stereo imaging, real-time volume graphics and interactive visualization.
Prerequisite: CSCI3260 or equivalent

CSCI5240 Combinatorial Search and Optimization with Constraints

Students will be exposed to various constraint-based combinatorial search and optimization techniques that arise in artificial intelligence, operations research, etc. Topics include, but are not limited to, local propagation, consistency algorithms, Boolean constraint solving, numerical constraint solving, linear programming, search, and their applications.

CSCI5320 Topics in Graph Algorithms

This course will discuss graph theory and graph algorithms with emphasis on the algorithmic aspects of graph theory. The course will cover classical topics such as search techniques, connectivity, colouring, matching and covering, network flows, planarity, traversability, perfect graphs, and NP-completeness of graph problems. The course will also discuss FPT algorithms for solving graph problems.
Pre-requisite: CSCI3160

CSCI5350 Advanced Topics in Game Theory

This course starts with a review of the basic concepts in game theory, including pure strategy and mixed strategy Nash equilibrium. It then discusses some specific types of games, such as zero-sum games and Bayesian games, and introduces other types of equilibriums including correlated equilibrium and evolutionary equilibrium. Rationalisability and the relation between knowledge and equilibrium will also be

discussed. After a review of subgame perfect equilibrium and sequential equilibrium, bargaining games, implementation theory, perfect Bayesian equilibrium and trembling hand perfect equilibrium will be discussed. Finally, the relationship between the core, the stable sets, and the Shapley value in coalitional games will be discussed.
Prerequisite: or ENGG2440 or ESTR2004, ENGG2040 or ENGG2430 or ESTR2002 and CSCI4220
Exclusion: CMSC5728

CSCI5440 Theory of Cryptography
The modern theory of cryptography studies the formal modelling and construction of computing systems that adDr.ess security concerns. This course aims to introduce the rigorous methodology that underlies the design of such systems. Topics include: * Computational foundations: Average-case hardness in NP, one-way functions * Pseudo-random number generators and pseudo-random functions * Zero-knowledge proofs and arguments * Generic protocols for secure multi-party computation * Trap-door permutations and public-key encryption * Black-box separations among cryptographic primitives

CSCI5460 Virtual Reality
This course introduces the fundamental and advanced research topics in virtual reality (VR), including VR tools & metaphors, multi-sensory interactions, geometric and behavior modeling, touch-enabled interfaces, real-time navigation, human factors in immersion, augmented reality systems, and internet-based VR applications. The web-based virtual reality interfaces plus 3D graphics engines build up the developing tools for testing the innovation for the advanced VR research and real-time applications.
Prerequisite: CSCI3260 or equivalent

CSCI5550 Advanced File and Storage Systems
This course aims to introduce important systems-level research topics in the design and implementation of practical file and storage systems. Topics include: (i) storage device organization (e.g., disk drives, disk arrays, RAID, solid state drives), (ii) file system design (e.g., log-structured file systems, distributed file systems), (iii) data availability (e.g., erasure coding techniques, data integrity protection), (iv) data consistency (e.g., journaling techniques), (v) data compression (e.g., deduplication), (vi) benchmarking (e.g., I/O metrics, benchmarking tools), etc. Depending on the current research trends, the course also discusses the latest applied storage topics, especially related to scalable and dependable big data management.
Prerequisite: CSCI3150 or ESTR3102

CSCI5570 Large Scale Data Processing Systems
This course introduces contemporary systems for large scale data processing. Topics to be covered include, but are not limited to: (1) advanced database systems (including distributed, parallel, columnar, in-memory systems, etc., for both OTLP and OLAP applications); (2) NoSQL and NewSQL systems; (3) distributed data stores; (4) big data analysis systems; (5) graph processing systems; (6) stream processing systems; and (7) data visualization. Advanced algorithms for data analytics (e.g., distributed machine learning

algorithms, streaming algorithms, etc.) that are implemented using the systems introduced in the course will also be discussed.

CSCI5580 Online Algorithms for Machine Learning and Optimizations

This course aims to cover topics in online learning and online optimization. Typical topics include multi-armed bandit (MAB) problems in various settings, online convex optimization (OCO) problems such as online linear regression, online classification, and certain general reinforcement learning problems. Different algorithms will be introduced to solve these problems and analysis of the performance and efficiency will be provided. Advisory: Students are expected to have taken CSCI3160.

ENGG5103 Techniques for Data Mining

Data mining provides useful tools for the analysis, understanding and extraction of useful information from huge databases. These techniques are used in business, finance, medicine and engineering. This course will introduce the techniques used in data mining. Topics will include clustering, classification, estimation, forecasting, statistical analysis and visualization tools.
Exclusion: CMSC5724

ENGG5189 Advanced Topics in Artificial Intelligence

The course introduces fuzzy logic and applications. Fuzzy expert systems. Fuzzy query. Fuzzy data and knowledge engineering. Fuzzy control. Genetic algorithms and programming and their applications. Parallel genetic algorithms. Island model and coevolution. Genetic programming. Introduction to emergent computing.
Exclusion: CMSC5707

Project List

Appendix III

Titles and Abstracts	Supervisor
Large Scale Machine Learning and Applications Description: Machine learning (especially deep learning) has been widely applied in industry today, and distributed machine learning techniques are in high demand in order to process massive data available in this big data era. In this project, you will learn popular and useful machine learning algorithms, and develop a toolkit for large scale machine learning on a general-purpose distributed platform. Such a toolkit will be very useful to numerous applications both in research and in industry. Students who are interested in this project may come and talk to me. My contact info and other info can be found here: http://www.cse.cuhk.edu.hk/~jcheng/ Remark: This project is particularly suitable for students who wish to work as a data engineer/analyst/scientist (all are highly demanded positions in today's market) in the future.	Prof. CHENG Sheung Chak
Advanced Techniques for Building Big Data Applications Description: Big data is one of the hottest topics today, and numerous companies in industry and almost all research institutes/universities are working on big data, especially on topics related to AI and machine learning. In this project, you will learn cutting edge techniques in processing big data, and develop new applications that are valuable to our industrial collaborators. Students who are interested in this project may come and talk to me. My contact info and other info can be found here: http://www.cse.cuhk.edu.hk/~jcheng/ Remark: This project is particularly suitable for students who wish to work as a data engineer/analyst/scientist (all are highly demanded positions in today's market) in the future.	Prof. CHENG Sheung Chak
Large-Scale Data Analytics with General-Purpose Distributed Platforms Description: Data analysis has become a do-or-die requirement for today's business world. General-purpose distributed platforms (e.g., Spark, Hadoop) have been very popular in research and also widely used in industry for large-scale data analytics. In this project, you will learn state-of-the-art general-purpose distributed platforms for large-scale data analytics, and develop new data analytics tools for real applications demanded by our industrial collaborators. Students who are interested in this project may come and talk to me. My contact info and other info can be found here: http://www.cse.cuhk.edu.hk/~jcheng/ Remark: This project is particularly suitable for students who wish to work as a data engineer/analyst/scientist (all are highly demanded positions in today's market) in the future.	Prof. CHENG Sheung Chak

A Low-latency system for Distributed Online Learning Description: Distributed machine learning is popular in both academia and industry. Current solutions, which are mostly based on Hadoop, Spark and specialized machine learning systems, process data in batches. When users implement online learning algorithms with existing systems, they usually suffer from high latency. In this project, you will learn many popular online learning algorithms such as online LSH, ALS, SGD and so on. We will further design and implement a distributed system for low-latency online learning. Students who are interested in this project may come and talk to me. My contact info and other info can be found here: http://www.cse.cuhk.edu.hk/~jcheng/ Remark: This project is particularly suitable for students who aim at pursuing master's or doctoral degrees in the future.	Prof. CHENG Sheung Chak
Dilated residual networks for automatic 3D biomedical image segmentation Automatic segmentation from 3D biomedical images, such as 3D computed tomography (CT) and magnetic resonance (MR) images, is of great significance for clinical practice. The state-of-the-art methods are based on 3D convolutional neural networks (3D CNNs). However, the 3D CNNs have a large number of parameters and are difficult for training. Nowadays, dilated convolution and skip connections are widely used in computer vision field and achieved success in other recognition problems. In this project, we will study the efficient network architecture for 3D biomedical image segmentation. Primarily, we will investigate the effectiveness of dilated residual 3D CNN for 3D biomedical image segmentation, which combines dilated convolution and skip connection techniques. The specific output of this project includes: 1) Study and compare the performance the different 3D CNNs for 3D biomedical image segmentation. 2) Proposed an efficient segmentation method based on dilated convolution and skip connections. 3) Implement the networks with one of deep learning libraries (e.g., TensorFlow, PyTorch or Caffe) 4) Show the performance of these networks on open challenge benchmark datasets (e.g., whole heart segmentation dataset, brain segmentation dataset).	Prof. HENG Pheng Ann
Deep neural netowrk compression for mobile devices Deep neural networks (DNN) have achieved success in various computer vision problems, like image classification, object detection, and semantic segmentation. However, these networks are both computationally intensive and memory intensive, and it is difficult to deploy them on mobile devices (e.g., phone) with limited computing resources. Recent efforts toward reducing these overheads involve pruning and compressing the weights of various layers. In this project, we will study and implement an effective model compression method to deploy the CNNs into mobile devices. The specific output of this project includes: 1) Implement an effective model compression technique (e.g., filter pruning) without hurting original network performance. 2) Learn a lightweight deep learning library designed for mobile devices (e.g., TensorFlow Mobile or Caffe2) 3) Develop a specific application (e.g., face detection, object detection) with this technique on mobile devices.	Prof. HENG Pheng Ann

A Virtual Reality based tool for standard plane annotation in volumetric ultrasound Standard plane is important for doctors to monitor fetal growth. Associated with broad field of view, volumetric ultrasound has unique advantages over traditional 2D planar ultrasound in capturing standard planes. However, limited by current tools, annotating standard planes in volumetric data with 2D view is indirect and inconvenient. To leverage the advantage of virtual reality (VR), we propose to develop a VR based annotation tool to greatly facilitate doctors to travel through the real 3D space, so they can annotate the standard planes more quickly and intuitively. With the project, you can learn to handle GUI, VR, volumetric ultrasound data, and the possible future work in machine learning based automatic standard plane detection.	Prof. HENG Pheng Ann
Automatic translation between multiple medical image modalities Different imaging modalities respond to specific tissues. Leveraging the complementary information between different modalities, especially the data representing the similar anatomy, is an interesting topic in medical image analysis field. Modality translation is a possible solution to the problem. In this project, we will focus on the volumetric data of whole heart, including CT and MR. Taking the Generative Adversarial Network (GAN) as a core, we propose to construct the bidirectional channel to translate CT into MR, and also MR into CT. The translation is general and may promote many applications, like volumetric segmentation.	Prof. HENG Pheng Ann
Diagnosis of skin cancer using convolutional neural networks Skin cancer is one of the most common human cancers. As the development of deep learning technique and dermoscopy technique, we can develop image analysis tools to enable the automated diagnosis of melanoma from dermoscopic images. In this project, we shall develop a deep learning method to automatically analyse the skin lesions from the dermoscopic images. We also need to develop a mobile app and integrate the model into it. The related techniques include convolutional neural networks (CNN) and tensorflow.	Prof. HENG Pheng Ann
Anomaly detection in telecommunication networks Anomaly detection is important in many domains. In this project, we will study a class of anomaly detection algorithms specifically for telecommunication networks. We will also implement them on AF-Stream, a distributed stream processing system developed by ourselves. We're also working with Huawei's Noah's Ark Lab to study the real applications of anomaly detection.. Pre-requisites: strong programming skills (in particular C/C++), interest in distributed systems Remarks: Students need to obtain prior approval from me before taking this project.	Prof. LEE Pak Ching
Geographical Information System The project aims at the development of a platform for geographical information on which information collected by users or mobile sensors (carried by volunteers, e.g. in mobile phones or mobile sensors) can be effectively captured and integrated with information amassed from software platforms such as Satellite data, Google earth, Flickers and Twitter for geo-referenced	Prof. LEUNG Kwong Sak

monitoring and decision making. Such platform will prove to be crucial for decision making in the future where ubiquitous sensors will be carried around by people and vehicles in space and time, and information has to be analyzed and courses of actions have to be decided in real time or near real time. Such platform is very instrumental to us for learning about phenomena such as pollution, microclimatic, environmental, and public health conditions of future cities, and to decide on appropriate improvement or prevention measures. This is a real-life project in collaboration with Institute of Future Cities of CUHK. The objective is to be accomplished by achieving the following sub-goals:, Goal 1. To develop a software system for the dynamic and real-time collection of multi-source information and social network from SNS (Social Network Service, e.g. Facebook, Google+) Goal 2. To develop a software platform to assimilate information collected from the mobile devices in Goal 1 with information collected by platforms such as Flicker and Twitter. The platform need to handle multi-sources, multi-temporal and spatial scale data with different visualization for effective decision making and analysis. We will also integrate the GIS with a Data Mining package such as Weka to carry out meaningful analysis and knowledge discovery. In this project, students will need to develop new software and enhance existing platform with new features according to user requirements from the Institute of Future Cities of CUHK.	
<u>Deep Neural Network for Informatics</u> Deep learning has been popularized in recent years due to the impressive performance in tasks like image recognition, speech recognition, and genome annotation. In some applications, such as human recognition in video and playing Go, and No-Limit Texas Hold'em poker, deep neural network actually outperforms human in terms of speed and capability. In this project, students will be asked to develop a specialized deep neural network system for a data analysis task. You can choose a project either in geoinformatics (e.g. pollution prediction from radar images, geo-annotation from satellite images), bioinformatics (e.g. drug repurposing, disease susceptibility via gene expression analysis) or other Chinese language related applications (e.g. Chinese speech synthesis, Chinese character recognition, Chinese text mining). In this project, you are required to do the following: • To study deep neural network (during summer holiday). • To find and understand the relevant dataset (during summer holiday). • To develop a deep neural network algorithm using either PyTorch or TensorFlow on our department GPGPU computer clusters. • To benchmark your system across a set of problems. • To extend the system to the web interface or mobile interface. This is a research-based project. Therefore, students are expected to be self-motivated. Prerequisite: Strong programming skills in Python and self-motivated to learn new things	Prof. LEUNG Kwong Sak

Equipment: GPGPU computational clusters	
SEED - Swarm and Evolutionary Enhanced Deep Intelligent System There has been overwhelming evidence to show that bees and ants are quite intelligent despite of their tiny brain. Ants can farm and cultivate crops while bees can understand abstract concepts, such as direction and combination. Furthermore, based on the Darwin's theory of evolution, the emergence of intelligence may be evolved from more simplistic ancestors naturally over time in order to overcome the challenges in different living environments. In this project, we want to create a swarm of intelligent beings via evolutionary computation. We called the whole system as the Swarm and Evolutionary Enhanced Deep (SEED) intelligent system. In this intelligent system, the swarm of intelligence is represented using a population of deep neural networks. Deep neural network is by far one of the most generic intelligent programs. These deep neural networks will collectively evolve to adapt to the fitness functions (i.e. the living environments). Some examples of evolution include optimizing the parameters, choosing the right optimizers or selecting suitable network structure responding. Since evolution takes time, a part of the system will be accelerated on multiple FPGAs. We hope that by applying evolutionary approach in a collaborative manner, we can cultivate even more intelligent programs in the SEEDI system. Lastly, the system will be apply to solve some real world problems (e.g. image recognition). Students will be asked to implement the SEED intelligent system. You are required to do the following: • To study deep neural network (during summer holiday). • To understand the library (see the list below) depending on the given FPGA hardware (during summer holiday). • To deploy a deep neural network accelerated by FPGAs. • To set up a (computer vision) benchmark for comparison across different systems. • To evolve the parameters of the deep neural network. • To evolve the structures of the deep neural network. • To evolve the training behaviors of the deep neural network. (optional) This is a research-based project. Therefore, students are expected to be self-motivated. Prerequisite: Strong programming skills in C++ and Python, and self-motivated to learn new things. Experience in OpenCL development is a plus Equipment: GPGPU computational clusters, FPGA boards	Prof. LEUNG Kwong Sak
EEG signal analysis for human pressure monitoring by AI techniques Electroencephalography (EEG) is used to record electrical activity of the brain. Many diseases can be diagnosis by EEG, like sleep disorders, coma and brain death. In this project, students are required to study how to use EEG signal to indicate pressure (stress). A wireless EEG sensor, EMOTIV Epoc+ 14 channel EEG sensor, will be used in this project, so that people can use the sensor at	Prof. LEUNG Kwong Sak

home to check their mental status. The goal of this project is to use machine learning to analyze how EEG signal related to pressure. We will be collaborating with experts form Psychology Department of CUHK. Equipment: EEG measuring device will be provided	
Free topic Any project proposed by student or agreed between the student and myself.	Prof. LEUNG Kwong Sak
Attention-driven video mashup With the popularization of mobile devices, people are taking videos of the same events everywhere. Even these videos may be uploaded and crowd sourced, the browsing experiences are very limited due to monotonous viewing and visual redundancy. The aim of this project is to develop the system that can automatically combine such videos and generate one video mashup, which summarizes the important events recorded by the videos. The system is able to determine which event is important by analyzing the joint attentions, select the efficient views according to the video quality, and automatically zoom in if necessary during the video browsing. Pre-requisite: C/C++ programming, OpenCV, Computer Vision	Prof. SUN Hanqiu
Video Super-resolution to FHD/4K Full High Definition (FHD) display is the majority nowadays, and 4K will be popular for rich-media devices. To display old video content in Standard Definition (SD) or HD resolution on these displays, interpolation used to up-scale the content will cause poor visual quality such as blurriness. In this project, we will investigate the feasibility of using image super-resolution technique to scale up the video contents. We will extend the sparse representation based image super resolution method to video super resolution problem. We will further utilize redundancy of adjacent images and update dictionary to improve sparse representation based reconstruction. Based on our previous work, we will test the extended approach on real-scene video, and measure the performance for video super resolution tasks. Prerequisite: image/video processing, C/C++ programming	Prof. SUN Hanqiu
Smart wearable glasses development Wearable glasses such as the google glass draw a lot of attention from the public because it provides novel applications to users. For example, it can display texts and information directly to users onto the glasses they are wearing. At CUHK, we designed a novel wearable glass hardware that enlarges the display area and can be built at low cost. The function is more versatile such that one can see through the display so that texts and graphics can be overlaid onto the real scene. This project is to develop software applications such as text recognition, language translation or face recognition applications that can be incorporated into our	Prof. WONG Kin Hong

wearable glass system. The project is suitable for those who are interested in system development and the creation of new products and ideas. We expect the system has high commercial potential and can be turned into a successful commercial product in future. The hardware design of our system can be found at http://www.cse.cuhk.edu.hk/~khwong/proj/msc_khw_1718.pptx	
Artificial Neural Network research for 3D computer vision Normally Artificial Neural Networks such as the Convolution or Deep Neural Networks can perform well in object recognition tasks, etc. We suggest we can use the networks for 3D reconstruction of models or camera pose tracking as well. The results can be useful in virtual reality systems and other applications such as automatic driving. Traditionally, a network takes the raw images as the input, our new idea is to feed the 2D feature tracking data to the network rather than the raw images to enable better results. We hope this can also enhance the efficiency of the network in terms of training and recognition. Our approach is not just preparing data for large scale training but rather to investigate ways to redesign a new network architecture for better performances. References are: http://www.cse.cuhk.edu.hk/~khwong/proj/msc_khw_1718.pptx, http://deeplearning.net/tutorial/lenet.html, https://www.tensorflow.org/	Prof. WONG Kin Hong
Health diagnosis by measuring pulses with a Smart-Watch It is believed pulse patterns reflect the health status of a person. Traditional, doctors of Chinese medicine diagnose diseases by measuring patterns of patient's pulses to find out the problems and cure. It is found that latest Smart-Watches can provide accurate measurements of such pulse data. By analyzing pulses using machine learning and other artificial intelligence techniques, it is hoped that it will provide enough information for accurate disease classification. The student will work on the development of machine algorithms tailored for this application.	Prof. WONG Kin Hong
Personalized recommendation systems. Recommendation systems have a wide range of applications in e-commerce, advertising, search engine, financial services, etc, Many emerging scenarios call for new recommendation algorithms. In this project, you'll first learn some basic recommendation methods, then work with me an a PhD student to try them on real-world data sets. You'll also get chances to design algorithms and fine-tuning them on real AppStore, news recommendation, and job recommendation systems. There is a large demand in recommendation system algorithm designer, and the knowledge and techniques learned in this project will make you well equipped for related job opportunities.	Prof. ZHANG Shengyu
Reinforcement learning algorithms In many practical scenarios, the data items comes in a stream and we need to make sequential decisions under uncertainty. With no or very limited knowledge about the future data, we need to explore around and at the same time aim to maximize the overall performance. Multi-Armed Bandit (MAB) and Markov Decision Process (MDP), two fundamental models in reinforcement learning, address this general problem. In this project, you'll read some of the simple but elegant algorithms with surprisingly good	Prof. ZHANG Shengyu

performance, as well as some recent development of the area, and then work with me an a PhD student to conduct some simulations to test their performances. If you have time and good ideas, you are also welcome to try to design new algorithms or heuristics to improve existing ones, and apply the algorithms on real-world datasets.	
Applications of artificial intelligence to finance Artificial intelligence (AI) has been successfully applied to many IT related companies, but much less to financial industry. What are typical applications of AI in finance and what will be the major next 5 to 10 years? In this project, you'll first survey the existing AI techniques applied in finance, and then focus on one or two of them to work with me an a PhD student to make further investigations. You'll have a chance to test your developed algorithms or methods on some real financial data.	Prof. ZHANG Shengyu
Matching and pricing on shared economics Consider the scenarios of shared taxi. At any moment, the taxi company receives a large collection of requests and taxi locations, and needs to assign (or recommend assignment) the taxis to requests. The requests may be paired up, meaning that two passengers share the same taxi. How to assign taxis and requests is an optimization problem, and how to price the passengers and charge the drivers is an interesting economical problem. In this project, you'll work with me an a PhD student to investigate this topic and try to design algorithms for different contexts. Your algorithms will then be tested on real datasets to evaluate its performance.	Prof. ZHANG Shengyu
Algorithms for fair allocations. Scenario 1: You are in a party with friends and it's time to cut the cake. You like the chocolate flag and some one else hopes to get a strawberry. In general, different people may have different preferences on different parts of the cake. Is there a way to divide the cake so that each person feels that s/he gets the best part? Scenario 2: You just arrived at HK and need to find an apartment with two other friends. After finding one, you guys face the problem of who takes which room and pays how much. Is there a way to assign the room and set their prices, so that each person feels that s/he gets the best deal in the assignment? Surprisingly, the answers to both questions are Yes! These are just two sample problems in a fascinating area called fair division or fair allocation. Do you like to know more stories, see how the fair allocation is actually done by an intriguing procedure, and even try to design some algorithms of your own? In this project, you'll read some classic problems in the area, to work with me an a PhD student to implement certain algorithms, to look for new applications, and if you have time and good ideas, also to try discovering new algorithms.	Prof. ZHANG Shengyu
Smart kiosk system This project aims to create a smart kiosk system. There are many ways to make the system smart: e.g. users may use speech (Cantonese) commands to do certain tasks; the system can analyse and figure out the usage patterns automatically via some machine learning techniques. Students can choose to make (certain parts of) the hardware and/or (certain parts of) the software.	Dr. LAM Tak Kei