

COMPUTING POSTGRADUATE (RESEARCH) MINI-CONFERENCE

Abstracts - Presentations on 9 Nov 2004

.....

9.15-9.30 *Feature Subset Selection Application of Data Mining*

Yihao ZHANG

Supervisor: A/Prof Mehmet ORGUN

Abstract

Feature subset selection has grown up to the core research area in multi-dimensional dataset analysis and database mining field. It is beneficial for data mining applications where datasets have lots of properties available such as data mining of application for which datasets with huge quantities properties are available. Generally, feature selection can perform two main functions in the whole data analysis procedure. On one hand, it creates a mature platform to provide faster and more cost-effective predictors. On the other hand, it supports a better understanding of infrastructure that generated the stream of data analysis.

9.30-9.45 *Online approaches to developing expertise in Computer Science learners*

Matthew BOWER

Supervisor: Prof Michael JOHNSON

Abstract

Compared to a discipline such as Mathematics, research into Computer Science Education comprises a relatively small body of literature. What's more the interaction between the technological context, the linguistic aspects and the applied nature of Computer Science means that developing expertise in the novice learner proposes a unique challenge to educators, one that is distinct from teaching in other learning domains.

This presentation presents an outline of how an integrated technological approach to developing expertise in Computer Science learners will be developed. Results of introductory research into student learning preferences, results of their response to multimedia lecture systems and virtual classrooms and results from Computer Science Education literature will be presented.

9.45-10.00 *Dynamic Topologies in Mobile Networking*

Nirisha SHRESTHA

Supervisor: A/Prof Bernard MANS

Associate Supervisor: Dr Mark DRAS

Abstract

A mobile ad hoc network (MANET) is an infrastructure-less network where wireless nodes move in arbitrary manner, dynamically creating and breaking connections, and communicating in multi-hops. Designing MANET protocols comes with challenging factors, in particular, wireless medium and mobility results in constraints in resources like bandwidth, capacity and energy. The objective of this research is to design distributed algorithms for dynamically finding energy efficient routes in presence of bandwidth and capacity constraints. In the talk, I will first present some preliminary results obtained in a study of the interference and collision impact on a well-established protocol. I will then conclude with my current work and future directions, which includes studying the effect of bandwidth constraints on power control in ad hoc topologies.

10.00-10.15 *Conceptual and constraints framework for semi-structured information domains*

Gillian MILLER

Supervisor: A/Prof Dominic VERITY

Associate Supervisor: Prof Michael JOHNSON

Abstract

My project is a framework for the specification and verification of conceptual and semantic constraints for information systems. As part of this framework I am developing ConCL - (Conceptual Constraint Language) - a data model and domain specific language for expressive system specification that can also be applied to validate data and check for conformity to this specification. The goals of the language are to be declarative, to be able to be used by information system practitioners, to be concise, to be readable, have a rich expressiveness but also be formally based and effectively implementable. The other goal is to make the model and language generic so as to be able to assimilate conceptual modeling ideas and disparate data representations for information that is being integrated from the oo, relational and semi-structured paradigms. The language is comparable to a query language, but with a primary concern of being able to concisely express logical expressions, invariants and structural constraints. One application of the overall framework is targeted at XML domains which have very limited semantics provided via XML schemas and DTDs.

This year I have been working on the detailed design of ConCL. In this presentation I will discuss the design and some of the key ideas that have emerged - the underlying data model, association path navigation, concept formation, extent constraints, association classification and constraints. I will also discuss some of the issues I have faced during the design process - why Yet Another Language, and why I have chosen expressiveness (and an algebraic approach) over tractability of reasoning.

10.15-10.30 *Automatic Vectorisation of Functional Code*

Matthew ROBERTS

Supervisor: A/Prof Dominic VERITY

Associate Supervisor: Dr Anthony SLOANE

Abstract

SIMD (Single Instruction, Multiple Data) and stream processing units are becoming common, cheap and powerful components in modern PCs. However, compiler technology has not been able to keep pace with the changes in hardware. It is very hard to get any performance gains with SIMD and stream processing hardware from modern compilers without hand coding the optimisations. In this project we take the approach that functional languages and their compilers provide a fertile and largely untouched ground for research into improving automatic compilation for these hardware units.

SIMD units are shipping in the CPUs of almost every computer sold today and cheap, consumer level graphics cards are very powerful stream processing units. Thus, the focus of this work will be to target this hardware from functional programming language compilers in the hope of making better use of the available hardware without requiring non-portable language extensions. We have chosen to work with functional languages because they have characteristics that are a natural fit to the problem and there has already been extensive work in this area for imperative languages.

11.00-11.15 Evaluation Framework for Tools to Support Requirements Inconsistency Management

Jyothi IYER

Supervisor: Prof Ray OFFEN

Associate Supervisor: Dr Deborah RICHARDS

Abstract

Inconsistency in software requirements is a well-acknowledged problem that has the potential to lead the project to failure, cost overruns or time delays. However, equally acknowledged is that inconsistency is inevitable in real life and in some cases desired. Therefore, the solution is not plain inconsistency detection and removal, but, that of managing inconsistency. Manually managing inconsistency is effort intensive. Automated support promises greater reliability and reduced effort. A survey of RE tools conducted reveals an essential lack of tools supporting requirements inconsistency management. This research presents an evaluation framework and a list of requirements, to enable users compare tools and facilitate tool vendors' develop/enhance products respectively.

11.15-11.30 Design Decisions SATEMA

David FOSTER

Supervisor: Prof Vijay VARADHARAJAN

Abstract

This paper investigates some of the design decisions encountered during the development of the Security and Trust Enhanced Mobile Agent. The topics for this paper includes Single or Multiple Hop Agents, Single or Multiple Agents, per agent owner, Single application Agent vs Generic Agent, Modular/ layered design, Security, Trust Policy, Privilege Policy and Realistic Business Model. Each of these topics will be defined and the method of implementation discussed.

11.30-11.45 Image watermarking for copyright protection

Kourosh FALLAH-MOSHFEghi

Supervisor: Dr Len HAMEY

Associate Supervisor: Prof Josef PIEPRZYK

Abstract

The exponential growth of using digital media has revolutionized the multimedia industry. Combined with the widespread use of the Internet it can be the perfect means for a new area of global trade. Unfortunately, digital media can be very easily and almost perfectly reproduced by the army of interested illegal users. Content providers are desperately seeking secure means to prevent or at least trace digital piracy.

Encryption is a well-established method. However once the content is decrypted to be played or used, it will be vulnerable to any kind of recording. Also decryption will be an extra load for promotional samples of media. In watermarking, some digital information is embedded imperceptibly inside the main content. In case of a copyright infringement, this information can be retrieved and used for owner identification. Other possible applications of watermarking include content authentication (e.g. to prevent faking images) and multimedia cataloguing. Although watermarking can be used for all types of digital media, I will focus on image watermarking.

This presentation consists of three parts. In the first part I introduce the basic concepts of watermarking. Considering the main requirements (imperceptibility, robustness and security and capacity) of watermarking algorithms we will see that different types of application have different requirements. Different watermark types are designed to meet these varying requirements. Robust watermarks are designed to survive virtually all attacks and are the best

choice for copyright protection. Fragile watermarks are destroyed by any attack and suit content authentication. Semi-fragile watermarks usually survive limited types of manipulation.

In the second part I talk about the state-of-the-art watermarking algorithms and attacks on them. Although a number of watermarking algorithms are commercially available now, no watermarking algorithm is yet known to be secure enough for widespread commercial use. Some researchers believe that watermarking will never be able to solve the copyright protection problem. I briefly consider main classes of watermarking algorithms and their relative advantages and shortcomings. A short review of attacks on watermarking algorithms will follow.

In the third part of the talk I'm going to explore a few research tracks that I intend to follow. Most watermarking algorithms are designed to survive global attacks, i.e. those that affect the whole picture or certain areas of it (e.g. perceptually significant parts). This makes them vulnerable to local random attacks that manipulate small portions of the image randomly. Stirmark, a previous attack in this category, remained unbeatable until recently. A new type of these attacks is yet to be tested on state-of-the-art watermarking algorithms. Artificial intelligence methods have been successfully used in a number of pattern recognition problems such as handwritten character recognition. Those might be useful in this application too. These methods can also be used to find the best local random manipulation to attack a watermarking algorithm. Other suggested tracks are finding the optimum marking space (DFT, DCT, Wavelet, etc) based on the content and application, resisting against protocol attacks (e.g. copy attack) and optimum watermark detection methods.

11.45-12.00 *Improvement of Precision and Recall for information Retrieval in a Narrow Domain: Reuse of Concepts by Formal Concept Analysis*

Woochul CHO

Supervisor: Dr Deborah RICHARDS

Associate Supervisor: Dr Stephen CASSIDY

Abstract

With the exponential growth of the World Wide Web (WWW), it has become the most popular place to gather information. However the size of the WWW makes it difficult for people to locate relevant information. About 85 % of all Web users use search engines of some kind for this purpose. However, existing search engines often do not return relevant information. The main focus of this paper is to improve search performance by using keywords and web pages which have been previously used or visited by other users. The Formal Concept Analysis (FCA) method has been adapted to maintain a concept map of keywords. This paper shows how both precision¹ and recall² has improved the specific-domain area, in which users can share the same knowledge.

¹ Percentage of total relevant documents retrieved from all documents.

² Percentage of total relevant documents retrieved that the searcher is actually interested in.

13.30-13.45 *Table Extraction From Free Text Documents*
Vanessa LONG

Supervisor: Prof Robert DALE

Associate Supervisor: Dr Rolf SCHWITTER

Abstract

Tables are popularly used in documents because of the expressive power they have in conveying information. The downside of tables is that they create new challenges to document processing. The purposes of my research are to address the challenges and to provide a theoretical framework for table processing. The success of our work relies heavily on having a reliable table detection algorithm. In this conference, we would like to present a solution for detecting tables from plain text documents based on layout features.

13.45-14.00 *Towards a Peer2Peer World-Wide-Web for the Broadband-enabled User Community*

Constantine MANTRATZIS

Supervisor: A/Prof Mehmet ORGUN

Abstract

This paper aims to study the concept of a distributed World Wide Web archive that complements the existing WWW and “lives” across a vast Peer-to-Peer network of broadband-connected user nodes.

It proposes the sharing of a web browser’s cached data with other peers in an effort to provide an alternative resource to “discontinued” web documents with [normally] short life spans such as video and audio content as well as frequently restructured text pages.

We have based this study on the success of existing file-sharing Peer-to-Peer networks and aim to extend their use further to facilitate content-oriented usage more appropriately while at the same time, addressing some of the major problems that arise from this.

14.00-14.15 *Biometric Authentication using a neural network model of facial recognition*

(Sharon) Wai Han HO

Supervisor: Dr Paul WATTERS

Associate Supervisor: A/Prof Dominic Verity

Abstract

Biometric authentication is the new direction in the area of security, and much attention has been given particularly to face recognition in recent years. Preliminary studies have showed that most of the previous researches had treated the problem solely as another problem of pattern recognition and image analysis. While there are merits in the approach, our research will take on another approach.

There are various issues affecting how a person looks in an image and the success recognition of him. The presentation will discuss the issues and the particular area that the research will work on.

14.15-14.30 *Model and Techniques for Solving Privacy/Security Issues in Medical Domain.*

Aungkoon KHUMKOMGOOL

Supervisor: A/Prof Mehmet ORGUN

Associate Supervisor: Prof Vijay VARADHARAJAN

Abstract

Unauthorised revealing of medical records, which contain private/sensitive information, can lead to prejudice or lost of job. An alerting process to inform users to the sensitive records should be employed in the system. This project is to develop a model that can warn the users before revealing the sensitive information, according to the Privacy Act Guideline. The model involves translating plain English privacy policy to XML based rule, then using Java based evaluation engine to match the rule with XML queries. If a query violates the rules, the tool will warn the user.

15.30-15.45 *A Two-Pronged Attack on the Dragon of Intractability*

Stephen GILMOUR

Supervisor: Dr Mark DRAS

Associate Supervisor: A/Prof Bernard MANS

Adjunct Supervisor: Dr Yusuf PISAN

Abstract

One approach to tractably finding a solution to an NP-complete optimisation problem is heuristic, where the solution is inexact but quickly found; another approach is to reduce the problem in such a way that the reduction has the same solution as the original but is simpler, and then to solve the reduction, noting that this reduction is still NP-complete. It is possible to combine the two approaches with the goal of taking advantage of both the speed of the heuristic approach and the exactness of the reduction, but this is typically done only in a simple way. The aim of my research is to explore the range of ways in which these two classes of approach can be combined. Part of this talk will include a discussion of work to be presented at ACSC'2005 where we investigate using reduction rules from parameterized complexity to preconfigure the search space for the ant colony optimization heuristic. I will show that under three different models of combination, this approach outperforms standard ACO for the unweighted vertex cover problem, and analyse how each model performs relative to each other.

15.45-16.00 *Learning Structural Characteristics in Machine Translation*

Simon ZWARTS

Supervisor: Dr Mark DRAS

Associate Supervisor: Prof Robert DALE

Abstract

Statistical approaches to machine translation have become popular again in the last decade, with the advent of more processing power and larger bodies of text for statistical training. Most of these statistical approaches use only surface characteristics of the text, but recent work has looked at how structural knowledge can be included in the process. The idea behind these research projects is that machine translation can benefit from knowledge of the rules that can be used to define the grammatical structure of the sentence, usually represented by parse trees. Most machine translation approaches rely heavily on large amounts of parallel text on which they should be trained. However there are very few parallel text available which are annotated with structure, which is necessary for statistical training for such models, in particular for languages outside those most widely spoken. Further, we still cannot rely on the fact that parsers would produce correct parse trees for arbitrary sentences. An additional problem is that most statistical approaches incorporating structure assume a reasonably close parallel structure and do not allow radical transformation. The idea of this Ph.D. project is to use partial information from parsers and elsewhere to address the problem of lack of annotated training data; these partial structures will also allow a wider range of transformation in order to describe translation between more typologically different languages.

16.00-16.15 *World Wide Web Question Answering*

Luiz PIZZATO

Supervisor: Dr Diego MOLLA-ALIOD

Abstract

In my talk, I will give an overview of my PhD research on World Wide Web Question Answering. First, I will describe what Question Answering is and how the web can assist the area. I will briefly show my research plan and explain the phase I am currently working. Finally, I will list this year achievements.

16.15-16.30 Constraint-based Strategies for Robust Semantic Parsing

Jean-Philippe PROST

Supervisor: Dr Diego MOLLA-ALIOD

Associate Supervisor: Dr Mark DRAS

Adjunct Supervisor: Dr Philippe BLACHE

Abstract

Interpreting Natural Language often requires dealing with linguistic information of different quality. The syntactic information can be minimal, like in "Monday washing Tuesday ironing Wednesday resting"; the degree of acceptability can vary as in "(a) Which book has Paul read? (b) Which book has read Paul? (c) Which book Paul has read?". When faced with a deviant construction it is necessary, in order to build an interpretation, to rely on partial information and to take into account information coming from different linguistic dimensions, like syntax, prosody, and pragmatics. Combined with the appropriate linguistic theory, the field of Constraint Satisfaction provides a natural way of representing and dealing at the same level with the multi-dimensional aspect of the linguistic information, but still lacks formal tools for dealing with (the semantics of) partial information. The aim of my research is to investigate the formal properties of a constraint-based computational model with respect to representing and interpreting partial information.

During this talk I will discuss a part of work in progress around the notion of density, which I presented at the Constraint Solving and Language Processing workshop (CSLP 2004). We will see how density, for example, can be used to quantify degree of acceptability, which will provide a way of assessing the accuracy of partial information.

16.30-16.45 A Statistical Account of Coherence and Verisimilitude in Summarization

Stephen WAN

Supervisor: Prof Robert DALE

Associate Supervisor: Dr Mark DRAS

Adjunct Supervisor: Dr Cecile Paris

Abstract

There are many professions today which involve sifting through archives of electronic documents in search of information. Summarisation technology can assist these information professionals with their research by helping to identify more accurately which documents are really worth reading in order to find the information they need. Current summarisation technology simply extracts a list of isolated sentences that are deemed important; however this is far from ideal. In order to better portray completeness of extracted information and to provide an overview of argument structure, a coherent abstract-like summary should be presented to the user. My current research to date has investigated the case of generating summaries for United Nations humanitarian funding proposals. I will argue that such summaries are far from the extraction-based summaries possible with the current state-of-the-art summarisers. The presentation will then describe work in progress which involves a novel hybrid approach that utilises domain independent statistical techniques and robustly acquired symbolic representations of sentences for computing the coherency and veracity of summaries.

~~~~~End of presentations on 9 Nov 2004~~~~~