



Technical University of Denmark
1st – 4th of July 2009



Workshop Program



Sponsors



Technical University of Denmark

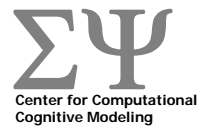


Table of Contents

| | |
|---------------------------------------|----|
| The Program at a Glance | 3 |
| Practical Information..... | 4 |
| Oral Presentations | 7 |
| Wednesday the 1 st of July | 7 |
| Thursday the 2 nd of July | 10 |
| Friday the 3 rd of July | 13 |
| Saturday the 4 th of July | 17 |
| Posters | 21 |
| List of Participants | 31 |

Organizers

Morten Mørup, Technical University of Denmark
Lek-Heng Lim, University of Berkeley, California
Michael Mahoney, Stanford University
Lars Kai Hansen, Technical University of Denmark
Gunnar Carlsson, Stanford University

Secretary

Ulla Nørhave, Technical University of Denmark

Contact:

mmds-organizers@imm.dtu.dk

Homepage:

<http://mmds.imm.dtu.dk>

The Program at a Glance

Wednesday, July 1

Morning Session

09:00 – 09:10 Opening: Kaj Madsen

09:10 – 10:10 Jerome Friedman

10:10 – 10:30 **Coffee Break**

Theme: Statistical Learning

10:30 – 11:30 Steffen Lauritzen

11:30 – 12:05 Yee Whye Teh

12:05 – 13:30 **Lunch Break**

Afternoon Session

13:30-14:30 Bernhard Schölkopf

Theme: Machine Learning

14:30 – 15:05 Klaus Mosegaard

15:05 – 15:25 **Coffee Break**

15:25 – 16:00 Nello Cristianini

16:00 – 16:35 Mikkel Schmidt

16:35 – 17:10 Ole Winther

17:20 **Welcome Drink**

Thursday, July 2

Morning Session

09:00 – 10:00 Edward Chang

10:00 – 10:20 **Coffee Break**

Theme: Multilinear Algebra for Data Analysis

10:20 – 10:55 Rasmus Bro

10:55 – 11:30 Pierre Comon

11:30 – 12:05 Lieven De Lathauwer

12:05 – 12:40 Lars Eldén

12:40 – 14:00 **Lunch Break**

Afternoon Session

14:00 – 15:00 Tomaso Poggio

15:00 – 17:45 **Poster Session**

18:00 **Bus from DTU to restaurant Ofelia**

Friday, July 3

Morning Session,

09:00 – 10:00 Ricardo Baeza-Yates

10:00 – 10:20 **Coffee Break**

Theme: Neuroscience

10:20 – 11:20 Scott Makeig

11:25 – 12:00 John Ashburner

12:00 – 13:30 **Lunch Break**

Afternoon Session

13:30 – 14:30 Joachim Buhmann

Theme: Clustering

14:30 – 15:05 Charles Elkan

15:05 – 15:25 **Coffee Break**

15:25 – 16:00 Neil Lawrence

16:00 – 16:35 Michael Mahoney

16:35 – 17:10 Morten Mørup

Saturday, July 4

Morning Session

09:00-10:00 Gunnar Carlsson

10:00 – 10:20 **Coffee Break**

Theme: New Mathematical Tools for Data Analysis

10:20 – 10:55 Risi Kondor

10:55 – 11:30 Samuel Kaski

11:30 – 12:05 Lek-Heng Lim

12:05 – 13:30 **Lunch Break**

Afternoon Session

13:30 – 14:30 Pedro Cano

Theme: Social computing

14:30 – 15:05 Joaquin Quiñero Candela

15:05 – 15:25 **Coffee Break**

15:25 – 16:00 Mark Herbster

16:00 – 16:35 Sune Lehmann

16:35 – 17:10 Lars Kai Hansen

17:30 **Bus from DTU to Cafe Ultimo in Tivoli**

Practical Information

Conference Venue: Glassalen at the Technical University of Denmark (DTU), Building 101

Address: Anker Engelunds Vej 1
DK-2800 Kongens Lyngby
Tel: (DTU) (+45) 45 25 25 25
Tel: (EMMDS) (+45) 60 81 31 78

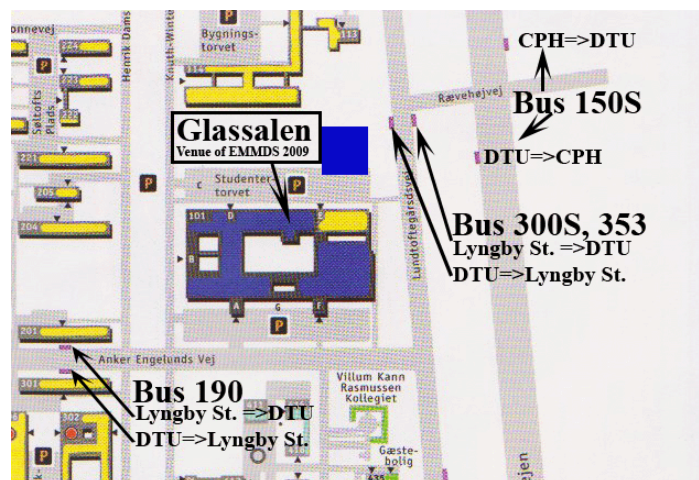


Getting from Copenhagen to the Technical University of Denmark:

The easiest is to take bus 150S from Nørreport station in direction Kokkedal station and get off at DTU/Rævehøjvej. The trip takes 25 minutes.

Alternatively from Copenhagen Central Station (København H.) or Nørreport Station take train B direction Holte og E direction Hillerød and get off at Lyngby station. From Lyngby station take either bus 353 direction Helsingør Station or 300S direction Nærum Station and get off at DTU/Rævehøjvej. Alternatively, take bus 190 Direction Holte Station and get off at Anker Engelundsvej.

The above trips will take approximately 30 minutes and require 5 zones, we suggest you purchase a 5 zone "clip card" valid for 10 trips (price 265 DKK) which can be obtained at any train station or kiosk around Copenhagen. Alternatively, one way tickets can be bought at the train station or on the bus (the bus only accepts cash) for 52.50 DKK.



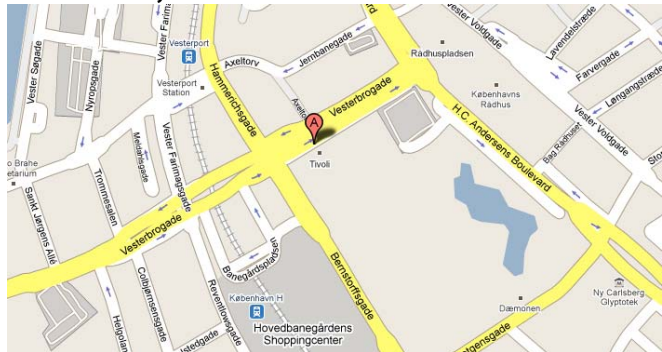
Dinner 2nd of July 18:45: Restaurant Ofelia at Skuespilhuset,

**Address: Sankt Annæ Plads 36
1250 København K
Tlf. (+45) 3369 3931**



Dinner 4th of July 18:30: Café Ultimo in Tivoli,

**Address: Vesterbrogade 3
1630 København
Tlf. (+45) 3375 0751**



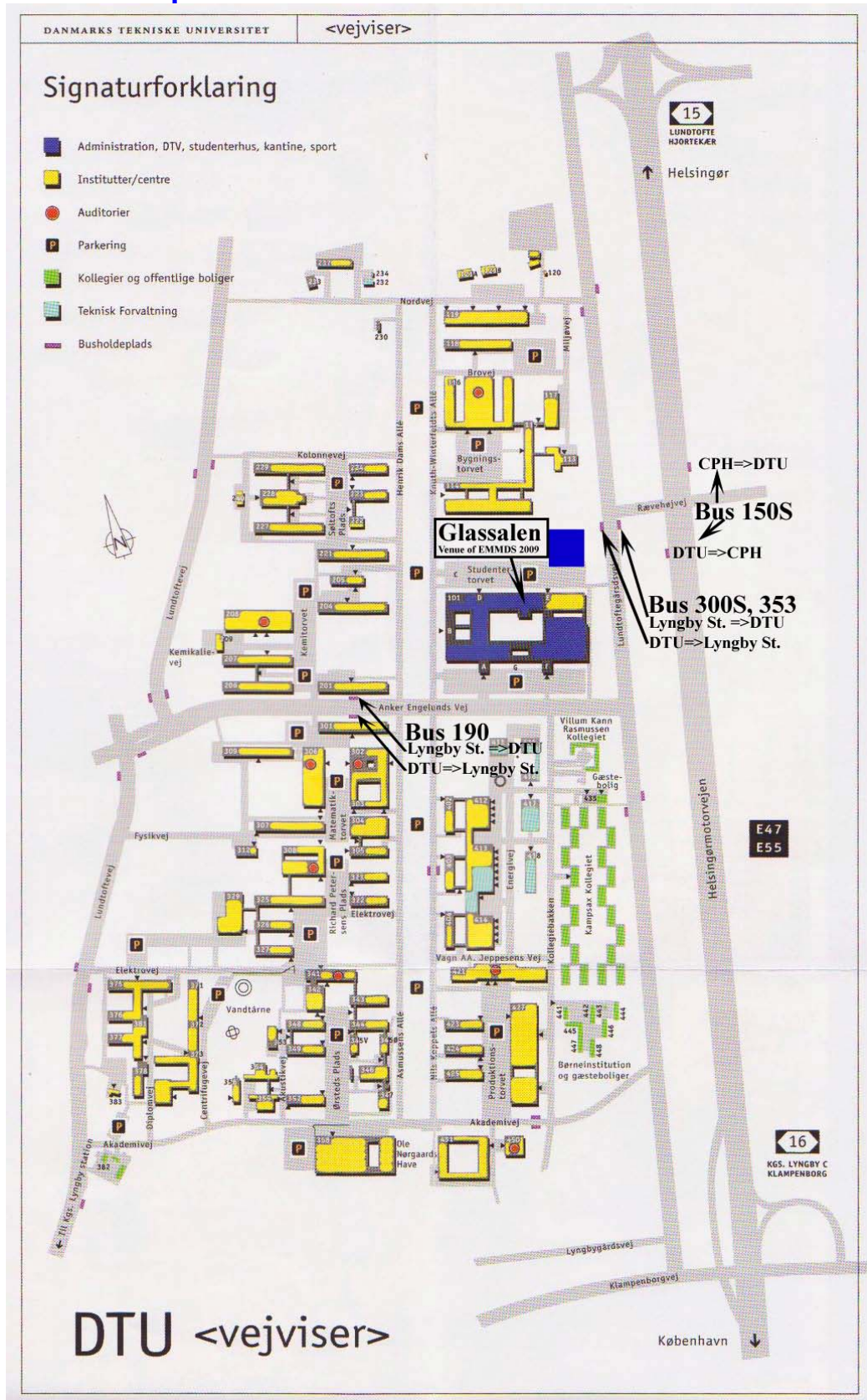
Internet Access

There will be free internet access at the conference venue. If you want to use the internet contact the registration desk to receive a username and password.

Taxa: To get a taxa you can contact any of the following services:

| | |
|-------------------------|--------------------------|
| DanTaxi | (+45) 38 79 08 50 |
| Amager-Øbro Taxa | (+45) 32 51 51 51 |
| Codan taxi | (+45) 70 25 25 25 |
| Taxamotor | (+45) 70 33 83 38 |
| TaxiNord | (+45) 48 48 48 48 |
| Taxa 4x35 | (+45) 35 35 35 35 |

Map of the DTU campus



Oral Presentations

Wednesday the 1st of July

Morning Session

09:00 – 09:10 Opening: Kaj Madsen, Technical University of Denmark

09:10 – 10:10 Jerome Friedman, Stanford University

Title: Predictive Learning via Rule Ensembles

Abstract: General regression and classification models are constructed as linear combinations of simple rules derived from the data. These rule ensembles are shown to produce predictive accuracy comparable to the best methods. However their principal advantage lies in interpretation. Because of its simple form, each rule is easy to understand, as is its influence on individual predictions. Similarly, the relevances of the respective input variables can be assessed globally, locally in different regions of the input space, or at individual prediction points. Techniques are presented for automatically identifying those variables that are involved in interactions with other variables, the strength and degree of those interactions, as well as the identities of the other variables with which they interact. Graphical representations are used to visualize both main and interaction effects.

* Joint work with Bogdan Popescu

10:10 – 10:30 Coffee Break

Theme: Statistical Learning

10:30 – 11:30 Steffen Lauritzen, University of Oxford

Title: Estimation of sparse conditional independence structures

Abstract: Conditional independence structures as represented by graphical models have a potential for revealing important relations between many variables. Methods for structure estimation and identification of different type have been suggested and studied empirically and to some extent theoretically. The lecture will briefly survey some of the more successful methods and discuss in what sense their potential is not yet fully exploited.

11:30 – 12:05 Yee Whye Teh, University College London

Title: Bayesian nonparametrics in document and language modeling

Abstract: Bayesian nonparametric models have garnered significant attention in recent years in both the machine learning and statistics communities. These are highly flexible models whose complexity grows with the amount of data, and are nice approaches to addressing the common problem of model selection. In this talk I shall first give a brief overview of Dirichlet processes and infinite mixture models, the cornerstone of Bayesian nonparametric models. Then I shall introduce the hierarchical Dirichlet process, a Bayesian nonparametric model for problems involving multiple related groups of data. I illustrate the use of hierarchical Dirichlet processes using some applications to document and language modelling.

12:05 – 13:30 Lunch Break

Afternoon Session

13:30-14:30 Bernhard Schölkopf, Max Planck Institute

Title: Machine learning with positive definite kernels

Abstract: Kernel methods have become one of the most widely used techniques in the field of machine learning. I will present my thoughts on what made them popular and what may (or may not) keep them going. I will discuss some recent algorithmic developments as well as applications in different domains.

Theme: Machine Learning

14:30 – 15:05 Klaus Mosegaard, University of Copenhagen

Title: Metaheuristics in science and engineering

Abstract: Metaheuristics play a major role in modeling algorithms for analysis of nonlinear, high-dimensional data sets in engineering and science. They are used when there is no satisfactory problem-specific algorithms available, or when it is not practical to implement such methods. Metaheuristics are considered to be problem-independent and to derive their efficiency from general ideas, such as genetic adaptation, thermodynamic relaxation, and other common-sense strategies.

Indeed, many metaheuristics perform well on certain problems, but we claim that this is only because they are modified to become problem specific (and hence not metaheuristics any more!). Modeling algorithms ignoring special properties of the problem to be solved will perform poorly. The more problem-specific they are, the more efficient.

The most common way to modify and improve the performance of a metaheuristic is to use information about the smoothness of the objective function of the considered problem. Although this improves the efficiency of the algorithm, we claim that most nonlinear, high-dimensional modeling problems remain hard even after this improvement.

15:05 – 15:25 Coffee Break

15:25 – 16:00 Nello Cristianini, University of Bristol

Title: Looking for Memes in Media Content

Abstract: The contents of the media sphere present various patterns that are of interest to social scientists. The extraction of this information by computational methods leads to large scale pattern-analysis questions. We will discuss efficient methods to detect surprising surprising memes in a textual data stream.

16:00 – 16:35 Mikkel Schmidt, University of Cambridge

Title: Bayesian matrix factorization approaches to blind source separation

Abstract: The blind source separation problem is to infer a set of hidden sources when only their mixtures are observable. This problem can naturally be represented as a matrix factorization problem. We present a general Bayesian approach to probabilistic matrix factorization, and present an efficient Markov chain Monte Carlo inference procedure based on Gibbs sampling. We discuss how adding relevant linear constraints can completely change the results of the algorithm. We demonstrate that our algorithm can be used to extract meaningful and interpretable features that are remarkably different from features extracted using existing related matrix factorization techniques.

16:35 – 17:10 Ole Winther, Technical University of Denmark / University of Copenhagen

Title: Hierarchical Bayesian modelling for collaborative filtering

Abstract: The Netflix prize has generated a lot of interest for collaborative filtering in machine learning. Collaborative filtering is a good example of the large scale learning problems of web era. Perhaps surprisingly, many machine learning methods scale well to giga-byte data sets. In this talk I will present joint work with Ulrich Paquet and Blaise Thomson, Cambridge on a Bayesian hierarchical approach to the Netflix problem. The model combines an ordinal likelihood for the discrete ranks with movie-viewer low-rank matrix factorization. Inference in this model, which has between 25 to 100 millions parameters (depending upon the rank of the matrix factorization), is carried out with Gibbs sampling. This confirms previous work (Salakhutdinov and Mnih, ICML-2008) that Bayesian averaging is viable for large models and data sets. I will discuss how parallelization can be used to further scale up the problem. Our best model achieves a 6.32% improvement in terms of RMSE over the Netflix's own CineMatch system. This is to our knowledge the best single model result, i.e. with no use of blending of different algorithms.

17:20 Welcome Drink

Thursday the 2nd of July

Morning Session

09:00 – 10:00 Edward Chang, Google Research, Beijing

Title: Parallel Algorithms for Collaborative Filtering

Abstract: Collaborative filtering has been widely used to predict the interests of a user. Given a user's past activities, collaborative filtering predicts the user's future preferences. This talk presents techniques and discoveries of our recent parallelization effort on collaborative filtering algorithms. In particular, parallel association mining and parallel latent Dirichlet allocation will be presented and their pros and cons analyzed. Some counter-intuitive results will also be presented to stimulate future parallel optimization research.

10:00 – 10:20 Coffee Break

Theme: Multilinear Algebra for Data Analysis

10:20 – 10:55 Rasmus Bro, University of Copenhagen

Title: Applications of Tensor Methods in Life Sciences

Abstract: In the last 30 years tensor methods have been increasingly used in chemistry and related areas, but the methods have only recently been more widely used. In this talk, some of the most used tensor methods are being discussed with a focus on some or the less known methods such as tensor regression and various constrained models. Examples are provided from different areas of chemistry and life science.

10:55 – 11:30 Pierre Comon, University of Nice, Sophia-Antipolis

Title: Tensor decompositions in statistical signal processing

Abstract: Tensor decompositions show up under different formulations in numerous application areas, including arithmetic complexity, data analysis and inverse problems. Depending on the application, the decompositions must be exact, or may be approximate. The problems of existence and uniqueness are often of crucial importance, but are yet incompletely solved, even if necessary conditions exist to obtain uniqueness up to scale and permutation ambiguities. Moreover in most cases, efficient numerical algorithms are still lacking. The focus of the talk is mainly on tensors enjoying symmetries, encountered in various problems of statistical signal processing. Independent Component Analysis may be seen as the approximation of a cumulant tensor by another of rank equal to its dimension. Blind Identification of linear mixtures of independent random variables may be seen as the exact decomposition of a tensor into a sum of rank-1 terms. Several other problems are shown to be of the same form as the two latter decompositions.

11:30 – 12:05 Lieven De Lathauwer, Katholieke Universiteit Leuven

Title: Tucker Compression, Parallel Factor Analysis and Block Term Decompositions: New Results

Abstract: The two most well-known generalizations of matrix rank to higher-order tensors are multilinear rank and (outer product) rank, respectively. The computation of the best low multilinear rank approximation of a tensor is sometimes called Tucker compression. The two main applications are compression of multi-way data and the computation of dominant subspaces of a higher-order tensor. In this talk, we touch on some new algorithms and discuss the problem of local minima. The (outer product) rank of a higher-order tensor is the minimal number of rank-1 terms in which it can be decomposed. The decomposition itself is sometimes called the Canonical or Parallel Factor decomposition (CANDECOMP/PARAFAC)(CP). The most well-known condition under which essential uniqueness of the CP decomposition is guaranteed, is the famous Kruskal bound. The rank is usually determined by means of trial-and-error or by using heuristic criteria. The most popular algorithm is of the alternating least squares type. However, in many practical cases of interest, the rank is equal to the matrix rank of a matrix representation of the tensor. In the case of exact data, CP may be computed by means of standard linear algebra under certain conditions on the rank. In the case of noisy data, one can often use algorithms for simultaneous matrix diagonalization. There also exists a uniqueness condition that is significantly more relaxed than Kruskal's, provided one of the tensor dimensions is greater than its rank. The latter condition is satisfied in many applications. Recently, we have proposed Block Term Decompositions (BTD) as a unifying framework for CP and Tucker. BTD also unifies multilinear rank and (outer product) rank. We discuss BTD generalizations of our CP results.

12:05 – 12:40 Lars Eldén, Linköping University

Title: Krylov methods for tensors

Abstract: Sparse tensors occur frequently in information sciences. We investigate different approaches of generalizing Krylov methods to tensors. We also discuss the applicability of Krylov-Schur methods. The purpose is to compute low-rank Tucker models of huge tensors.

12:40 – 14:00 Lunch Break

Afternoon Session

14:00 – 15:00 Tomaso Poggio, Massachusetts Institute of Technology

Title: From Neuroscience to Hierarchical Learning Architectures

Abstract: In learning theory an effective and broad class of algorithms is provided by regularization in Reproducing Kernel Hilbert Spaces. After a brief overview of their mathematical structure, I will list additional “learning principles” that could be used in addition to RKHS-like smoothness to capture prior information in the data in addressing the problem of “sample complexity”. One of these new “learning principles” is reflected in the hierarchical architecture of a learning algorithm. I will then sketch a new attempt (with S. Smale, L. Rosasco and J. Bouvrie) to develop a mathematics for hierarchical kernel machines – centered around the notion of a recursively defined “derived kernel” – and directly suggested by the neuroscience of the visual cortex.

15:00 – 17:45 Poster Session

18:00 Bus from DTU to restaurant Ofelia

Friday the 3rd of July

Morning Session,

09:00 – 10:00 Ricardo Baeza-Yates, Yahoo! Research, Barcelona

Title: The Power of Data

Abstract: In this presentation we survey several examples of mining large Web related data sets, including queries, text, images, etc. We will survey different ways to use and analyze this data to improve different applications and to evaluate the results. As a corollary we show that many times the quality and volume of data is more important than algorithmic improvements.

10:00 – 10:20 Coffee Break

Theme: Neuroscience

10:20 – 11:20 Scott Makeig, University of California, San Diego

Title: Multiscale brain/body imaging: Towards a single brain electrophysiology

Abstract: Human brain dynamics are inherently multiscale, their structures at different spatial scales (e.g., cortical regions, columns, neurons, synapses, molecules) having different dynamics and connectivity structure. One of the most pressing current basic research questions in brain dynamics concern the active coordination of these dynamics at multiple space and time scales, a subject long neglected during the experimental era dominated by single microelectrode data recording and analysis.

Scalp EEG (or its near-equivalent MEG) can only see the far-field projections of locally synchronized field activity, chiefly across cortical domains of poorly understood size, location, and dynamics. Ideal measures of cortical field dynamics therefore should be multi-resolution. A unique window of opportunity is afforded by the current clinical practice of invasive monitoring of cortical (and/or sub-cortical) activities in subjects with complex cases of intractable epilepsy for the purpose of planning remedial brain surgery. I and my colleagues have begun analyzing human multi-resolution electrophysiological data recorded by Dr. Greg Worrell at Mayo Clinic. Though adequate joint analysis and modeling of simultaneous recordings at multiple spatial scales is not simple, it will be necessary to understand electrophysiological data recorded at any scale.

However, high-density imaging modalities that view brain activity from the greatest distance (e.g., from the scalp) may paradoxically also be better suited to studying its distributed dynamics. Our research into new methods for human EEG analysis suggests that high-density EEG will soon become a true functional cortical imaging modality with high temporal and adequate spatial resolution. EEG is also the only lightweight, low cost, low energy, and therefore portable brain imaging modality. We are therefore pioneering the use of simultaneous high-density portable EEG, body motion capture, and eye tracking to image brain dynamics supporting our active behavior in the natural 3-D world, a modality I call Mobile Brain/Body Imaging (MoBI).

At least since the advent of response averaging computers in the early 1960s, electrophysiological research at the neuron and scalp scales has been separated into two camps (neurobiology and psychophysiology) with little to say to each other. More adequate analysis methods, applied to model the complexities of unaveraged data at all spatial scales, should

produce a reunification of brain electrophysiology as a single topic. In this process, public data resources including the new Human Electrophysiology and Anatomic Data with Integrated Tools (HeadIT) resource that Jeff Grethe and I have been funded to build may play a useful role.

11:25 – 12:00 John Ashburner, University College London

Title: Brain Morphometrics from MRI Scans

Abstract: This talk is about procedures for modelling the anatomical variability among human brains, using MRI scans. Currently within the brain imaging field, a number of modellers make their algorithms freely available for use by neuroscientists, neurologists etc, who then apply these models to their own relatively small datasets. Most neuroscientists wish to localise anatomical differences between populations, identifying regions of the brain that may be larger or smaller in one population, compared to another. Such simple mass-univariate characterisations can be compactly described, and easily communicated in the literature. Unfortunately, the models used to achieve such characterisations have poor predictive accuracy.

More recently, a trend may be emerging, whereby neuroscientists, neurologists etc are more prepared to share their data. Such datasets can be pooled, which should allow a much deeper understanding of neuroanatomical variability to be achieved. Projects such as ADNI, OASIS, IXI etc are making scans of human brains freely available via the web, and multivariate pattern recognition approaches are being applied to such data, where the aim is to develop biomarkers or diagnostic tools. This is especially the situation for Alzheimer's Disease, which is becoming more prevalent as the population ages. Ideas developed by Grenander, Miller, Younes, Mumford and others can be used to model neuroanatomical variability, and may provide more parsimonious representations of the data. These could allow greater predictive accuracies to be achieved.

12:00 – 13:30 Lunch Break

Afternoon Session

13:30 –14:30 Joachim Buhmann, Swiss Federal Institute of Technology (ETH), Zürich

Title: Structure validation in clustering by stability analysis

Abstract: Partitioning of data sets into groups defines an important preprocessing step for compression, prototype extraction or outlier removal. Various criteria of connectedness or proximity have been proposed to group data according to structural similarity but in general it is unclear which method or model to use. In the spirit of information theory we propose a decision process to determine the extractable information from data conditioned on a hypothesis class of structures. Maximizing the amount of information which can be reliably learned from data in the presence of noise selects appropriate models. Empirical evidence for this model selection concept is provided by cluster validation in bioinformatics and in computer security, i.e., the analysis of microarray data and multilabel clustering of Boolean data for role based access control.

Theme: Clustering

14:30 – 15:05 Charles Elkan, University of California, San Diego

Title: Accounting for burstiness in topic models

Abstract: A topic model is a statistical model of a collection of documents that identifies the multiple themes that are present in the documents. Topic modeling is more useful than clustering, because clustering assumes that each document belongs to a single theme, which is often not true in reality. All current topic models assume that each theme is represented by a multinomial distribution, an assumption that is also not true in reality, since it contradicts the fact that words in topics and in documents are bursty: if a word is used once, it is more likely to be used again. In this talk, I will present a new topic model that uses the Dirichlet compound multinomial (DCM) distribution to allow for burstiness. Experimental results show that the DCM-based model fits real-world data much better than the corresponding traditional topic model. I will also describe a new extended topic model that lets us learn correlations between topics using a generalization of the DCM. Note:
Joint work with Gabe Doyle from Linguistics at UCSD.

15:05 – 15:25 Coffee Break

15:25 – 16:00 Neil Lawrence, University of Manchester

Title: Nonlinear Matrix Factorization with Gaussian Processes

Abstract: A popular approach to collaborative filtering is matrix factorization. In this talk we consider the "probabilistic matrix factorization" and by taking a latent variable model perspective we show its equivalence to Bayesian PCA. This inspires us to consider probabilistic PCA and its non-linear extension, the Gaussian process latent variable model (GP-LVM) as an approach for probabilistic non-linear matrix factorization. We apply our approach to benchmark movie recommender data sets. The results show better than previous state-of-the-art performance.

This is joint work with Raquel Urtasun.

16:00 – 16:35 Michael Mahoney, Stanford University

Title: Community structure in large social and information networks

Abstract: The concept of a community is central to social network analysis, and thus a large body of work has been devoted to identifying community structure. For example, a community may be thought of as a set of web pages on related topics, a set of people who share common interests, or more generally as a set of nodes in a network more similar amongst themselves than with the remainder of the network. Motivated by difficulties we experienced at actually finding meaningful communities in large real-world networks, we have performed a large scale analysis of a wide range of social and information networks. Our main methodology uses local spectral methods and involves computing isoperimetric properties of the networks at various size scales -- a novel application of ideas from statistics and scientific computation to internet data analysis. Our empirical results suggest a significantly more refined picture of community structure than has been appreciated previously. Our most striking finding is that in nearly every network dataset we examined, we observe tight but almost trivial communities at very small size scales, and at larger size scales, the best possible communities gradually "blend in" with the rest of the network and thus become less "community-like." This behavior is not explained, even at a qualitative level, by any of the commonly-used network generation models. Moreover, this behavior is exactly the opposite of what one would expect based on experience with and intuition from expander graphs, from graphs that are well-embeddable in a low-dimensional structure, and from small social networks that have served as testbeds of community detection algorithms. Possible mechanisms for reproducing our empirical observations will be discussed, as will implications of these findings for clustering, classification, and more general data analysis in modern large social and information networks.

16:35 – 17:10 Morten Mørup, Technical University of Denmark

Title: Clustering on the Simplex

Abstract: Continuous relaxation of hard assignment clustering problems can lead to better solutions than greedy iterative refinement algorithms. However, the validity of existing relaxations is contingent on problem specific fuzzy parameters that quantify the level of similarity between the original combinatorial problem and the relaxed continuous domain problem. Equivalence of solutions obtained from the relaxation and the hard assignment is guaranteed only in the limit of vanishing "fuzziness". This paper derives a new exact relaxation without such a fuzzy parameter which is applicable for a wide range of clustering problems such as the k-means objective and pairwise clustering as well as graph partition problems, e.g., for community detection in complex networks. In particular we show that a relaxation to the simplex can be given for which the extreme solutions are stable hard assignment solutions and vice versa. Based on the new relaxation we derive the sr-clustering algorithm that has the same complexity as traditional greedy iterative refinement algorithms but leading to significantly better partitions of the data. We finally propose a new type of clustering methods based on extracting the principal convex hull formed by imposing two types of simplex constraints. We demonstrate how the features extracted are defined by the distinct aspects of the data.

Saturday the 4th of July

Morning Session

09:00-10:00 Gunnar Carlsson, Stanford University

Title: Topology and Data

Abstract: Topological methods have in recent years shown themselves to be capable of identifying a number of qualitative geometric properties of high dimensional data sets. In this talk, we will describe philosophical reasons why topological methods should play a role in the study of data, as well as give several examples of how the ideas play out in practice.

10:00 – 10:20 Coffee Break

Theme: New Mathematical Tools for Data Analysis

10:20 – 10:55 Risi Kondor, University College London / California Institute of Technology

Title: Non-commutative harmonic analysis in machine learning: the skew spectrum and the graphlet spectrum of graphs

Abstract: There has recently been a surge of interest in applying representation theoretical tools to machine learning problems. I give a brief overview of these developments and describe a specific application, which is the construction of efficiently computable graph invariants for classifying medium sized graphs, such as the chemical structure of organic molecules. Joint work with Karsten Borgwardt and Nino Shervashidze

10:55 – 11:30 Samuel Kaski, Helsinki University of Technology

Title: Probabilistic retrieval and visualization of relevant experiments

Abstract: Given a databank of data-intensive scientific experiments, a useful exercise is to find earlier experiments relevant to a new study. Our application is in modern cellular biology, where databanks of genome-wide measurement data from earlier experiments exist, and data content-based search is likely to produce novel insights. The goal is to retrieve experiments in which the same biological processes are activated. This can be due either to experiments targeting the same biological question, or to as-yet unknown relationships. We use a combination of existing and new probabilistic machine learning techniques to extract information about the biological processes differentially activated in each experiment, to retrieve earlier experiments where the same processes are activated, and to visualize and interpret the retrieval results.

This is joint work with José Caldas, Nils Gehlenborg, Ali Faisal, and Alvis Brazma.

11:30 – 12:05 Lek-Heng Lim, University of California, Berkeley

Title: Principal Cumulant Components Analysis

Abstract: Multivariate Gaussian data is completely characterized by its mean and covariance, yet modern non-Gaussian data makes higher-order statistics such as cumulants inevitable. For univariate data, the third and fourth scalar-valued cumulants are relatively well-studied as skewness and kurtosis. For multivariate data, these cumulants are tensor-valued, higher-order analogs of the covariance matrix capturing higher-order dependence in the data. In addition to their relative obscurity, there are few effective methods for analyzing these cumulant tensors. We propose a technique along the lines of Principal Component Analysis and Independent Component Analysis to analyze multivariate, non-Gaussian data motivated by the multilinear algebraic properties of cumulants. Our novel subspace selection method relies on finding principal cumulant components that account for most of the variation in all higher-order cumulants, just as PCA obtains varimax components. An efficient algorithm based on limited-memory quasi-Newton maximization over a Grassmannian, using only standard matrix operations, may be used to find the principal cumulant components. Numerical experiments include forecasting higher portfolio moments and image dimension reduction.

This is joint work with Jason Morton.

12:05 – 13:30 Lunch Break

Afternoon Session

13:30 – 14:30 Pedro Cano, Barcelona Music and Audio Technologies

Title: Music Recommendation Systems: A Complex Networks Perspective

Abstract: Nowadays it is possible to access millions of music tracks. In order to ease users to search and discover the music, a number of recommendation systems have been proposed. We analyze the topology of several commercial recommendation systems from a Complex networks perspective. We observe structural properties that provide some hints on searchability and behaviour of music recommendation systems. For example, is the Long Tail truly exploited? How much of the network structure is due to actual musical similarity and how much to other network growth processes?. The properties derived allows us to compare different recommendation approaches: Expert-based, Collaborative filtering and Content-based. Finally, we provide possible optimizations when designing music information systems.

Theme: Social Computing

14:30 – 15:05 Joaquin Quiñonero Candela, Microsoft Research, Cambridge

Title: Probabilistic Machine Learning in Computational Advertising

Abstract: In the past few years online advertising has grown at least an order of magnitude faster than advertising on all other media. This talk focuses on advertising on search engines, where accurate predictions of the probability that a user clicks on an advertisement crucially benefit all three parties involved: the user, the advertiser, and the search engine. We present a fully probabilistic classification model that has the ability of learning from terabytes of web usage data. The model explicitly represents uncertainty allowing for fully probabilistic predictions. Observing 2 positives out of 10 instances or 200 positives out of 1000 instances leads in both cases to an average of 20%, but in the first case the uncertainty about the prediction should be larger. The only way of learning about an advertisement is to show it to the user, which often comes at an opportunity cost since this might not be the best ad to show. The talk describes the way in which the model presented allows for a natural way of exploring the available ad inventory.

15:05 – 15:25 Coffee Break

15:25 – 16:00 Mark Herbster, University College London

Title: Resistive geometry for graph-based transduction.

Abstract: We give an overview of well-known methods for graph-based transduction. In graph-based transduction we are given a fixed set of objects, some of which are labeled and some of which are unlabeled, and we wish to predict the unlabeled objects. A graph is then defined where an edge between objects indicates similarity between objects. If the graph is weighted then the weights indicate the degree of similarity. These include the min-cut method of [2] and the harmonic energy minimization procedure of [5] (also [1]). We interpret these methods as specific instances of the minimization of a p-energy [3]. When $p = 2$ the analogy is that the graph is an electrical network; the edges are now resistors whose resistance is reciprocal of the similarity. The fixed labels from $\{-1, 1\}$ now correspond to potential (voltage) constraints and the algorithm for labeling the graph is then to find the set of consistent voltages which minimize the energy dissipation and then to predict with the "sign" of the voltages. We extend the analogy for general p , which leads to natural analogues of Kirchoffs laws, Ohms law, the conservation of energy principle, and the rules of resistors in series and parallel.

We exploit this network analogy in two ways. First, we will show how by choosing a $p \in (1, 2)$ this leads to an algorithm that obtains advantageous performance guarantees that are unobtainable for the special cases of $p = 1, 2$. Second, when $p = 2$ we will show how to treat this energy minimization procedure as a kernel method, and we will find that the kernel matrix can be derived by a transformation of the matrix of effective resistances between pairs of vertices in the network. By approximating the network with a tree, we will develop a fast method [4] to compute the kernel matrix. This allows us to scale our method to large graphs. We present experiments on two web-spam classification tasks, one of which includes a graph with 400,000 vertices and more than 10,000,000 edges. The results indicate that the accuracy of our technique is competitive with previous methods using the full graph information.

References

- [1] M. Belkin, I. Matveeva, and P. Niyogi. Regularization and semi-supervised learning on large graphs. In Proc. of the 17th Annual Conf. on Learning Theory (COLT'04), Banff, Alberta, 2004.
- [2] A. Blum and S. Chawla. Learning from labeled and unlabeled data using graph mincuts. In Proc. 18th International Conf. on Machine Learning, pages 19-26. Morgan Kaufmann, San Francisco, CA, 2001.
- [3] M. Herbster and G. Lever. Predicting the labelling of a graph via p -norm interpolation. In Proc. of the 22nd Annual Conf. on Learning Theory (COLT'09), 2009.
- [4] M. Herbster, M. Pontil, and S. R. Galeano. Fast prediction on a tree. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, Advances in Neural Information Processing Systems 21, pages 657-664. 2009.
- [5] X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In 20th International Conf. on Machine Learning (ICML-2003), pages 912-919, 2003

16:00 – 16:35 Sune Lehmann, Northeastern University / Harvard University

Title: Connections Matter. Communities of Links in Complex Networks

Abstract: Identifying modular structure in a complex network is generally considered a problem of assigning a community membership to each node. We propose a new perspective for the problem of community detection. Rather than defining a community as a set of densely interconnected nodes, we define a community as a set of (related) links. We show how this alternative viewpoint incorporates important features of real world networks, including overlapping communities, multi-partite structure, and hierarchical organization in a natural way. A quantitative framework for detecting and evaluating such link partitions is introduced.

16:35 – 17:10 Lars Kai Hansen, Technical University of Denmark

Title: Machine Learning in Complex Networks

Abstract: Social and other complex networks show non-trivial structure at all scales from local to global. The presence of implicit 'communities' is basic to understand network based collaboration and collective problem solving. The community detection problem can be phrased as an inference problem in a Potts model with adaptive connection strengths representing the intra- and inter-community link probabilities (Hofman & Wiggins, 2007). I will discuss community detection thresholds in the adaptive Potts model based on approximate inference using deterministic and Monte Carlo search.

17:30 Bus from DTU to Restaurant Ultimo in Tivoli

Posters

1: Michael Kai Petersen, DTU Informatics, Technical University of Denmark

Title: Sparse but emotional 3D decomposition of lyrics

Abstract: Both in music and language we rely on syntax for parsing sequences of words or tones, which based on hierarchically nested grammatical structures allow us to express and share the meaning contained within a sentence or a melodic phrase. Recent neuroimaging studies indicate that core elements of lyrical music in particular appear to be treated by the brain in a fashion similar to those of language. And as both low-level semantics of lyrics and our affective responses can be encoded in words, a simplified cognitive model can be constructed which uses LSA latent semantic analysis to emulate how we might perceive the emotional context of media. Extending the derived matrices into 3-way arrays, emotional patterns in the lyrics are aligned and compared over time based on a selection of songs. Fitting the decompositions using a sparse regression algorithm to prune excess components, and applying automatic relevance determination to determine the amount of sparsity imposed on the core array, a small number of emotional time-series contours are identified, which appear to provide generalized structures underlying the song lyrics.

This is joint work with Morten Mørup and Lars Kai Hansen, DTU Informatics, Technical University of Denmark

2: Alexander Dimitrov, Center for Computational Biology Montana State University Bozeman MT

Title: Soft clustering decoding of neural codes

Abstract: methods based on Rate Distortion theory have been successfully used to cluster stimuli and neural responses in order to study neural codes at a level of detail supported by the amount of available data. They approximate the joint stimulus-response distribution by soft-clustering paired stimulus-response observations into smaller reproductions of the stimulus and response spaces. An optimal soft clustering is found by maximizing an information-theoretic cost function subject to both equality and inequality constraints, in hundreds to thousands of dimensions.

This analytical approach has several advantages over other current approaches: it yields the most informative approximation of the encoding scheme given the available data (i.e. it gives the lowest distortion, by preserving the most mutual information between stimulus and response classes); the cost function, which is intrinsic to the problem, does not introduce implicit assumptions about the nature or linearity of the encoding scheme; it incorporates an objective, quantitative scheme for refining the codebook as more stimulus-response data becomes available; and it does not need repetitions of the stimulus under mild continuity assumptions, so the stimulus space may be investigated more thoroughly.

The method of annealing has been used to solve the corresponding high dimensional non-linear optimization problems. The annealing solutions undergo a series of bifurcations, which we study using bifurcation theory in the presence of symmetries. In this contribution we describe these symmetry breaking bifurcations in detail, and indicate some of the consequences of the form bifurcations. We present results of annealing applied to neurophysiological data.

3: Per Christian Hansen, DTU Informatics, Technical University of Denmark

Title: A parameter-choice method that exploits residual information

Abstract: All regularization methods - used for computing stable solutions to inverse problems - rely on the choice of the regularization parameter that balances the solution's smoothness and how well it fits the data. Most algorithms for choosing this regularization parameter are based on the norm of the residual vector.

However, the residual vector carries important information about the regularized solution, and therefore it can be advantageous to seek to extract more of the information available there.

We present important relations between the residual components and the amount of information that is available in the noisy data, and we show how to use statistical tools and fast Fourier transforms to extract this information efficiently. This approach leads to a computationally inexpensive parameter-choice rule based on the normalized cumulative periodogram, which is particularly suited for large-scale problems.

Joint work with Misha E. Kilmer, Tufts University, MA

4: Eric Yu-En Lu, University of Cambridge

Title: On Personality and Cognitive Constraints in Social Networks

Abstract: In the last decade, there has been a massive increase in network research across both the social and physical sciences. In the more physics-oriented studies, there have been extensive work on phenomenological models and, to a lesser extent, generative models concerning large networks with applications to social and biological networks. In sociology, on the other hand, much is devoted to the study of individual links and its social characteristics. This often involve theoretical constructs from diverse fields such as psychology and anthropology. In this poster, we present our social network models and results due to ideas from anthropology --- the influences of individual personality and human cognitive constraints. We show that the distribution of links can have significant evolutionary implications as well as act as its construct.

5: Finn Årup Nielsen, DTU Informatics, Technical University of Denmark

Title: Multivariate analysis in the Brede Toolbox

Abstract: The Brede Toolbox is a Matlab toolbox that implements some of the standard multivariate analysis algorithms, such as K-means, independent component analysis and non-negative matrix factorization (NMF). The Matlab functions operate on annotated matrices and plotting and HTML generation functions can use this annotation when reporting analysis results, e.g., such that results of hierarchical NMF can be plotted with a single function. The Brede Toolbox have been used for text mining PubMed abstracts, analysis of Wikipedia citations and for analyzing the content of the neuroinformatics Brede Database. The latter analysis involves the formation of multiple kernel density estimates of brain coordinates and subsequent multivariate analyses of these densities with the results containing 3D visualization in brain space added to the Brede Database homepage.

6: Arnold Skimminge, DTU Informatics, Technical University of Denmark

Title: Long-term regional atrophy and association with clinical outcome following severe traumatic brain injury: a tensor based morphometry study

Abstract: Two structural MR scans with a 1 year interval were acquired in 24 adult patients with severe TBI and in 14 healthy control subjects. We used tensor based morphometry to evaluate the regional distribution of brain volume change between the two scan time points. We find late atrophy to be significantly more pronounced in TBI patients with unfavourable outcome as compared to those with favourable outcome in areas related to consciousness and motor control.

7: Victoria Yanulevskaya, ISLA, Informatics Institute, University of Amsterdam, The Netherlands

Title: From massive high to compact low: towards exploiting statistical regularities in image description.

Abstract: In the state of the art computer vision algorithms for the challenging task of image categorization large set of high-dimensional features are used to describe an image. The combination of many and diverse features often yields better performance. As a result, modern algorithms have to learn from massive high dimensional data. We investigate how natural image statistics can be used to reduce high-dimensional feature spaces to lower dimensional ones.

We explore the link between visual content and edge distribution of natural images at the global level. The edge distribution can be adequately characterized by the two-parameter Weibull distribution [1]. We distinguish four different regimes of the Weibull distribution: power-law, exponential, Gaussian, and the case when the Weibull distribution is not appropriate [2]. With model selection techniques from information theory, we can determine the probability for every sub-model to explain the statistical properties of an image. We experimentally explore the occurrence of the content classes in PASCAL VOC2009 data base [3], which contains 7,819 natural images.

Power law distribution is chosen as an appropriate sub-model for 20% of the images. These images tend to have well separated foreground and uniform background regions. Only 2% of the images are Gaussian distributed, these are images which contain mostly high frequency texture. Most of the images (58%) follow the exponential distribution, which refers to moderate contrast contents. These images usually contain a lot of details at different scales. The rest of the images (20%) do not follow the Weibull distribution according to the g-test ($\alpha = 0.05$). These images tend to be composed of a few parts, each following a different sub-model. Thus, each part seems to conform the Weibull distribution, but the mixture does not collapse into one of the sub-models.

Our results show that natural image statistics explain important aspects of visual content. Each sub-model reflects a specific type of visual content, at the global level. As such, we explicitly link natural image statistics and visual content.

This is joint work with J.M. Geusebroek, ISLA, Informatics Institute, University of Amsterdam, The Netherlands

[1] Geusebroek J. M. & Smeulders A. W. M. (2003). Fragmentation in the vision of scenes. In ICCV.

[2] Yanulevskaya V. & Geusebroek J. M. (2009). Significance of the Weibull distribution and its

sub-models in natural image statistics. In VISAPP.

[3] Everingham M., Van Gool L., Williams C. K. I., Winn J. & Zisserman A. (to be prepared after the challenge workshop). The PASCAL Visual Object Classes Challenge 2009 (VOC2009) Results.

8: Carlos Alzate, ESAT-SISTA, K.U.Leuven

Title: Sparse spectral clustering with out-of-sample extensions for large-scale data

Abstract: The spectral clustering formulation is within a constrained optimization framework with primal and dual model representations allowing the clustering model to be extended to out-of-sample points together with model selection. Two approaches are used in order to obtain a sparse model which becomes important for large-scale data. The first scheme uses the incomplete Cholesky decomposition to compute low-rank approximations of a modified Laplacian which contains cluster information. A reduced set method is also presented to compute efficiently the cluster indicators for out-of-sample data. The second scheme uses an approximation of the quadratic Renyi entropy in order to actively select a subset of the data points which will be used for training. A validation subset is used for model selection to determine the number of clusters and the kernel parameters. The cluster indicators of the remaining data points are inferred using the out-of-sample extension. Experimental results with large-scale toy data and images are shown to illustrate the performance and computation times of the proposed approaches.

9: Cesar Caiafa, LABSP, RIKEN Brain Science Institute, Wako, Saitama 351-0198, JAPAN.

Title: Generalizing the CUR Matrix Decomposition to Large-Scale Multi-way Arrays

Abstract: It is known that, given a matrix $Y \in \mathbb{R}^{I \times J}$ with $\text{rank}(Y) = R$, one can perfectly reconstruct it by choosing only $P = R$ rows and columns determining a non singular intersection submatrix W and by calculating the corresponding CUR decomposition, i.e. $Y = CUR$, with $U=W^{-1}$, where matrices $C \in \mathbb{R}^{I \times P}$ and $R \in \mathbb{R}^{P \times J}$ are the selected rows and columns respectively. In this work, we extend the idea of CUR decomposition to multi-way arrays. In particular, we present a result that provides a Tucker representation of an N-way tensor $Y \in \mathbb{R}^{I_1 \times I_2 \dots \times I_N}$ in terms of a reduced subset of n-mode fibers with a core tensor obtained from the entries of the intersection sub-tensor. In this way, this Fiber Based Tensor Decomposition (FBTD), is completely determined by selecting subsets of R indices in each dimension and it is proven to be exact if the original tensor has an exact Tucker representation of order R. We provide an algorithm for the selection of proper fibers which does not require to access to all the entries of the tensor which is useful for large scale applications. Numerical results show that our FBTD model can also provide good approximations by using less indices than the order of the original Tucker tensor ($P < R$).

This is joint work with Andrzej Cichocki, LABSP, RIKEN Brain Science Institute, Wako, Saitama 351-0198, JAPAN.

10: Ricardo Henao, DTU Informatics, Technical University of Denmark

Title: Learning Graphical Model Structure with Bayesian Sparse Linear Factor Models

Abstract: In recent years, sparse linear models have been a very active field of research in machine learning community and it is also well known that learning the structure of graphical models, specifically directed acyclic graphs (DAG), is rather a difficult task to solve. In this paper we present an approach to learn DAG structure based on some recent developments to Bayesian sparse factor models. Assuming that the observed variables can be ordered in such way they can be represented as a DAG and that the value of each variable is a linear combination of values already taken by previous variables plus a driving signal, we can write a

data vector \mathbf{x} with d variables as $\mathbf{x} = \mathbf{PAP}\mathbf{x} + \mathbf{z}$, where \mathbf{A} is a strictly lower triangular weight matrix, \mathbf{P} is a permutation matrix encoding the correct order of the variables and \mathbf{z} is the driving signal. If \mathbf{A} is square we can rewrite the problem as $\mathbf{x} = \mathbf{Bz} = \mathbf{P}(\mathbf{I} - \mathbf{A})^{-1}\mathbf{Pz}$ and we end up with a linear factor model with two restrictions, (i) \mathbf{B} must be permutable to a triangular form since $(\mathbf{I} - \mathbf{A})^{-1}$ is triangular and (ii) \mathbf{z} must be non-Gaussian independent variables to ensure identifiability (up to scaling and permutations of columns, \mathbf{P}_c). We have developed a three-step algorithm which uses Bayesian slap and spike sparsity priors to estimate an ensemble of factor models, permutations and DAG models. It avoids combination of discrete (permutation) and continuous estimation (parameters) which are prone to local minima. In order to estimate the factor model we specify a Bayesian model where \mathbf{B} has a slap and spike mixture prior to allow for sparsity, \mathbf{z} has a Laplace distribution and the inference process is carried out by Gibbs sampling on the complete model hierarchy which includes a residual term. The factor model is invariant to \mathbf{P} and \mathbf{P}_c but we can make a stochastic search for \mathbf{P} and \mathbf{P}_c within the Gibbs sampling by accepting new permutations matrices according log likelihood ratios (Metropolis-Hastings) for $\mathbf{PB}\mathbf{P}_c$ masked to be lower triangular. This produces a list of candidate orderings that can be used in the DAG estimation step by specifying a similar Bayesian model on $\mathbf{x} = \mathbf{P}'\mathbf{A}\mathbf{P}'\mathbf{x} + \mathbf{z}$, where \mathbf{P}' is a candidate ordering and \mathbf{A} is strictly lower triangular with slap and spike priors. The final outcome of the algorithm is an ensemble of factor and DAG models. Model selection among these are performed using a test log likelihood. Results are presented on artificial and real data (flow cytometry measurements). We compare our model with the ICA based related method LINGAM (Linear Non-Gaussian Acyclic Model for Causal Discovery), showing that we can achieve better results with less number of examples and also that our model scales better with the number of variables.

This is joint work with Ole Winther, DTU Informatics, Technical University of Denmark and Copenhagen University

11: Lasse L. Mølgaard, DTU Informatics, Technical University of Denmark

Title: Castsearch - Context Based Spoken Document Retrieval

Abstract: This poster describes work on the development of a system for retrieval of relevant stories from broadcast news. The system utilizes a combination of audio processing and text mining. The audio processing consists of a segmentation step that partitions the audio into speech and music. The speech is further segmented into speaker segments and then transcribed using an automatic speech recognition system, to yield text input for clustering using non-negative matrix factorization (NMF). We find semantic topics that are used to evaluate the performance for topic detection. Based on these topics we show that a novel query expansion can be performed to return more intelligent search results. We also show that the query expansion helps overcome errors inferred by the automatic transcription.

12: Christian Walder, DTU Informatics, Technical University of Denmark

Title: Support Vector Machines without the Nuisance Parameter

Abstract: This paper considers kernels invariant to translation, rotation and dilation. We show that no non-trivial positive definite (p.d.) kernels exist which are radial and dilation invariant, only conditionally positive definite (c.p.d.) ones. Accordingly, we discuss the c.p.d. case and provide some novel analysis, including an elementary derivation of a c.p.d. representer theorem. On the practical side, we give a support vector machine (s.v.m.) algorithm for arbitrary c.p.d. kernels. For the thin-plate kernel this leads to a classifier with only one parameter (the amount of regularisation), which we demonstrate to be as effective as an s.v.m. with the Gaussian kernel, even though the Gaussian involves a second parameter (the length scale).

13: Rasmus Amossen, IT University of Copenhagen

Title: Faster Join-Projects and Sparse Matrix Multiplications

Abstract: Computing an equi-join followed by a duplicate eliminating projection is conventionally done by performing the two operations in serial. If some join attribute is projected away the intermediate result may be much larger than both the input and the output, and the computation could therefore potentially be performed faster by a direct procedure that does not produce such a large intermediate result. We present a new algorithm that has smaller intermediate results on worst-case inputs, and in particular is more efficient in both the RAM and I/O model. It is easy to see that join-project where the join attributes are projected away is equivalent to boolean matrix multiplication. Our results can therefore also be interpreted as improved sparse, output-sensitive matrix multiplication.

14: Line H. Clemmensen, DTU Informatics, Technical University of Denmark

Title: Supervised classification in high-dimensional sparse spaces

Abstract: Classification in high-dimensional feature spaces where interpretation and dimension reduction are of great importance is common in biological and medical applications. For these applications standard methods as microarrays, 1D NMR, and spectroscopy have become everyday tools for measuring thousands of features in samples of interest. Furthermore, the samples are often costly and therefore many such problems have few observations in relation to the number of features. Traditional linear discriminant analysis (LDA) is a widely used classification method when the sample size is larger than the number of features. I propose a sparse discriminant analysis which extends LDA to the case where the number of features exceeds that of the samples and at the same time imposes sparseness to the solution. The method uses optimal scoring to get the discriminant analysis on a regression form and then adds a ridge and a lasso constraint on the regression parameters to obtain robust and sparse solutions via the elastic net. The method is illustrated on data of microbiological fungi.

15: Morten Hansen, DTU Informatics, Technical University of Denmark

Title: Efficient Minimum-Phase Prefilter Computation Using Fast QL-Factorization

Abstract: This paper presents a novel approach for computing both the minimum-phase filter and the associated all-pass filter in a computationally efficient way using the fast QL-factorization. A desirable property of this approach is that the complexity is independent on the size of the matrix which is QL-factorized, and thereby the complexity scales with the required precision of the filters and the filter length.

16: Peter Mondrup Rasmussen, DTU Informatics, Technical University of Denmark and Center of Functionally Integrative Neuroscience

Title: Visualization of non-linear predictive models in functional neuroimaging - sensitivity maps

Abstract: During a typical neuroimaging experiment several hundreds of scans are acquired each comprising tens of thousands of spatial locations in the human brain. Traditionally, the analysis of neuroimaging data has been conducted through an univariate approach aiming at functional localization, i.e. identification of brain regions with a significant coupling between the bloodoxygenation dependent signal (BOLD) and explanatory variables in term of some sensory, motor or cognitive condition. In recent years, machine learning and pattern recognition techniques has increasingly been used to learn the statistical relationship between patterns of brain activation and the experimental conditions. Kernel based learning techniques such as support vector machines (SVM) and relevance vector machines (RVM) has successfully been applied in the analysis of various types of neuroimaging experiments. In the context of model evaluation not only classification performance is to be considered but also an informative model visualization is important: i) in evaluation of preprocessing strategies model visualization may rival the generalization performance in importance. ii) model visualization reveals where in the brain the discriminative information resides. We investigate a model visualization for non-linear kernel methods based on sensitivity analysis. The resulting sensitivity maps highlights voxels that are found relevant by the model for discrimination between brain states. Our visualization scheme is not restricted to linear models but also generalizes to non-linear kernel methods. In an XOR-like simulation we demonstrate that the visualization is a quite powerful non-linear signal detector. On in-vivo data from simple motor and visual fMRI experiments we show, that the sensitivity map is capable of generation of meaningful global summary maps of non-linear kernel methods.

This is joint work with Kristoffer H. Madsen², Torben E. Lund³ and Lars K. Hansen¹

¹ DTU Informatics, Technical University of Denmark

² Center of Functionally Integrative Neuroscience, Aarhus C

³ Danish Research Center for Magnetic Resonance, Hvidovre

17: Ole Winther, DTU Informatics, Technical University of Denmark and Copenhagen University

Title: Robust Processes for Latent Variables in Dynamical Factor Models

Abstract: In a previous work, we introduced an algorithm to learn factor models and directed acyclic graphs (DAG) within the same framework. Initially, we considered a heavy-tailed independent identically distribution for the factors and furthermore extended the model to handle smoothness of the data by using instead a Gaussian process prior. The original algorithm has three components, (i) inference of an identifiable Bayesian sparse factor model, (ii) stochastic search over variable and latent factor orderings to produce a candidate set of variable permutations and (iii) inference of a Bayesian sparse DAG model. Here we only focus on the prior distribution for the factors because we are still interested in handling smoothness of the data but without reducing the performance of our algorithm due to the presence of outliers – which are very frequent in real situations. For this purpose, we want to investigate the effect of allowing for a non-Gaussian process for the factors and to present some comparative results both using artificial and real data.

This is joint work with Ricardo Henao, DTU Informatics, Technical University of Denmark

18: Morten Arngren, DTU Informatics, Technical University of Denmark

Title: Kernel based subspace projection of near infrared hyperspectral images of maize kernels

Abstract: We present an exploratory analysis of hyperspectral 900-1700 nm images of maize kernels. The imaging device is a line scanning hyper spectral camera using a broadband NIR illumination. In order to explore the hyperspectral data we compare a series of subspace projection methods including principal component analysis and maximum autocorrelation factor analysis. The latter utilizes the fact that interesting phenomena in images exhibit spatial autocorrelation. However, linear projections often fail to grasp the underlying variability on the data. Therefore we propose to use so-called kernel version of the two afore-mentioned methods. The kernel methods implicitly transform the data to a higher dimensional space using non-linear transformations while retaining the computational complexity. Analysis on our data example illustrates that the proposed kernel maximum autocorrelation factor transform outperform the linear methods as well as kernel principal components in producing interesting projections of the data.

19: Matthew G. Liptrot, Danish Research Centre for Magnetic Resonance (DRCMR), Copenhagen University Hospital (www.drcmr.dk)

Title: Addressing the problem of path-length dependency in probabilistic tractography- The ICE-T framework

Abstract: Diffusion MRI, or diffusion weighted imaging (DWI), is a technique that measures the restricted diffusion of water molecules within brain tissue. Different reconstruction methods quantify water-diffusion anisotropy in the intra- and extra-cellular spaces of the neural environment. Fibre tracking models then use the directions of greatest diffusion as estimates of white matter fibre orientation. Several fibre tracking algorithms have emerged in the last few years that provide reproducible visualizations of three-dimensional fibre bundles. One class of these algorithms is probabilistic tractography. Probabilistic tractography [1] affords a potentially quantitative method for non-invasively generating a “connection probability map”, a frequency measure of successful vs. attempted fibre-tracking trials to remote brain areas from a given seed region. However, one of the limitations with the approach is that of path-length dependency, where the values thus produced are modulated by a gradual decline in the likelihood of successful propagation as the distance from the seed region increases [2]. This bias is unfortunately inherent to the method, as it is derived from propagation, along a potential fibre, of the very uncertainty per voxel which is required in order to generate the probability maps. Hence, in so favouring short-range connections over longer-range ones, the method's ability to perform fibre-tracking to distal regions is compromised. Surprisingly, little focus has been directed towards the path-length dependency problem (e.g. [2]). Here, we address the issue with a novel tractography framework, Iterative Confidence Enhancement for Tractography (ICE-T). This retains the stages of conventional probabilistic streamlining tractography, but then attempts to overcome the path-length dependency via the introduction of an extra stage, and a subsequent feedback loop to allow iteration of the fibre tracking stage. In this way, the ICE-T framework can be used with any conventional streamlining tractography method. Fibre-tracking was performed from a somatosensory seed region, and whereas a traditional approach with 60,000 streamlines failed to penetrate far into the contralateral hemisphere, the ICE-T version succeeded with good agreement with in-vivo tracer results. We hope that ICE-T will allow future comparison of quantitative fibre-tracking results across subjects.

[1] G. Parker, et al, A framework for a streamline-based probabilistic index of connectivity (PICO) using a structural interpretation of MRI diffusion measurements, J Magn Reson Imaging, 18(2), 2003, pp242-54

[2] D. Morris, et al, Probabilistic fibre tracking: Differentiation of connections from chance events, Neuroimage 42, 2008, pp1329-1339

This is joint work with Tim Bjørn Dyrby¹, Karam Sidaros¹

¹ Danish Research Research Centre for Magnetic Resonance (DRCMR), Copenhagen University Hospital, Hvidovre, Denmark. (www.drcmr.dk)

20: Mariya Ishteva, Department of Electrical Engineering ESAT/SCD, Katholieke Universiteit Leuven

Title: Best low multilinear rank approximation of tensors and its local minima

Abstract: We study the best low multilinear rank approximation of higher-order tensors. It is widely used as a tool for dimensionality reduction and signal subspace estimation in chemometrics, biomedical signal processing, telecommunications, etc. Searching for reliable and efficient ways of computing this approximation, we have developed a method based on the trust-region scheme. The tensor approximation problem is presented as a minimization problem on a quotient manifold and superlinear convergence speed is achieved. We compare our method with the well-known higher-order orthogonal iteration method and with a class of recently proposed Newton methods. Finally, we discuss the issue of local optima.

This is joint work with L. De Lathauwer^{1,2}, P.-A. Absil³, S. Van Huffel¹

¹ Department of Electrical Engineering ESAT/SCD, Katholieke Universiteit Leuven, Belgium,

² Group Science, Engineering and Technology, Katholieke Universiteit Leuven Campus Kortrijk, Belgium,

³ Department of Mathematical Engineering, Université catholique de Louvain, Belgium.

21: Junming Shao, University of Munich and Technical University of Munich

Title: Automated Density-based Clustering White Matter Tracts with Dynamical Time Wrapping

Abstract: Diffusion magnetic resonance imaging (MRI) provides a promising way of estimating the neural fiber pathways in the human brain non-invasively via white matter tractography. However, it is currently difficult to visualize and analyze these millions of resulting tracts quantitatively. Automated clustering these tracts would be particularly useful for neuroscience community, such as accurate neurosurgical planning, tract-based analysis, white matter atlas creation, etc. In this paper, we propose a novel framework for automatic clustering white matter tracts. The pair-wise similarity of fibers is first calculated using dynamical time wrapping, an effective and efficient distance measurement for fiber curves which indicated by the extensive experiments. Considering the inevitable noise of the data and the inherent imperfectness of fiber tractography, an outlier robust density-based clustering is applied to cluster these fibers. The imperfect fibers (noise) are easily excluded due to the vast dissimilarity with normal fibers. Extensive experiments demonstrate that our method can identifies corresponding white matter regions effectively.

22: Isabel de Sousa Amorim, DTU Informatics, Technical University of Denmark

Title: Monte Carlo based test for inferring about the unidimensionality of a sensory panel

Abstract: Sensory panels are used extensively within food research and industry as a tool for measuring human perceptual product characteristics. A sensory panel is said to be unidimensional for one variable when all assessors evaluate the attributes in the same way. In turn, reliability on scores in sensory analysis depends on the unidimensionality level of the panel. However, once estimated, the level of unidimensionality should be tested. This research describes a test based on Monte Carlo simulation process for evaluating the unidimensionality of a sensory panel. This evaluation is made using principal component analysis (PCA). For describing the Monte Carlo unidimensionality test (MCUT) virtual and real panels were used: one is a simulation study to evaluate the performance of the test (type I error rates and power) and the other is an illustration of the method using an evaluation of an actual sensory panel. The Monte Carlo based test can be adopted as a tool for checking the efficiency of the training process of a sensory panel. The MCUT revealed to be powerful for large panel and sample sizes and efficient for evaluating unidimensionality of sensory panels.

23: Andrea Campagna, IT University of Copenhagen

Title: Computing cosine and lift measures via biased sampling

Abstract: Sampling-based methods have previously been proposed for the problem of finding interesting associations in data, even for low-support items. While these methods do not guarantee precise results, they are vastly more efficient than approaches that rely on exact counting. However, for some of the most important similarity measures such as the lift, cosine, and all_confidence measures, no such methods have been known. In this paper we show how these measures can be supported by a simple biased sampling method. We demonstrate theoretically that our method is superior to exact methods when the threshold for "interesting similarity" is above the average pairwise similarity. Our method is particularly good when transactions contain many items. We confirm in experiments on both real and synthetic data that this gives a significant speedup (sometimes much larger than the theoretical guarantees). Reductions in computation time of over an order of magnitude are observed on standard benchmark data sets. Reductions in space usage are also often observed.

24: Risi Kondor, University College London, England

Title: The skew spectrum of graphs

Abstract: The central issue in representing graph-structured data instances in learning algorithms is designing features which are invariant to permuting the numbering of the vertices. We present a new system of invariant graph features which we call the skew spectrum of graphs. The skew spectrum is based on mapping the adjacency matrix of any (weighted, directed, unlabeled) graph to a function on the symmetric group and computing bispectral invariants. The reduced form of the skew spectrum is computable in $O(n^3)$ time, and experiments show that on several benchmark datasets it can outperform state of the art graph kernels.

joint work with Karsten Borgwardt and Nino Shervashidze

List of Participants

| | | |
|-----------------|--------------|--|
| Trine Julie | Abrahamsen | DTU Informatics, Denmark |
| Tommy Sonne | Alstrøm | DTU Informatics, Denmark |
| Carlos | Alzate | K.U. Leuven, ESAT-SCD, Belgium |
| Isabel | Amorim | DTU Informatics, Denmark |
| Rasmus | Amossen | IT University of Copenhagen |
| Tobias | Andersen | DTU Informatics, Denmark |
| Morten | Arngren | DTU Informatics / FOSS Analytical A/S, Denmark |
| John | Ashburner | University College London, England |
| Vadim | Axelrod | Tel-Aviv University, Psychology Department Israel |
| Ricardo | Baeza-Yates | Yahoo! Research, Barcelona, Spain |
| Philip | Bille | DTU Informatics, Denmark |
| Rasmus | Bro | University of Copenhagen, Faculty of Life Sciences, Denmark |
| Per Bruun | Brockhoff | DTU Informatics, Denmark |
| Joachim | Buhmann | ETH Zürich, Compartment of Computer Science, Switzerland |
| Andrius | Butkus | DTU Informatics, Denmark |
| Cesar | Caiafa | Brain Science Institute (BSI), Riken, Japan |
| Andrea | Campagna | IT University of Copenhagen, Denmark |
| Joaquin Q. | Candela | Microsoft Research Cambridge, England |
| Pedro | Cano | Barcelona Music & Audio Technologies, Spain |
| Gunnar | Carlsson | Stanford University, USA |
| Edward | Chang | Google Research Beijing, P.R. China |
| Line | Clemmensen | DTU Informatics, Denmark |
| Pierre | Comon | University of Nice Sophia-Antipolis, France |
| Nello | Cristianini | University of Bristol, England |
| Anders Bjørholm | Dahl | DTU Informatics, Denmark |
| Sune | Darkner | DTU Informatics, Denmark |
| Lieven | De Lathauwer | K.U. Leuven, Belgium |
| Ayhan | Demiriz | Sakarya University, Dept. Industrial Engineering, Turkey |
| Alexander | Dimitrov | Montana State University, USA |
| Tim Bjørn | Dyrby | Danish Researach Centre for Magnetic Resonance, Denmark |
| Lars | Eldén | Linköping University, Sweden |
| Charles | Elkan | University of California San Diego, USA |
| Bjarne | Ersbøll | DTU Informatics, Denmark |
| Jerome | Friedman | Stanford University, USA |
| Søren | Halling | Atlantstøkni, Faroe Islands |
| Lars Kai | Hansen | DTU Informatics, Denmark |
| Mads Fogtmann | Hansen | DTU Informatics, Denmark |
| Morten | Hansen | DTU Informatics, Denmark |
| Per Christian | Hansen | DTU Informatics, Denmark |
| Ricardo | Henao | DTU Informatics, Denmark |
| Mark | Herbster | University College London, England |
| Anders | Haaning | Jyske Bank, Denmark |
| Mariya | Ishteva | K.U. Leuven, Belgium |
| Bjørn Sand | Jensen | DTU Informatics, Denmark |
| Søren Holdt | Jensen | Aalborg University, Dept. Electronic Systems, Denmark |
| Kristjana Yr | Jonsdottir | Center of Functional Integrative Neuroscience, Denmark |
| Seliz | Karadogan | DTU Informatics, Denmark |
| Samuel | Kaski | Helsinki University of Technology, Finland |
| Gitte Moos | Knudsen | Center Integrated Molecular Brain Imaging, Rigshospitalet, Denmark |
| Risi | Kondor | University College London, England |

| | | |
|-----------------|----------------|--|
| Konrad | Krysiak-Baltyn | DTU Systembiology, Denmark |
| Søren | Kyllingsbæk | University of Copenhagen, Department of Psychology, Denmark |
| Jakob Eg | Larsen | DTU Informatics, Denmark |
| Jan | Larsen | DTU Informatics, Denmark |
| Rasmus | Larsen | DTU Informatics, Denmark |
| Steffen | Lauritzen | University of Oxford, England |
| Neil | Lawrence | University of Manchester, England |
| Sune | Lehmann | Northeastern University / Harvard University, USA |
| Lek-Heng | Lim | University of California, Berkeley, USA |
| Matthew | Liptrot | Danish Research Centre for Magnetic Resonance, Denmark |
| Yu-En | Lu | University of Cambridge, Computer Laboratory |
| Kaj | Madsen | DTU Informatics, Denmark |
| Kristoffer H. | Madsen | Danish Research Centre for Magnetic Resonance, Denmark |
| Michael | Mahoney | Stanford University, USA |
| Scott | Makeig | University of California San Diego, USA |
| Peter | Mannfolk | Lund University, Dept. Medical Radiation Physics, Sweden |
| Wassel | Mashagbeh | US Department of Commerce, USA |
| Gemma | Monte | CIBERSAM, Spain |
| Klaus | Mosegaard | University of Copenhagen, Denmark |
| Lasse Lohilahti | Mølgaard | DTU Informatics, Denmark |
| Morten | Mørup | DTU Informatics, Denmark |
| Christian | Mårtenson | Swedish Defence Research Agency, Sweden |
| Allan Aasbjerg | Nielsen | DTU Informatics, Denmark |
| Finn Årup | Nielsen | DTU Informatics, Denmark |
| Simon | Nielsen | DTU Informatics, Denmark |
| Hildur | Ólafsdóttir | DTU Informatics, Denmark |
| Benjamin | Pedersen | Atlantstøkni, Faroe Islands |
| Kaare Brandt | Petersen | Epital ApS, Denmark |
| Michael Kai | Petersen | DTU Informatics, Denmark |
| Christian | Pipper | University of Copenhagen, Faculty of Life Sciences, Denmark |
| Tomaso | Poggio | Massachusetts Institute of Technology, USA |
| Krzysztof | Potempa | Lundbeck, Denmark |
| Philippe | Preux | INRIA, France |
| Morten | Proschowsky | DTU Informatics, Denmark |
| Peter Mondrup | Rasmussen | DTU Informatics, Denmark |
| Christian | Ritz | University of Copenhagen, Faculty of Life Sciences |
| Mikkel | Schmidt | University of Cambridge, England |
| Bernhard | Schölkopf | Max Planck Institute for Biological Cybernetics, Tübingen, Germany |
| Junming | Shao | University of Munich/Technical University of Munich, Germany |
| Arnold | Skimminge | DTU Informatics, Denmark |
| Carsten | Stahlhut | DTU Informatics, Denmark |
| Jeremy | Taylor | FINA Technologies Cambridge, MA, USA |
| Yee Whye | Teh | University College London, England |
| Philipp | Trénel | AgroTech A/S, Denmark |
| Christian | Walder | DTU Informatics, Denmark |
| Manda | Winlaw | University of Waterloo, Canada |
| Ole | Winther | DTU Informatics, Denmark |
| Victoria | Yanulevskaya | University of Amsterdam, The Netherlands |

