THE UNIVERSITY OF SYDNEY

FACULTY OF ENGINEERING

Engineering Vacation Research Internship Program

COMPUTER SCIENCE RESEARCH PROJECTS – WINTER 2023

THE UNIVERSITY OF SYDNEY

FACULTY OF ENGINEERING

COMPUTER SCIENCE PROJECTS

CS2023/1 Multilabel Classification and Melanoma Diagnosis of Skin Lesions using Dermoscopy, Clinical Images, and Metadata

Supervisors: Prof Jinman Kim, Xiaohang Fu

Eligibility: WAM>75 and Undergraduate candidates must have already completed at least 96 credit points towards their undergraduate degree at the time of application.

Project Description:
Melanoma is the deadliest form of skin cancer, and Australia has the highest incidence of melanoma in the world. The identification of melanoma involves an integrated analysis of skin lesion images acquired using clinical and dermoscopy modalities. Dermoscopic images provide a detailed view of the subsurface visual structures, while clinical images provide macroscopic details. Visual melanoma diagnosis is commonly based on the 7-point visual category checklist (7PC), which involves identifying specific visual characteristics of skin lesions. There is an opportunity for automated methods to aid in diagnostic decision support to overcome the subjectivity and variability of manual classification. This project will focus on building a deep learning model to leverage three modalities (dermoscopic images, clinical images, and the patient's metadata) to improve classification performance of the 7PC and melanoma. Various machine learning model architectures using convolutional neural networks and graph neural networks will be investigated.

Requirement to be on campus: Yes * *dependent on government's health advice*

CS2023/2 Apply Process mining in electronic health records

Supervisor: Associate Prof Simon Poon

Eligibility:
Strong interest in medical and health informatics, and process-oriented data science. Good knowledge in data mining and machine learning, and programming skills. Students interested in honours and postgraduate research studies are encouraged to apply.

Project Description:
In recent years, process mining has been increasingly applied to healthcare to gain insights on complex, multidisciplinary clinical pathways. It has been used to inform efforts to improve patient quality of care, improve coordination between departments, and optimize the use of resources [1] It bridges the gap between clinical pathways prescribed by guidelines and the actual processes patients are undertaking. This project aims to apply process mining (PM) methods to generate process models for clinical pathways from electronic health records (eMR). These process models will allow us to study causes of pathway constraints and improvement strategies based on data-driven performance analysis. Process mining algorithms (like Heuristics, inductive and Fuzzy miners) would be used to extract process models from the event logs extracted from eMR. Different data-driven process models will be systematically evaluated to terms of accuracy and fitness in representing clinical pathways.

[1] Munoz-Gama, J. et al. (2022), Process mining for healthcare: Characteristics and challenges, Journal of Biomedical Informatics 127, 103994. URL - https://www.sciencedirect.com/science/article/pii/S1532046422000107

Requirement to be on campus: No

CS2023/3 Causal Complexity: Augmentation of set-theoretics for machine learning

Supervisor: Associate Prof Simon Poon

Eligibility Criteria: Good knowledge in data science and statistical techniques. Students with good programming skills and are interested in multi-disciplinary studies (in conjunction social science) are encouraged to apply.

Project Description:
Assessing effects from complex interactions of multiple study factors are common in empirical studies. Changing one factor may have little effect on study outcome if other factors remain unchanged. Furthermore, such interactions may extend to different configurations with complex synergistic relationships amongst many seemingly unrelated factors. From the view of statistical association analysis, especially correlation analysis with the well-known correlation coefficients has been a useful technique for identifying relationships from observational data. However, conventional statistical methods cannot account for situations in which only specific combinations of variables reveal their impact on the outcome (conjunctural causation) or all paths that lead to an outcome need to be simultaneously uncovered (equifinality). These methods also fall short in explaining situations in which a given combination of variables contributes to the presence of an outcome but at the same time is irrelevant for the absence of that outcome (causal asymmetry). In this project, we address these issues by integrating set-theoretic approaches in conjunction with statistical approaches (like regressions) for discovering meaning configurations from observational data with limited diversity.

Reference: A Configurational Analysis of Risk Patterns for Predicting the Outcome After Traumatic Brain Injury **https://pubmed.ncbi.nlm.nih.gov/29854144/**

Requirement to be on campus: Can be done remotely, but slight preference might be given to on campus students.

CS2023/4 Evaluation of AI Imaging Diagnostic Medical Decision Support Systems in Radiology

Supervisors: Associate Prof Simon Poon, Dr Neysa Petrina

Eligibility: Good knowledge of Health Information. Interest in usability research of technology in clinical context. Good knowledge in mixed method research. Students interested in multi-disciplinary studies (in conjunction with Law or Business) are more suitable for this project.

Project Description:
The implementation of artificial intelligence (AI) technology in radiology systems as support tools for clinical decision-making promises to substantially expedite the medical screening process, reduce workload burdens, and improve the quality of care provided. However, the utility of AI in a real-world radiology context is significantly impacted by human factors related to daily clinical practice which go beyond diagnostic accuracy such as user acceptability and system usability. Neglect of these in the adoption of past healthcare technologies (e.g. electronic health records, computer-aided detection software) has resulted in suboptimal outcomes such as increased clinician distrust and frustration, and degraded decision-making quality even where the technology was not functionally flawed.

This is part of an ongoing industry-collaboration research project. The aim is to systematically investigate factors contributing to Radiologists acceptance or rejection of AI, in particular Radiologists behavioural responses to AI technology in the diagnostic decision-making process. This is a health information systems project. Student enrolled in combined social/health science and engineering/computing degrees and planning to pursue honours or masters by research are encouraged to apply.

Requirement to be on campus: Can be done remotely, but slight preference might be given to on campus students.

CS2023/5 Delta Debugging for Parsers

Supervisor: Dr Rahul Gopinath

Eligibility: You will work with the supervisor directly for this project. You should be a fast learner. Excellent skills in programming (Basic Python & Java knowledge is necessary), problem solving, as well as the ability to work independently are required.

Project Description:
In cybersecurity, your fuzzers can often come up with a very large input that causes a crash. To debug the crash, one often needs a much smaller input that reproduces the crash. HDD is an algorithm used to reduce the input size if the input conforms to a grammar. However, it does not work well if the grammar is incomplete such as in many handwritten parsers and the input can't be parsed.

In this project, we will explore input-repair based techniques to repair and then parse the input and compare it with using original delta debugging (which does not require a grammar) and coverage-based strategies to identify hierarchies in the partially parsed input.

This project, if completed successfully, may be extended for a paper in one of the A/A* conferences in software engineering.

Requirement to be on campus: No

CS2023/6 Enable Design by Contract in JDK 19

Supervisor: Dr Rahul Gopinath

Eligibility: The project requires a quick learner, Furthermore, the project is to update a Java program. Hence, this calls for an expert Java programmer.

Project Description:
Design by contract is a technique which allows one to specify the pre and post conditions as well as invariants for classes and methods. This was popularized by Bertrand Meyer for Eiffel. While this enables programmers to ensure that their code is bug free, the programming language support for design by contract has been lacking. Fortunately, for Java, there has been a limited support with a previous project called cofoja from google. The library supports adding annotations to support pre-post conditions and invariants to Java source files. Unfortunately, the project is no longer compatible with new Java versions and libraries. This project asks to (1) update Cofoja so that it is up to date with new JDK and dependencies (2) Compare the contracts supported in Cofoja with those that are supported by Eiffel (3) Extend Cofoja with some of the unsupported methods (4) Verify the utility of Cofoja by converting simple opensource Java project to use Cofoja.

This project, when finished can spark the attention of industry giants such as Google, Oracle, Parasoft etc. with a significant Java codebase. The project can also serve as a foundation for future research in design by contract, which is a well-regarded software design technique.

Requirement to be on campus: No

CS2023/7 Identifying Buggy Patterns

Supervisor: Dr Rahul Gopinath

Eligibility:
You will work with the supervisor directly for this project. You should be a fast learner. Excellent skills in programming (Basic Python & Java knowledge is necessary), problem solving, as well as the ability to work independently are required.

Project Description:
Bugs are often induced by similar inputs. If we can identify and isolate such inputs, we can prevent bug inducing code from being exercised, and hence prevent vulnerabilities. The requirement for getting this to work is a large set of test cases. Fortunately, we know how to generate a large number of test cases through fuzzing. This project will explore how to learn regular expressions from such bug inducing tests such that any input that matches such regular expressions is guaranteed to trigger the bug. Such regular expressions can be then used to combine different bugs to produce more novel buggy behavior.

This project, if completed successfully, may be extended for a paper in one of the A/A* conferences in software engineering.

Requirement to be on campus: No

CS2023/8 Mutation Analysis for Fuzzers

Supervisor: Dr Rahul Gopinath

Eligibility: You will work with the supervisor directly for this project. You should be a fast learner. Excellent skills in programming (Basic Python & Java knowledge is necessary), problem solving, as well as the ability to work independently are required.

Project Description:
Fuzzers are test generators that specialize on producing millions of inputs at a time. The next frontier of fuzzing is to incorporate better oracles, that is verifiers of behaviour. However, for this to happen, we need evaluators that can verify behaviour. Mutation analysis is the gold standard for evaluating behaviour. It accomplishes this by generating thousands of possibly buggy variants of the given program and checking how many of these are detected. In order to use mutation analysis for fuzzing, its computational expenditure need to be brought under control. One way to do that is to evaluate multiple bugs at the same time in each execution. However, to do that, we need to ensure that the inserted bugs do not interact with each other. This project will explore the best way to identify independent bugs so that they can be combined.

This project, if completed successfully, may be extended for a paper in one of the A/A* conferences in software engineering.

Requirement to be on campus: No

CS2023/9 Extracting parse trees from context-free recognition matrix

Supervisor: Dr Rahul Gopinath

Eligibility: This project requires a fast learner, who is comfortable with system languages such as C and C++

Project Description:
Parsing is one of the mainstays of software engineering. We use parsing to convert inputs into internal data structures and to quickly reject invalid inputs. Hence, performance is of high importance. Traditional parsers (LR, Recursive Descent) are limited to subsets of context-free grammars and are difficult to use. Two algorithms targeting general-context

free parsing are CYK and Valiant. These are also highly parallelizable. Unfortunately, while recognizing a string is easy by constructing a recognition matrix, extracting a parse tree is more difficult. Ruzzo in 1979 showed that it is possible to extract parse trees from the recognition matrix in an extra effort of O(log(n)). This project will involve implementing this algorithm and will explore the performance of this algorithm with respect to a naïve extractor, and an extractor using backlinks using real-world workloads on real world grammars.

Requirement to be on campus: No

CS2023/10 Solving Test Reduction Slippage

Supervisor: Dr Rahul Gopinath

Eligibility: You will work with the supervisor directly for this project. You should be a fast learner. Excellent skills in programming (Basic Python & Java knowledge is necessary), problem solving, as well as the ability to work independently are required.

Project Description:
Fuzzers routinely produce very large and incomprehensible inputs that can crash a program. Such inputs can't be debugged unless a much smaller input that reproduces the crash can be obtained. Delta debugging is an algorithm used to reduce the input of a test case to make it comprehensible.

One of the problems with delta debugging is that while reducing the input, new inputs may be produced that induces an unrelated crash. This is called test reduction slippage. The problem is that slippage can result in programmers time being wasted on an unrelated problem. This project will explore how to solve the test reduction slippage using execution grammars and compare it to the effectiveness of using simple coverage.

This project, if completed successfully, may be extended for a paper in one of the A/A* conferences in software engineering.
.
Requirement to be on campus: No

CS2023/11 Extracting Key Value Access Patterns from Applications

Supervisor: Dr Rahul Gopinath

Eligibility: This project requires a fast learner, who is comfortable with working with existing code bases, and using a system language such as C.

Project Description:
This short project is within a larger funded activity that aims to improve the performance of key-value stores, which provide a simple get/put interface and are used in many applications to store data. It is important to know about the workload pattern, for example whether there are more reads (gets) compared to writes (puts) or whether some keys are accessed very frequently, because different implementations of a key-value store are beneficial depending on the kind of workload.

In this project we will seek to show how one can capture the workload pattern. To begin, we will dynamically instrument an application by wrapping external interfaces to data stores and tracking the workload. We will measure the overhead of doing so. If time allows, we may explore ways to detect changes in the access pattern over time or compare to other ways of capturing workload pattern.

Requirement to be on campus: No

CS2023/12 Error Correction with Grammars (Valiant)

Supervisors: Dr Rahul Gopinath and Sasha Rubin

Eligibility: This project requires a fast learner, who is comfortable with doing literature search, understanding the algorithms, along with the implementation.

Project Description:
Data aggregation and cleaning is one of the most important steps in data science. The data may come from multiple sources and hence, may not match the required format exactly. This is especially an issue if the required format is a rich format like JSON, XML, S-Expr etc. In such cases, we may have to rely on available error correction algorithms (at best) or manual labour (at worst). While numerous error correction algorithms exist, error correction for context-free grammars is still lagging with the best-known algorithm from 1972 from Aho et al. which extends Earley parsing. However, the algorithmic complexity of the error correcting parser is O(n^3).

This project will involve (1) understanding the background by doing literature review, (2) understanding the algorithm from the research paper, (3) implementing the algorithm for error correction using Valiant's recognizer (4) empirical analysis of this algorithm to identify its actual performance characteristics.

Requirement to be on campus: No

CS2023/13 Applying Operator Precedence to Parsed Output

Supervisor: Dr Rahul Gopinath

Eligibility: This project requires a fast learner, who is comfortable with doing literature search, understanding the algorithms, along with the implementation.

Project Description:
Parsers are one of the key interfaces between computers and human beings. A general parser takes a specification (typically provided as a grammar) the input string and generates a derivation tree. A key difficulty when using parsers is that in many languages, specific operators have specific precedence. For example, multiplication (*) typically has higher precedence than addition (+). A typical way to account for this is to try to encode this precedence in the grammar itself. However, this is error-prone, and has led to a bad reputation for general context-free parsers. The alternative is to extract the first parse tree and transform the tree to the derivation tree with the correct precedence bracketing. LaLonde et al showed how to do this for simple languages (LaLonde 1981).

This project will involve (1) doing the background literature review (2) understanding and implementing LaLonde's algorithm (3) applying this algorithm on common programming language patterns that typically are solved with handwritten precedence parsers and evaluating the effectiveness.

Requirement to be on campus: No

CS2023/14 Visualizing Parsing Algorithms

Supervisor: Dr Rahul Gopinath and and Dr. Zhanna Sarsenbayeva

Eligibility: You will work with the supervisor directly for this project. You should be a fast learner. Excellent skills in programming (Basic Python & Java knowledge is necessary), problem solving, as well as the ability to work independently are required.

Project Description:
A parser converts a given string into a parse tree, which is then used for further processing. Parsers are one of the main stays of software engineering because they form the interface

between the human world and the program. However, programmers still find the operation of advanced parsers complex and mysterious and falls back on traditional recursive descent or regular expression-based parsing. This project is an attempt to make the operation of advanced parsers more intuitive. It will investigate how to visualize the operation of multiple parsing algorithms with the aim of making the algorithm operation visible and intuitive for the programmers.

This project, if completed successfully, may be extended for a paper in one of the top conferences such as "The Visual Computer"

Requirement to be on campus: No

CS2023/15 Privacy reserving federated recommendation systems

Supervisor: Dr Wei Bao

Eligibility: Python programming experience. Knowledge in machine learning. Knowledge in recommendation systems or social networks.

Project Description:
Recommendation systems are widely adopted in a variety of areas to predict users' preference and rating on items. It is extremely useful for online stores and social networks. Traditional recommendation system based on conventional techniques, such as collaborative filter, have become mature. However, conventional recommendation systems used users' historical rating to build recommendation models. They require users to exchange their ratings explicitly, causing privacy issues. Federated learning (FL) addresses this issue. A server and users collaboratively train a machine-learning model by privacy-aware interactions.

We are motivated to employ FL for recommendation systems. We plan to investigate how traditional mathematical models (e.g., collaborative filters) can be adaptively employed in the FL environment. In additional, we target to exam how the convergence performance is influenced in the FL environment. Finally, we aim to study the robustness of federated recommendation systems against malicious user attack (e.g., model poisoning).

Requirement to be on campus: Yes * *dependent on government's health advice*

CS2023/16 Online Bipartite Matching in Social Networks

Supervisor: Dr Wei Bao

Eligibility: Knowledge in online algorithms. Good programming skills

Project Description:
Bipartite matching is a fundamental problem in graph theory and computer science, which involves finding the largest possible matching between two sets of vertices in a bipartite graph. In the online version of bipartite matching problem, the vertices and edges of the bipartite graph are revealed one at a time, and the goal is to maintain a matching as new vertices and edges are added to the graph.

The online bipartite matching problem is important in many social network applications such as ride sharing (e.g., Uber), job seeking (e.g., Seek) and dating matchmaking (e.g., Hinge) because it provides a way to match individuals in a fair and efficient manner, even when the preferences and characteristics of the individuals are not known in advance.

In this project, we expect to investigate several state-of-the-art algorithms to achieve optimal online bipartite matching in social networks under different assumptions (e.g., known adversary distribution, fully adversary, etc.) The performance of the algorithms will be tested and verified in real-world datasets.

Requirement to be on campus: Yes * *dependent on government's health advice*

CS2023/17 Economic Mechanism for Personalised Federated Learning

Supervisor: Dr Wei Bao

Eligibility: Knowledge in federated learning, game theory, and optimisation.

Project Description:
Federated learning is a distributed machine learning paradigm, in which many clients collaboratively train a shared learning model under a server's coordination. It protects clients' data privacy by allowing clients to keep their data on their own devices and only share intermediary model parameters with the central server. However, with statistical data heterogeneity, the trained global model may not generalise well on the local data of each individual client. Therefore, if a client wants to obtain a good, personalised model, it requires the server to take some actions (e.g., adjust learning objective) instead of solely optimizing for the global model accuracy.

Besides technical considerations for the model personalisation in federated learning, in this project, we also aim to tackle some economic issues to incentivise clients to participate personalised federated learning. For example, the server may design a pricing mechanism for personalised models with different qualities. A client with a large amount of local data can train a local model with good performance by himself, so that the server should pay extra. We aim to design such economic mechanisms for personalisation issues in federated learning.

Requirement to be on campus: Yes * *dependent on government's health advice*

CS2023/18 Neural Interfaces for Visually Impaired as Tactile Enabling Technology

Supervisor: Dr Anusha Withana

Eligibility: You will work with the supervisor and a PhD student, and we expect you are a fast learner. Excellent skills in programming, skills in embedded systems, machine learning, knowledge in design and fabrication, and human computer interaction are added benefits.

Project Description:
Modern computers frequently use visual and auditory interfaces. For example, we can see the visual information on our screens and hear the audio computers generate. However, one important modality is less explored, that is, what we feel through our skin. We feel vibrations, temperature, and pressure through sensory receptors in our skin. This modality is particularly important to create enabling interfaces, for instance computer interfaces used by visually impaired people. In this project, we will explore how we can create enabling technologies using novel tactile interfaces, particularly an interface directly communicates with our neural system.

Requirement to be on campus: Yes * *dependent on government's health advice*

CS2023/19 ShoeSense: Non-optical methods for tracking in-shoe foot motion

Supervisors: Dr Anusha Withana, Dr Alycia Fong Yan and Prof Alistair McEwan

Eligibility:
You will work with the supervisor and a PhD student, and we You will work with three supervisors (computer science/health and medicine/bio med), and we expect you are a fast learner. Excellent skills in programming, skills in embedded systems, machine learning, knowledge in design and fabrication, and human computer interaction are added benefits.

Project Description:
Understanding the effect of footwear on foot motion has many health benefits for all populations. It is also important to inform footwear designers and manufacturers. The current gold standard optical method of motion capture uses retro-reflective or active markers captured with multiple infrared cameras. This requires cutting holes in the shoe to place the markers and cannot be used outside lab settings. In this project we plan to develop a novel, magnet based wearable sensor to track foot movements inside the shoe without any modification to the shoe. Currently we have developed initial prototypes using sock worn magnets. The goal of this project is to improve the sensor and collect data and analyse them.

Requirement to be on campus: Yes * *dependent on government's health advice*

CS2023/20 Comparing Graph Query Languages

Supervisor: Prof Alan Fekete

Eligibility: Experience writing SQL queries is expected. Understanding of graph concepts (from discrete maths or data structures) is desirable.

Project Description:
Many applications need to maintain data that can be described by a graph (with labels on nodes and edges). For examples, a social network, a supply chain, a transport network, these can all be captured as graphs. Several platforms offer database capability for graph data (e.g., Neo4j, Amazon Neptune, ArangoDB, dgraph, and many more). There are many query languages proposed for accessing graph data, some similar to SQL and others more imperative in style. This exploratory project will be done in the context of an emerging collaboration with dgraph, to compare different approachs, by taking sample data schema, and user needs, and writing queries in the various languages to extract the same information.

Requirement to be on campus: No

CS2023/21Co-word analysis of literature on emotion tracking

Supervisor: Dr Zhanna Sarsenbayeva

Eligibility: You will work with the supervisor directly for this project. You should be a fast learner. Excellent skills in programming (Basic Python & Java knowledge is necessary), knowledge of stats, problem solving, as well as the ability to work independently are required.

Project Description:
Co-word analysis is widely used to analyse textual content. It focuses on understanding the relationship between terms in a text and is used for mapping patterns and trends of associated words. Given these features, co-word analysis offers a powerful bibliometric approach to map the evolution and assess the structure of scientific disciplines using publication data, including metadata, titles, abstracts, and keywords. Ideally, keywords are used to describe the content of a research article. Thus, co-word analyses of keywords can reveal the conceptual structure and evolution of the research topics within the area or field in focus, based on the interaction of the respective keyword. To define a conceptual structure and the characteristics of a research area or field, we use keyword networks and clusters of keywords. In this work you are required to perform co-word analysis of literature on emotion tracking. You will learn how to program in R and apply statistical analysis and hierarchical clustering in the real-world scenario.

This project, if completed successfully, may be extended for a paper in one of the top venues of the HCI research field.

Requirement to be on campus: No

CS2023/22 Using advanced IoT-based machine learning for in-home quality ageing

Supervisor: Prof Athman Bouguettaya

Eligibility: Experience in programming in Python or Matlab; Able to implement machine learning algorithms; Experience with backend developments (preferably NodeJS); Working knowledge of database technologies (e.g., MySQL).

Project Description:
With the fast development of IoT sensors and other digital technologies, it is now possible to collect privacy-preserving health data from private dwellings and homes. The project aims to design and develop trustworthy IoT-based machine learning algorithms to leverage in-home behavioural data coupled with contextual data related to age group, ethnicity, etc., to provide a highly personalised approach to detecting and preventing adverse events, which would negatively affect the graceful and quality of in-home ageing. The aim is to accurately predict and prevent adverse health events using advanced and highly personalised techniques using non-obstructive and invisible technologies.

Requirement to be on campus: Yes **dependent on government's health advice*

CS2023/23 WebIoT: A portal for dataset collection, visualization and analytics

Supervisor: Prof Athman Bouguettaya

Eligibility: Good knowledge and skills in web application development, intuitive UI design, experience with frontend and backend (e.g., Mongo DB and Node JS).

Project Description:
The data generated from IoT devices may provide valuable information to provide efficient services such as drone delivery in smart cities. Moreover, it can help researchers to test and validate their proposed systems and solutions. Therefore, there is a need to build a secure web portal that manages multiple data sets collected from multiple IoT devices for data analytics. The IoT devices include thermal cameras, smart plugs, drones, and context-aware devices. The goal of this project is to provide researchers and developers with quick access to request and download data sets. The portal will function as a centralized platform for all data collected. In addition, the portal will periodically collect data from context-aware IoT devices (e.g., motion and weather sensors) installed at the school of computer science campus. The portal will be designed with a focus on data analytics, featuring an intuitive and user-friendly interface for data exploration and visualization. The portal will also have the capability to integrate additional data sets in the future. Security is a top priority in this project, and measures should be taken to protect sensitive information. This includes implementing robust authentication and authorization protocols and secure data storage and transmission. This project offers hands-on experience in designing and building a secure data integration portal.

Requirement to be on campus: Yes **dependent on government's health advice*

CS2023/24 Trust in Crowdsourced Social Media Images

Supervisor: Prof Athman Bouguettaya

Eligibility: Programming in Python, writing scripts to pre-process data, basic machine learning skills, writing and communication skills

Project Description:
Social media has become an essential part of our lives. Social media users upload a lot of

content on different social media platforms. The content shared on social media may contain untrustworthy images. Existing approaches relies on image processing to detect fake images. These solutions are costly and computationally intensive. Some recent solutions investigate comments on a post to assess credibility of an image. These solutions may not accurately determine the credibility of an image because fake posts on social media can still get supportive comments. We leverage the metadata of an image and the related posted information to determine the trustworthiness of a crowdsourced image. This information may reflect some changes in the image. For instance, there may be some discrepancies in the metadata which might be an indication of modifications in the image. Therefore, investigating this meta-information may provide useful insights about the changes introduced in the image.

Requirement to be on campus: Yes **dependent on government's health advice*

CS2023/25 Efficient Algorithms for Large-Scale Social Network Analysis

Supervisor: Associate Prof Lijun Chang

Eligibility: WAM>75 and Undergraduate candidates must have already completed at least 96 credit points towards their undergraduate degree at the time of application.

Project Description:
Graph model has been widely used to represent the relationships among entities in a wide spectrum of applications such as social networks, communication networks, and biological networks. We are nowadays facing a tremendous amount of large-scale real-world graphs with millions or billions of edges. Thus, there is a need of designing efficient algorithms for processing large graphs. In this project, our aim is to design efficient algorithms to speed up graph processing on the ever-growing large graph datasets. The problems that we will be investigating can be (1) dense subgraph (e.g., clique, k-plex) computation over a large sparse graph which finds one dense subgraph of the maximum size, or (2) dense subgraph enumeration which enumerates all maximal dense subgraphs.

Requirement to be on campus: No

CS2023/26 Understanding Non-transferrable NFT

Supervisor: Dr Qiang Tang

Eligibility: Familiar with blockchain or cryptography

Project Description:
NFT is a popular token in the blockchain space that represents broader version of assets. Recently, people have proposed Soulbound Tokens that are not transferrable/exchangable as usual tokens. We will do a systematic study of the rationale, design principle and potential new applications of such non-transferrable tokens.

One reference point could be
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4105763

Requirement to be on campus: No

CS2023/27 Understanding Federated Consensus in Global Decentralized Payment Systems of Stellar and Ripple

Supervisor: Dr Qiang Tang

Eligibility: Theory oriented and is comfortable with formal reasoning

Project Description:
New Decentralized Payment Systems such as Stellar and Ripple gets more and more popular for cheaper, faster and more resilient international payment and transfer applications.

However, their consensus mechanism is still under investigation whether they can indeed provide needed security. We will try to understand how they work and analyze its security.

Ref:

https://assets.websitefiles.com/5deac75ecad2173c2ccccbc7/5df2560fba2fb0526f0ed55f_stellar-consensus-protocol.pdf

https://link.springer.com/article/10.1007/s00446-022-00430-0

And there are many blog posts on relevant topics.

Requirement to be on campus: No

CS2023/28 Enhancing Robustness and Efficiency of Deep Neural Architectures with Neural Dilation and Neural Drop

Supervisor: Dr. Chang Xu

Eligibility: We are looking for a student who has the following qualifications:

1. Strong knowledge of Deep Learning and Computer Vision: The student should have a solid understanding of the principles and techniques used in deep learning and computer vision. It would be a plus if the student is familiar with the key concepts, such as convolutional neural networks (CNNs), Neural Architecture Search (NAS), and adversarial robustness.
2. Proficient in Python and PyTorch: The student should have a strong programming background, particularly in Python and PyTorch. They should be able to write clean, efficient code, and have experience with implementing deep learning algorithms using PyTorch.
3. Ability to read papers: The student should be able to read and understand papers related to deep learning and computer vision. They should have experience in critically evaluating research papers, identifying the key contributions, and understanding the limitations of the proposed approaches.

Project Description:
This research project addresses the challenge of adversarial attacks on deep learning models and proposes a method to improve their adversarial robustness while maintaining accuracy on clean inputs. Adversarial robustness refers to the ability of deep models to make correct predictions when small perturbations are added to the input. The proposed approach involves adding new blocks to an existing convolutional neural network backbone to improve the adversarial robustness-accuracy trade-off. However, adding new blocks can increase computational overheads, so a method of "drop" is proposed to remove some of these blocks. By removing certain blocks, the computational overhead can be reduced without significantly impacting the performance of the model. This approach provides a practical way to enhance the reliability of deep learning models in real-world applications where adversarial attacks are a significant concern.

Requirement to be on campus: Yes **dependent on government's health advice*

CS2023/29 Study of social dynamics for health management

Supervisors: Dr Sue Ch'ng (Primary) and Simon Poon (associate)

Eligibility: Good programming skill and interest in simulation studies, network science and social network analysis are encouraged to apply.

Project Description:
The Guardian Angel concept proposes the idea that a large, decentralised social support network can more effectively motivate individual in a connected community to collectively improve a population's health and lifestyle habits than a traditional centralised system of a few localized hubs, e.g., health care professionals (e.g., clinicians), monitoring a large number of spokes (e.g., patients). The Guardian Angel idea is to organise patients, so that each individual has a guardian angel (or more than 1 guardians) from among the other patients (defined as children) and is assigned to be a guardian angel for someone else. The angel has the ability to engage with the other person's activities and then encourage and motivate the person to continue on a positive trajectory (to accomplish health goal). Every person tries to motivate someone and is motivated by someone else so that the health and lifestyle of the whole community improves.

The aim of this project is to develop an agent-based model using simulation to study the dynamics of different social network topologies.

Reference: Socialising Health Burden Through Different Network Topologies: A Simulation Study. Stud Health Technol Inform. 2017; 239:112-118. PMID: 28756445.
URL: https://pubmed.ncbi.nlm.nih.gov/28756445/

Requirement to be on campus: No

CS2023/30 DB4ML – In-Database Machine Learning

Supervisor: Associate Prof Uwe Roehm

Eligibility: Strong academic background in Computer Science, especially databases, and C programming skills.

Project Description:
Machine Learning algorithms are not well supported by existing database systems as they typically iterate over the dataset until a convergence criterion is met. This is not supported, e.g., by SQL. Hence extensions with user-defined functions (UDF) are needed, which however don't parallelise very well.

We have developed a novel approach to parallelise machine learning algorithms inside databases with multi-version storage layer, called DB4ML. In this project, we want to benchmark our existing DB4ML prototype against existing UDF-based approaches such as MADlib for PostgreSQL. We will use simple machine learning algorithms, such as clustering or PageRank. This project will give you a great introduction into machine learning with databases and has the potential for being used in a publication.
Some database and programming skills will be needed for installing and preparing both approaches for benchmarking.

Requirement to be on campus: No

CS2023/31 Benchmarking a DRL-based Query Advisor for MongoDB

Supervisor: Associate Prof Uwe Roehm

Eligibility: Strong academic background in Computer Science, especially databases and ideally also machine learning (the latter can be picked up during the project too). This project also requires good Python programming skills. Experience in benchmarking and tuning of systems would be desirable.

Project Description:
MongoDB is a popular NoSQL database system which allows to partition data distributed over multiple nodes for fast performance and scalability. An important design decision is how data gets indexed so that only specific partitions need to be contacted for a given query. In previous work, we have shown that MongoDB's own query optimiser sometimes selects a sub-optimal query execution plan by wrongly estimating the benefit of indexes.

Together with MongoDB, we are working on an external query advisor that is using deep reinforcement learning (DRL) to direct MongoDB towards the optimal plan by monitoring past query executions and then adding suitable execution 'hints' to new queries. In this project, the task is to take our existing query advisor and benchmark its performance against the latest version of MongoDB. Most of the required code, especially the DRL part, already exists from a previous project, written in Python. This project will give good insight into the usage of machine learning for the auto-tuning of database systems, as well as into MongoDB in particular – the project is a collaboration with MongoDB, and its outcome is planned for being used in a publication about our approach.

Requirement to be on campus: No

CS2023/32 Faithful Visual Analytics of Big Complex Data

Supervisors: Prof. Seokhee Hong, Dr. Amyra Meidiana

Eligibility: Skills Required: Data Structure and Algorithms and Programming (Java, C++, Python, Javascript)

Project Description:
Technological advances have increased data volumes in the last few years, and now we are experiencing a "data deluge" in which data is produced much faster than it can be understood by humans.
These big complex data sets have grown in importance due to factors such as international terrorism, the success of genomics, increasingly complex software systems, and widespread fraud on stock markets.

Visualisation is a powerful tool to compute good geometric representation of abstract data to support analysts to find insights and patterns in big complex data sets.

This project aims to design, implement, and evaluate new visualisation algorithms for faithful visualisation of big complex data, to enable humans to find ground truth structure in big complex data sets, such as social networks and biological networks.
These new visualisation methods are in high demand by industry for the next generation visual analytic tools.

Requirement to be on campus: Yes **dependent on government's health advice*

CS2023/33 Evaluating MongoDB with LinkBench+

Supervisors: Associate Prof. Uwe Roehm and Prof Alan Fekete

Eligibility: Strong academic background in Computer Science, especially databases and ideally NoSQL databases. This project also requires good Java programming skills. Experience in benchmarking and tuning of systems would be desirable but can be picked up during the project.

Project Description:
LinkBench is a database benchmark for document stores, such as MongoDB, a popular NoSQL database system. This benchmark is used to test the performance of a document store for both queries and updates, but with the limitation that it does not check for the correctness of the updated documents. In fact, in its original form, it can produce some incorrect executions.

We are currently working together with MongoDB on an improved version of LinkBench that does not have this limitation anymore. The task of this project is to take this improved version and conduct an experimental performance comparison of different document schemas – such as nested documents versus linked document instances – using the improved LinkBench version. Most of the required benchmark code, written in Java, already exists from a previous project. This project will give good insight into the benchmarking of NoSQL database systems in general, and MongoDB in particular. The results of this project are planned for being used in a publication about our approach.

Requirement to be on campus: No