Session 1 Discrete Optimisation Algorithms

Mark Jones
Max-Cut Parameterized Above the Edwards-Erdos Bound

We study the boundary of tractability for the Max-Cut problem in graphs. Our main result shows that Max-Cut above the Edwards-Erdos bound is fixed-parameter tractable: we give an algorithm that for any connected graph with n vertices and m edges finds a cut of size m/2 + (n-1)/4 + k in time $2^O(k)n^4$, or decides that no such cut exists. This answers a long-standing open question from parameterized complexity that has been posed several times over the past 15 years. Our algorithm is asymptotically optimal, under the Exponential Time Hypothesis, and is strengthened by a polynomial-time computable kernel of polynomial size.

Robert Crowston
Parameterized Study of the Test Cover Problem

Given a set [n], a test cover is a collection {T_1,...,T_m} of distinct subsets of [n], such that for every pair of integers i,j, there exists a T_p, such that |{i,j} \cap T_p|=1. In the Test Cover problem we are given a set [n] of items, together with a collection $\cal T$ of distinct subsets of these items, called tests. We assume this collection form a test cover. The objective of the problem is to find a subset $\cal T' \subseteq \cal T$ such that $\cal T'$ is also a test cover. In the talk I will discuss several natural parametrizations of this problem.

Gabriele Muciaccia
(Non)-existence of Polynomial Kernels for the Test Cover Problem

The input of the Test Cover problem consists of a set V of vertices, and a collection E = fE1; : : : ;Emg of distinct subsets of V , called tests. A test Eq separates a pair vi; vj of vertices if jfvi; vjg \ Eqj = 1: A subcollection T _ E is a test cover if each pair vi; vj of distinct vertices is separated by a test in T . The objective is to _nd a test cover of minimum cardinality, if one exists. This problem is NP-hard.

We consider two parameterizations the Test Cover problem with parameter k: (a) decide whether there is a test cover with at most k tests, (b) decide whether there is a test cover with at most jV j□k tests. Both parameterizations are known to be _xed-parameter tractable. We prove that none have a polynomial size kernel unless NP _ coNP=poly. Our proofs use the cross-composition method recently introduced by Bodlaender et al. (2011) and parametric duality introduced by Chen et al. (2005). The result for the parameterization (a) was an open problem (private communications with Henning Fernau and Jiong Guo, Jan.-Feb. 2012). We also show that the parameterization (a) admits a polynomial size kernel if the size of each test is upper-bounded by a constant.

Languages

Robert Walsh
Exploring Abstract Syntax Tree Generation

Tear-Insert-Fold (TIF) Grammars provide a powerful method of transforming a concrete derivation tree into a more abstract form – referred to as an abstract syntax tree.
In this introductory talk (designed to be accessible even for those without a background in compiler theory) we examine the design of compilers and how we can automatically generate parts of a compiler.

We provide motivation for abstract syntax trees as the basis for defining the semantics of the language, and look at how TIF grammars provide a concise means for specifying the set of abstract syntax trees. We will then briefly look at how an abstract syntax can be automatically constructed from a given concrete syntax.

Session 2 Type Theory

Tao Xue
Dot-types and Their Implementation

Dot-types, as proposed by Pustejovsky and studied by many others, are special data types useful in formal semantics to describe interesting linguistic phenomena such as copredication. We present an implementation of dot-types in the proof assistant Plastic base on their formalization in modern type theories.

Bioinformatics

Sandra Smieszek
Across the Bridges and into the Trees (with Haixuan Yang, Paul Devlin and Alberto Paccanaro)

Circadian clocks are ubiquitous and are found in bacteria, fungi, plants and animals. They phase cellular processes and behavior to specific times of day and anticipate daily diurnal changes providing fitness advantage.The difficulty in elucidating the circadian clock is rooted in the notion that two independent factors contribute to controllability: the systems architecture (which components interact with each other) and the dynamic rules that capture the time dependent interactions between components. The identification of the principal actors (several have been characterized) inside a molecular and phenotypic network or inside a community is important to understand the topology of a complex network. Previously, we developed novel methods for the identification of circadian genes from short time-course microarray data and for the identification of the individual regulatory motifs which aggregate into coherent motif clusters capable of predicting the phase of a clock gene with high fidelity. Such motifs form the backbone of regulatory networks and play central role in defining its global topological organization. Here, we integrate gene expression profiles and protein interaction maps to provide a systematic and global view of combinatorial network modules underlying representative circadian programs. Furthermore we integrate the newly discovered cis regulatory modules into the circadian regulatory networks. This study forms the beginning of analytical framework that should allow one to study the controllability of a complex system like the circadian clock in plants through the combination of driver nodes with their time dependent control reflecting the systems dynamic logic. Such a network will provide a quantitative yet holistic outlook upon a complex modular network of great agronomic importance.

Horacio Caniza Vierci
GoCharts: a new holistic approach for Biological Process discovery in transcriptomic data

Vital functions in a living organism are the result of a number of processes carried out by its cells. These biological processes are carried out through the activation of specific groups of genes.

Microarrays can monitor individual genes in a cell, identifying the expression level of each gene at a given time. Microarray data thus allow us to ask an important question: given a cell at a particular time, what are the biological processes being carried out by its genes? This question can be answered using terms from the Gene Ontology, a well defined manually curated vocabulary, containing terms for each of the biological processes in an organism.

During the first year of my PhD, I have contributed in the development GO-Charts (Gene Ontology Charts) a novel, holistic approach for the discovery of biological processes from microarray data. GO-Charts constitutes a fundamental paradigm shift from the current state of the art methods, as it is the first method to explicitly make use of the Gene Ontology structure

to identify the biological processes. The use of the GO structure also allows us to use multi-dimensional scaling (MDS) on a distance matrix between processes to embed them in a 2D plane. This gives raise to "charts" of GO terms, an intuitive way of representing the result of a microarray experiment, that the biologists can easily navigate.

Session 3 Agent Technology

Paulo Ricca Goncalves
The Open Object

We live in a world that is becoming increasingly populated by computationally-capable objects (such as your college card or your mobile phone) connected in various ways between them (such as wifi or bluetooth). One question that often arises is "can we connect these objects in ways that they can cooperate and help us with what we do in a our daily life?" As a first attempt to answer this question, we present the concept of the Open Object, a lightweight decentralised model for orchestrating the collective behaviour of physical objects with computational capabilities. The work also touches the area of End-User Development, as in the long terms we wish to make it easy for anyone to define ways in which objects co-operate in order to achieve a user defined goal.

Bedour Al-Rayes
Adaptive Decision-Making Model for Negotiation in Dynamic Environment

We present some initial ideas towards developing an adaptive bargaining agent that negotiates with other agents on behalf of its user to allocate resources, such as buying and selling goods or services over the Internet. Although the number of people conducting business transactions over the Internet is constantly increasing, e-commerce market places are normally based on fixed prices (e.g. Amazon books) or mediated via negotiation mechanisms based on auctions (e.g. e-Bay). However, there are many limitations of auctions. They are time consuming while communication is unidirectional and participants cannot adapt their strategies to the behaviour of their negotiation opponent. Therefore, agents that are capable of bargaining may allocate goods faster by communicating in bidirectional way, through offers and counter offers. Until now, there are no existing bargaining agents available on the Internet. Consequently, this suggests the need to conduct more research in this area and study where bargaining agents with adaptive strategies can be of use.

Ulrich Schaechtle
Learning Relations with Hypotheses in First-Order Logic in a Medical Domain

Current systems for medical diagnosis are either stuck in inflexible, symbolic-reasoning approaches characteristic of the expert-systems technology of a decade or two ago or in specialised machine-learning techniques whose solutions are hard to bring into a logical context. We present research in the field of statistical relational learning trying to overcome this problem by combining computational logic with probabilistic graphical models. Our research focuses on learning relations in graphical models under explicit incorporation of domain knowledge. Since the knowledge is represented in a logical form, new relations and structures may give way to insights and rules with regards to how to interpret observations of interest. We briefly discuss how to apply these ideas in medical diagnosis and more specifically on monitoring diabetic patients. The work is supported by the EU funded Commodity12 project.

Session 4 Machine Learning

Antonis Lambrou
Reliable probability estimates based on Support Vector Machines for large multiclass datasets

Venn Predictors are machine learning algorithms that can provide calibrated multiprobability outputs for their predictions. Each Venn Predictor is defined by a taxonomy which separates the data into categories. Typically, a taxonomy is derived from the use of an underlying classical algorithm. In this paper, we apply the Venn Prediction framework on a simple taxonomy derived from the multiclass Support Vector Machine (SVM). We compare our method with other probabilistic methods for SVMs, namely Platt's method, SVM Binning, and SVM with Isotonic Regression. We show that these methods do not always provide well calibrated results, while Venn Predictors will always guarantee this property under the i.i.d. assumption. We also use the inductive framework on Venn Predictors in order to be able to compete with the time efficiency of the other algorithms on large datasets.

Dmitry Adamskiy
A Closer Look at Adaptive Regret

For the prediction with expert advice setting, we consider methods to construct algorithms that have low adaptive regret. The adaptive regret of an algorithm on a time interval [t1, t2] is the loss of the algorithm there minus the loss of the best expert. Adaptive regret measures how well the algorithm approximates the best expert locally, and it is therefore somewhere between the classical regret (measured on all outcomes) and the tracking regret, where the algorithm is compared to a good sequence of experts.

Valentina Fedorova
Plug-in martingales for testing exchangeability on-line

Real-life data are complex and a challenge is to check their properties. Machine learning algorithms have been developed to learn and make prediction for such data. To state and prove properties of the algorithms usually it is necessary to assume that the examples of data are drawn independently from a probability distribution (the same for all the examples), that is equivalent to the exchangeability assumption. Thus it is important to check that the assumption is satisfied for the data otherwise the properties of algorithms can be violated and their results can be misleading. This talk is about an on-line approach for testing exchangeability. The on-line mode means that examples of data arrive one after another and after observing every example the testing procedure outputs a number reflecting a deviation from the assumption. It is done by calculating exchangeability martingales. Martingales are special random variables with a property that if they grow up to a large value we can reject our assumptions. I present the last results about a new way of construction martingales. The new martingales are flexible to detect different deviations from the exchangeability. And it has been proven that asymptotically they provide a better growth rate than other known martingales. The talk outlines the background theory, explains the new results and shows experiments for two benchmark datasets.

Jiaxin Kou
Information Visualisation based on t-SNE

The aim of our research is to develop new visualization methods as a practical tool in machine learning and data mining, which can be used to show the structure of high-dimensional data and the fit of a machine learning algorithm to a dataset. On the existing research of t-SNE and NNP, recently we develop a tool to intestate the performance and fitness of Support Vector Machine to various of datasets, which is just a glimpse of new machine learning experience.

Meng Yang
Learning with additional information

In many supervised learning applications, training data contains additional information not reflected in training pairs. Vapnik introduced a new paradigm LUPI which provides the learning paradigm with additional information and take advantage of it through SVM+ method in off-line mode. Following his idea, we would like to use additional information by conformal predictor in on-line mode, which allows to make prediction with reliable measures of confidence, to improve the prediction performance.

Chenzhe Zhou
Implementing Multi-Classification SVM Algorithms into Venn Machine

An introduction to the implementation of multi-classification support vector machine algorithms into venn machine and some comparisons between the results.

Tim Scarfe
A Dynamic Programming Approach to Segmenting DJ Mixed Music Shows

DJ mixes are an essential part of electronic dance music. Whether you listen to your favourite DJ on the radio, live or downloaded from the internet -- the individual tracks will be mixed together. More recently, podcasts (through Apple iTunes) have become the de facto standard empowering people to download radio shows for offline listening automatically. Users can play the mixes on their computers and/or mobile devices. Many shows are also broadcast on internet radio stations such as Digitally Imported and syndicated world wide on terrestrial radio.

The technique we used in this paper is to segment the audio file into discrete time intervals (called tiles), and perform a transformation on those tiles into a domain where two tiles from the same tracks would be distinguishable by their cosine. We extracted two core sets of features; Fourier based and autocorrelation based. Intuitively speaking the Fourier features can be thought of as useful for distinguishing instruments (after filtering to remove noise and accentuate relevant features) and the autocorrelation features useful for distinguishing rhythmic patterns. A dynamic programming approach is used to segment the shows back into their original tracks.