NEURAL INFORMATION PROCESSING SYSTEMS WORKSHOP 2010

Whistler,
British Columbia
December 10 -11th

Neural Information
Processing Systems
Foundation

2010 WORKSHOP BOOK

NIPS 2010

NEURAL INFORMATION PROCESSING SYSTEMS WORKSHOP

TUTORIALS
December 6, 2010
Hyatt Regency
Vancouver, BC, Canada

CONFERENCE SESSIONS
December 6-9, 2010
Hyatt Regency
Vancouver, BC, Canada

SAM ROWEIS SYMPOSIUM
December 9, 2010
Hyatt Regency
Vancouver, BC, Canada

WORKSHOP
December 10-11, 2010
The Westin Resort & Spa
The Hilton Whistler Resort & Spa
Whistler, BC Canada

At the Vancouver symposium in honor of Sam Roweis, five invited speakers will discuss Sam's contributions to machine learning and other areas of science and engineering. In Whistler there are 28 workshops covering a wide range of topics in parallel sessions. They provide an exciting preview for future trends in Neural Information Processing Systems, complimenting the main conference. The workshops were organized by Neil Lawrence.

Neural Information Processing Systems Foundation

TABLE OF CONTENTS

Neural Information Processing Systems Foundation

ORGANIZING COMMITTEE

General Chairs	**John Lafferty**, Carnegie Mellon University, **Chris Williams**, University of Edinburgh
Program Chairs	**Richard Zemel**, University of Toronto, **John Shawe-Taylor**, University College London
Tutorials Chair	**Kevin Murphy**, University of British Columbia
Workshop Chair	**Neil D Lawrence**, University of Manchester
Demonstration Chair	**Isabelle Guyon**, Clopinet
Publications Chair & Electronic Proceedings Chair	**Aron Culotta**, Southeastern Louisiana University
Workflow Master	**Laurent Charlin**, University of Toronto

PROGRAM COMMITTEE

Maria Florina Balcan, Georgia Institute of Technology
Mikhail Belkin, The Ohio State University
Shai Ben-David, University of Waterloo
Andrew Blake . Microsoft Research
David Blei. Princeton University
Rich Caruana, Microsoft Research
Olivier Chapelle, Yahoo Research
Corinna Cortes, Google, Inc
Hal Daume III, University of Utah
Tijl De Bie, University of Bristol
Nando de Freitas, UBC
Ofer Dekel, Microsoft Research
David Dunson, Duke University
Li Fei-Fei, Stanford University
Rob Fergus, New York University
Kenji Fukumizu, Institute of Statistical Mathematics
Sally Goldman, Google, Inc. & Washington University in St. Louis
Matthias Hein, Saarland University
Katherine Heller, University of Cambridge
Tony Jebara, Columbia University
Charles Kemp, Carnegie Mellon University
John Langford, Yahoo Research
Tai Sing Lee, Carnegie Mellon University
Wee Sun Lee, National University of Singapore
Mate Lengyel, University of Cambridge
Chih-Jen Lin, National Taiwan University
Bernabe Linares-Barranco, Instituto de Microelectronica de Sevilla
Phil Long, Google, Inc
Jon McAuliffe, UC Berkeley
Remi Munos, INRIA Lille - Nord Europe
Iain Murray, University of Toronto/Edinburgh
Yael Niv, Princeton University
William Noble, University of Washington
Rob Nowak, University of Wisconsin-Madison
Jan Peters, Max-Planck Institute of Biological Cybernetics
Jonathan Pillow, University of Texas at Austin
Pascal Poupart, University of Waterloo
Pradeep Ravikumar, University of Texas at Austin
Irina Rish, IBM T.J. Watson Research Center
Matthias Seeger, Saarland University
Nathan Srebro, TTI-Chicago
Ben Taskar, University of Pennsylvania
Bill Triggs, CNRS
Raquel Urtasun, TTI-Chicago
Kilian Weinberger, Washington University in St. Louis
Tong Zhang, Rutgers
Dengyong Zhou, Microsoft Research
Xiaojin (Jerry) Zhu, University of Wisconsin-Madison

NIPS would like to especially thank Microsoft Research for their donation of Conference Management Toolkit (CMT) software and server space.

NIPS FOUNDATION OFFICERS & BOARD MEMBERS

CORE LOGISTICS TEAM

The running of NIPS would not be possible without the help of many volunteers, students, researchers and administrators who donate their valuable time and energy to assist the conference in various ways. However, there is a core team at the Salk Institute whose tireless efforts make the conference run smoothly and efficiently every year. This year, NIPS would particularly like to acknowlege the exceptional work of:

Lee Campbell - IT Manager
Chris Hiestand - Webmaster
Sheri Leone - Volunteer Coordinator
Mary Ellen Perry - Executive Director

AWARDS

OUTSTANDING STUDENT PAPER AWARDS
Construction of Dependent Dirichlet Processes based on Poisson Processes
Dahua Lin*, Eric Grimson and John Fisher

A Theory of Multiclass Boosting
Indraneel Mukherjee* and Robert E Schapire

STUDENT PAPER HONORABLE MENTIONS
MAP estimation in Binary MRFs via Bipartite Multi-cuts
Sashank Jakkam Reddi*, Sunita Sarawagi and Sundar Vishwanathan

The Multidimensional Wisdom of Crowds
Peter Welinder*, Steve Branson, Serge Belongie and Pietro Perona

* Winner

SPONSORS

NIPS gratefully acknowledges the generosity of those individuals and organizations who have provided financial support for the NIPS 2010 conference. The financial support enabled us to sponsor student travel and participation, the outstanding student paper awards, the demonstration track and the opening buffet.

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
UNITED STATES AIR FORCE

Microsoft Research

ELSEVIER
Artificial Intelligence

Google™

IBM Research

PASCAL2
Pattern Analysis, Statistical Modelling and Computational Learning

BRAIN PRODUCTS
Solutions for neurophysiological research

TWO SIGMA

TOYOTA

D E Shaw & Co

YAHOO! LABS

intel®

Springer
Machine Learning Journal

PROGRAM HIGHLIGHTS

Thursday, December 9th

The Sam Roweis Symposium
2.00 - 5.00 — Hyatt Vancouver

Registration — Westin Emerald foyer
4:00 - 8:00

Reception (Hors d'œuvres and Non-alcoholic Beverages)
6:30 - 8:00 — Westin: Emerald

Opening Remarks
7:00 - 7:15 — Westin: Emerald

FRIDAY, DECEMBER 10TH

Breakfast
6:30 - 8:00 — Westin Emerald

Registration — Westin Emerald foyer
7:00 - 11:00

Friday Workshops

All workshops run from 7:30 to 10:30AM and from 3:30 to 6:30PM
with Coffee breaks from 9:00 to 9:30AM and 5:00 to 5:30PM

WS2	Beyond classification: Machine Learning for next generation Computer Vision challenges	Westin: Alpine BC
WS5	Coarse-to-Fine Learning and Inference	Westin: Alpine DE
WS6	Computational Social Science and the Wisdom of Crowds	Westin: Callaghan
WS7	Decision Making with Multiple Imperfect Decision Makers	Hilton: Black Tusk
WS8	Deep Learning and Unsupervised Feature Learning	Hilton: Cheakmus
WS13	Machine Learning for Assistive Technologies	Westin: Glacier
WS15	Machine Learning in Computational Biology	Hilton: Sutcliffe B
WS16	Machine Learning in Online Advertising	Hilton: Diamond Head
WS18	Modeling Human Communication Dynamics	Westin: Alpine A
WS19	Monte Carlo Methods for Bayesian Inference in Modern Day Applications	Hilton: Mt Currie North
WS23	Optimization for Machine Learning	Westin: Emerald A
WS24	Practical Application of Sparse Modeling: Open Issues and New Directions	Hilton: Sutcliffe A
WS26	Robust Statistical Learning	Hilton: Mt Currie South
WS27	Tensors, Kernels, and Machine Learning	Westin: Nordic

SATURDAY, DECEMBER 11TH

Breakfast 6:30 - 8:00	Westin Emerald
Registration 7:00 - 11:00	Westin Emerald foyer

Saturday Workshops

All workshops run from 7:30 to 10:30AM and from 3:30 to 6:30PM
with Coffee breaks from 9:00 to 9:30AM and 5:00 to 5:30PM

WS1	Advances in Activity-Dependent Synaptic Plasticity	Westin: Callaghan
WS3	Challenges of Data Visualization	Westin: Alpine DE
WS4	Charting Chemical Space: Challenges and Opportunities for AI and Machine Learning	Westin: Glacier
WS9	Discrete Optimization in Machine Learning: Structures, Algorithms and Applications	Hilton: Cheakmus
WS10	Learning and Planning from Batch Time Series Data	Hilton: Sutcliffe A
WS11	Learning on Cores, Clusters, and Clouds	Hilton: Mt Currie South
WS12	Low-rank Methods for Large-scale Machine Learning	Westin: Alpine BC
WS14	Machine Learning for Social Computing	Westin: Alpine A
WS17	Machine Learning meets Computational Photography	Hilton: Black Tusk
WS20	Networks Across Disciplines: Theory and Applications	Westin: Nordic
WS21	New Directions in Multiple Kernel Learning	Hilton: Mt Currie North
WS22	Numerical Mathematics Challenges in Machine Learning	Hilton: Diamond Head
WS25	Predictive Models in Personalized Medicine	Hilton: Sutcliffe B
WS28	Transfer Learning Via Rich Generative Models	Westin: Emerald A

*** * * PLEASE NOTE * * ***
SOME WORKSHOPS RUN ON DIFFERENT SCHEDULES.
PLEASE CHECK TIMINGS ON THE SUBSEQUENT PAGES.
INFORMATION ABOUT LIFT TICKETS AND BUS TICKETS WILL BE POSTED AT THE REGISTRATION DESK IN THE WESTIN LOBBY.

The Sam Roweis Symposium

LOCATION
Hyatt Regency
Thursday, 2:00 - 5:00

Session Chair
Maneesh Sahari

SCHEDULE

14:00-14:30	Unifying Views in Unsupervised Learning Zoubin Ghahramani
14:30-15:00	Manifold Learning Lawrence Saul
15:00-15:30	A Probabilistic Approach to Data Visualization Geoffrey Hinton
16:30-17:00	Coffee Break
16:00-16:30	Learning Structural Sparsity Yoram Singer
16:30-17:00	Automating Astronomy David W. Hogg

INVITED SPEAKERS

Unifying Views in Unsupervised Learning
Zoubin Ghahramani, University of Cambridge & CMU

The NIPS community has benefited greatly from Sam Roweis' insights into the connections between different models and algorithms. I will review our work on a 'unifying' framework for linear Gaussian models, which formed the backbone of the NIPS Tutorial Sam and I gave in 1999. This framework highlighted connections between factor analysis, PCA, mixture models, HMMs, state-space models, and ICA, had the EM algorithm as the all-purpose swiss-army-knife of learning algorithms, and culminated in a 'graphical model for graphical models' depicting the connections. Though perhaps well-known now, those connections were surprising at the time (at least to us) and resulted in a more coherent and systematic view of statistical machine learning that has endured to this day. Inspired by this approach, I will present some newer unifying views, of kernel methods, and of nonparametric Bayesian models.

Manifold Learning
Lawrence Saul, UC San Diego

How can we detect low dimensional structure in high dimensional data? Sam and I worked feverishly on this problem for a number of years. We were particularly interested in analyzing high dimensional data that lies on or near a low dimensional manifold. I will describe the algorithm, locally linear embedding (LLE), that we developed for this problem. I will conclude by relating LLE to more recent work in manifold learning and sketching some future directions for research.

Learning Structural Sparsity
Yoram Singer, Google

In the past years my work focused on algorithms for learning high dimensional yet sparse models from very large datasets. During the two years that Sam spent at Google, he greatly influenced my course of research on large scale learning of structural sparsity. He was too humble and too busy to formally co-author any of the papers that constitute the talk (see http://magicbroom.info/Sparsity.html). Yet, many parts of this talk would not have materialized without his encouragement, feedback, and ideas. In the talk I review the design, analysis and implementation of sparsity promoting learning algorithms, including coordinate and mirror descent with non-smooth regularization, forward-backward splitting algorithms, and other recently devised algorithms for sparse models. I will conclude with an overview of new work on learning self pruning decision trees and structured histograms by combining exponential models with sparsity promoting regularization.

A Probabilistic Approach to Data Visualization
Geoffrey E. Hinton, University of Toronto

Dimensionality reduction methods allow us to visualize the structure of large, high-dimensional datasets by giving each data-point a location in a two-dimensional map. Sam Roweis was involved in the development of several different methods for producing maps that preserve local similarity by displaying very similar data-points at nearby locations in the map without worrying too much about the map distances between dissimilar data-points. One of these methods, called Stochastic Neighbor Embedding, converts the problem of finding a good map into the problem of matching two probability distributions. It uses the density under a high-dimensional Gaussian centered at each data-point to determine the probability of picking each of the other data-points as a neighbor. It then uses exactly the same method to determine neighbor probabilities using the two-dimensional locations of the corresponding map points. The aim is to move the map points so that the neighbor probabilities computed in the high-dimensional data-space are well-modeled by the neighbor probabilities computed in the low-dimensional map. This leads to very nice maps for a variety of datasets. I will describe some further developments of this method that lead to even better maps.

Automating Astronomy
David W. Hogg, NYU

Telescopes and their instruments produce digital images. This makes astronomy a playground for computer vision. It is "easy" because the viewpoint is fixed, the illumination is fixed (on human timescales), and the range of objects viewed is limited (at finite resolution). It is hard because the data are beyond multi-spectral (current data span 17 orders of magnitude in wavelength) and we care deeply about the objects imaged at extremely low signal-to-noise. In precise contexts we are compelled to model the data probabilistically; this requires techniques of machine learning. Successes in this area have implications for our understanding of the fundamental laws of physics, the dark matter, and the initial conditions of the Universe.

IN MEMORIAM

Sam Roweis 1972-2010

The Neural Information Processing Systems Foundation mourns the untimely death of Sam T. Roweis on January 12, 2010. Sam was a brilliant scientist and engineer whose work deeply influenced the fields of artificial intelligence, machine learning, applied mathematics, neural computation, and observational science. He was also a strong advocate for the use of machine learning and computational statistics for scientific data analysis and discovery.

Sam T. Roweis was born on April 27, 1972. He graduated from secondary school as valedictorian of the University of Toronto Schools in 1990, and obtained a bachelor's degree with honors from the University of Toronto Engineering Science Program four years later. In 1994 he joined the Computation and Neural Systems PhD program at the California Institute of Technology, working under the supervision of John J. Hopfield. After earning his PhD in 1999, Sam took a postdoctoral position in London with the Gatsby Computational Neuroscience Unit. After his postdoc, some time at MIT, and a stint with the startup company WhizBang! Labs, Sam took a faculty job at the University of Toronto. In 2005 Sam spent a semester at MIT and in 2006 he was named a fellow of the Canadian Institute for Advanced Research (CIfAR) and received tenure at Toronto. He joined Google's research labs in San Francisco and Mountain View in 2007. Sam moved to the Computer Science Department at NYU's Courant Institute as an Associate Professor in September 2009.

Among Sam's many achievements in machine learning was a new form of Independent Component Analysis (ICA) that could be used to separate multiple audio sources from a single microphone signal, the Locally Linear Embedding algorithm (LLE) with Lawrence Saul, which revolutionized the field of dimensionality reduction, Stochastic Neighborhood Embedding (SNE) and Neighborhood Component Analysis (NCA), and the astrometry.net system with David W. Hogg that can take any picture of the sky from any source, and instantly identify the location, orientation, and magnification of the image, as well as name each object (star, galaxy, nebula) it contains.

Sam had a singular gift: to him, any complex concept was naturally reduced to a simple set of ideas, each of which had clear analogies in other (often very distant) realms. This gift allowed him to explain the key idea behind anything in just a few minutes. Combined with contagious enthusiasm, this made him an unusually gifted teacher and speaker. His talks and discussions were clear and highly entertaining. His tutorial lectures on graphical models and metric learning, available on video at videolectures.net, have been viewed over 25,000 times. He would often begin group meetings by giving a puzzle, the solution of which was always beautiful, enlightening, or hilarious.

Many members of the research community became friends with Sam, because of his warm and friendly personality, his communicative smile, and his natural inclination to engagement and enthusiasm. Sam inspired many students to pursue a career in research, and to focus their research on machine learning and artificial intelligence because of his broad interests, his clear-sightedness, his sense of humor, his warmth and his infectious enthusiasm.

In the last year of his life, Sam was battling severe depression but his wonderfully professional demeanor concealed this from most of his friends and colleagues. He is greatly mourned by his colleagues and students at NYU, who extend their sympathy to his many friends in the broader research community, especially at the University of Toronto, the Gatsby Neuroscience Unit, and Google Research. Most of all, we express our deepest sympathy to his wife Meredith, his twin baby daughters Aya and Orli, and his father Shoukry.

Yann LeCun, David Hogg, Zoubin Ghahramani, Geoffrey Hinton.

Advances in Activity-Dependent Synaptic Plasticity

http://www.pitt.edu/~pwm/Plasticity2010/

LOCATION
Westin: Callaghan
Saturday, 07:30 - 10:30 and 3:30 - 6:30

Paul W. Munro pwm@pitt.edu
University of Pittsburgh

Abstract

Since Hebb articulated his Neurophysiological Postulate in 1949 up to the present day, the relationship between synapse modification and neuronal activity has been the subject of enormous interest. Laboratory studies have revealed phenomena such as LTP, LTD, and STDP. Theoretical developments have both inspired studies and been inspired by them. The intent of the proposed workshop is to foster communication among researchers in this field. The workshop is intended to be of interest to experimentalists and modelers studying plasticity from the neurobiological level to the cognitive level. The workshop is targeted toward researchers in this area, hopefully drawing a 50/50 mix of experimental results and theoretical ideas. Another goal is to bring together established researchers with grad students and postdocs

SCHEDULE

7:30-7:45	Opening remarks
7:45-8:15	Activity-dependent modulation and modification of electrical synapses Julie Haas
8:15-8:45	A theory of loop formation and elimination by STDP James Kozlowski
8:45-9:00	Discussion
9:00-9:30	Spatio-temporal credit assignment in population learning Johannes Friedrich
9:30-10:00	Functional Requirements for Reward-modulated STDP Nicolas Fremaux
10:00-10:30	Discussion
15:45-16:15	Long-term synaptic modification of cortical synapses improves sensory perception Rob Froemke
16:15-16:45	Plasticity mechanisms underlying optimal odor representation in insect olfaction Maxim Bazhenov
16:45-17:00	Coffee Break
17:00-17:30	Why is connectivity in barrel cortex different from that in visual cortex? --A plasticity model Claudia Clopath
17:30-18:30	Panel discussion

INVITED SPEAKERS

Plasticity mechanisms underlying optimal odor representation in insect olfaction
Maxim Bazhenov, U. C. Riverside

Ideally, a coding strategy used by a sensory system should provide an optimal representation across the full possible range of stimulation conditions. For the olfactory system, this task involves optimally encoding odors at different concentrations, an ability critical in many species for survival. How successive layers of neural circuits in the olfactory system regulate sensory input to maintain stable odor representations across broad ranges of concentration remains a mystery. Drawing on results obtained with biophysical network models of insect olfaction, I will discuss synaptic plasticity mechanisms that contribute to optimal encoding olfactory information at different levels of odor processing. I will present a hypothesis how circuit properties of the olfactory system allow adopting an optimal strategy of information processing, shifting from coincidence detection to temporal integration as the odor concentration changes. Plasticity operating on synapses between different processing levels ensures sparse odor representation and optimal discrimination performance between odors. The ability to shift operational modes between coincidence detection and temporal integration can be achieved readily by a combination of environmental contingencies and network interactions

Why is connectivity in barrel cortex different from that in visual cortex? --A plasticity model
Claudia Clopath, CNRS

Electrophysiological connectivity patterns in cortex often show a few strong connections in a sea of weak connections. In some brain areas a large fraction of strong connections are bidirectional, in others they are mainly unidirectional. In order to explain these connectivity patterns, we use a model of Spike-Timing-Dependent Plasticity where synaptic changes depend on presynaptic spike arrival and the postsynaptic membrane potential, filtered with two different time constants. The model describes several nonlinear effects in STDP experiments, as well as the voltage dependence of plasticity under voltage clamp and classical paradigms of LTP/LTD induction. We show that in a simulated recurrent network of spiking neurons our plasticity rule leads not only to development of localized receptive fields, but also to connectivity patterns that reflect the neural code: for temporal coding paradigms with spatio-temporal input correlations, strong connections are predominantly unidirectional, whereas they are bidirectional under rate coded input with spatial correlations only. Thus variable connectivity patterns in the brain, mainly unidirectional in barrel cortex versus bidirectional in visual cortex, could reflect different coding principles across brain areas.

Functional Requirements for Reward-modulated STDP
Nicolas Fremaux, EPFL

As a biologically plausible paradigm for learning in spiking neural networks, spike-timing dependent plasticity (STDP) has been shown to perform well in unsupervised learning tasks such as receptive field development. However, STDP fails to take behavioral relevance into account, and as such is inadequate to explain a vast range of learning tasks in which the final outcome is conditioned on the prior execution of a series of actions. In this talk, I will show that the addition of a third, global, reward-based factor to the pre-and post-synaptic factors of STDP is a promising solution to this problem, with strong experimental and theoretical motivations. I will derive simple functional requirements for these rules, and illustrate them in a motor sequence learning task. These requirements suggest that the brain needs a ''critic'' structure, constantly evaluating the potential for rewarding events. I will propose a biologically plausible implementation of such a structure, that performs motor or navigational tasks.

Spatio-temporal credit assignment in population learning
Johannes Friedrich, University of Bern

In learning from trial and error animals need to relate behavioral decisions to environmental reinforcement even though it may be difficult to assign credit to a particular decision when outcomes are uncertain or subject to delays. When considering the biophysical basis of learning, the credit assignment problem is compounded because the behavioral decisions itself result from the spatio-temporal aggregation of many synaptic releases. We present a model of plasticity induction for reinforcement learning in a population of leaky integrate-and-fire neurons which is based on a cascade of synaptic memory traces. Each synaptic cascade correlates presynaptic input first with postsynaptic events, next with the behavioral decisions and finally with external reinforcement. For operant conditioning, learning succeeds even when reinforcement is delivered with a delay so large that temporal contiguity between decision and pertinent reward is lost due to intervening decisions which are themselves subject to delayed reinforcement. Hence, in contrast to temporal difference (TD) learning approaches, our rule can deal with non-Markovian decision processes where actions depend on the past history. Simulations of other tasks where TD learning fails highlight the robustness of the proposed scheme. Our rule is also explicit with regard to the assumed biophysical mechanisms.

Long-term synaptic modification of cortical synapses improves sensory perception
Rob Froemke, NYU

Synapses and receptive fields of the cerebral cortex are plastic. However, changes to specific inputs must be coordinated within neural networks to ensure that excitability and feature selectivity are appropriately configured for perception of the sensory environment. Here we measured long-term modifications of auditory cortical excitatory synapses in vivo, induced by pairing acoustic stimuli with activation of the nucleus basalis neuromodulatory system. Positive and negative changes to specific inputs were precisely orchestrated across entire receptive fields, conserving mean excitation while reducing overall variance. Computational analysis indicated that decreased variability should increase detection and discrimination of near-threshold stimuli, and this was confirmed psychophysically in behaving animals. Thus direct modification of specific cortical inputs leads to wide-scale synaptic changes, which collectively support improved sensory perception and enhanced behavioral performance.

Activity-dependent modulation and modification of electrical synapses
Julie Haas, Harvard University

Activity-dependent forms of synaptic plasticity have been widely studied at chemical synapses, but the effects of activity on electrical synaptic strength are unknown. Electrical synapses, mediated by gap junction channels, are widespread in both vertebrate and invertebrate neuronal systems. In mature mammals, electrical synapses are most often comprised of connexin 36 and couple inhibitory neurons of similar functional and biochemical profiles. The thalamic reticular nucleus (TRN), which provides the main source of feedback inhibition to thalamocortical projections, is composed exclusively of GABAergic inhibitory neurons that express an extremely high density of gap junctions. Like other thalamic neurons, TRN cells have two voltage-dependent firing modes: burst and tonic spiking, which correspond to different states of alertness and attention. Using dual whole-cell recordings in vitro, we demonstrate that electrical synaptic strength is modified by induced paired activity in coupled pairs of recorded cells. We also show that changes in electrical coupling strength evolve within minutes and are long-lasting. These activity-dependent modifications of electrical coupling strength may exert state-dependent influences on the nature and synchrony of postsynaptic inhibition provided by these neurons.

A theory of loop formation and elimination by STDP
James Kozlowski, IBM

We show that the local spike timing-dependent plasticity (STDP) rule has the effect of regulating the trans-synaptic weights of loops of any length within a simulated network of neurons. We show that depending on STDP's polarity, functional loops are formed or eliminated in networks driven to normal spiking conditions by random, partially correlated inputs, where functional loops comprise synaptic weights that exceed a positive threshold. We further prove that STDP is a form of loop-regulating plasticity for the case of a linear network driven by noise. Thus a notable local synaptic learning rule makes a specific prediction about synapses in the brain in which standard STDP is present: that under normal spiking conditions, they should participate in predominantly feed-forward connections at all scales. Our model implies that any deviations from this prediction would require a substantial modification to the hypothesized role for standard STDP Given its widespread occurrence in the brain, we predict that STDP could also regulate long range functional loops among individual neurons across all brain scales, up to, and including, the scale of global brain network topology.

Beyond Classification: Machine Learning for Next Generation Computer Vision challenges

http://sites.google.com/site/mlngcvc/

LOCATION
Westin: Alpine BC
Friday, 07:30 - 10:30 and 3:30 - 5:30

Craig Saunders, Xerox Research Centre Europe

Jakob Verbeek, INRIA

Svetlana Lazebnik, University of North Carolina

ABSTRACT

This workshop seeks to excite and inform researchers to tackle the next level of problems in the area of Computer Vision. The idea is to both give Computer Vision researchers access to the latest Machine Learning research, and also to communicate to researchers in the machine learning community some of the latest challenges in computer vision, in order to stimulate the emergence of the next generation of learning techniques. The workshop itself is motivated from several different points of view:

1. There is a great interest in and take-up of machine learning techniques in the computer vision community. In top vision conferences such as CVPR, machine learning is prevalent: there is widespread use of Bayesian Techniques, Kernel Methods, Structured Prediction, Deep Learning, etc.; and many vision conferences have featured invited speakers from the machine learning community.

2. Despite the quality of this research and the significant adoption of machine learning techniques, often such techniques are used as ‘‘black box’’ parts of a pipeline, performing traditional tasks such as classification or feature selection, rather than fundamentally taking a learning approach to solving some of the unique problems arising in real-world vision applications.

3. Beyond object recognition and robot navigation, many interesting problems in computer vision are less well known. These include more complex tasks such as joint geometric/semantic scene parsing, object discovery, modeling of visual attributes, image aesthetics, etc.

4. Even within the domain of ‘‘classic’’ recognition systems, we also face significant challenges in scaling up machine learning techniques to millions of images and thousands of categories (consider for example the ImageNet data set).

5. Images often come with extra multi-modal information (social network graphs, user preference, implicit feedback indicators, etc) and this information is often poorly used, or integrated in an ad-hoc fashion.

This workshop therefore seeks to bring together machine learning and computer vision researchers to discuss these challenges, show current progress, highlight open questions and stimulate promising future research.

SCHEDULE

7:30-7:45	Opening remarks, welcome, overview of objectives of the day and talks & posters/demos
7:45-8:45	Invited talk: Large Scale Image Annotation: Learning to Rank with Joint Word-Image Samy Bengio
8:45-8:55	Contributed Challenge: Gesture Recognition Competition Isabelle Guyon, Vassilis Athitsos, Jitendra Malik, Ivan Laptev
8:55-9:10	Poster / Demo *Spotlights*
9:10-9:45	Break - Poster/Demo/Challenge session
9:45-10:15	Contributed talk: Sifting through Images with Multinomial Relevance Feedback Dorota Glowacka, Alan Medlar, John Shawe-Taylor
10:15-10:45	Food for thought: Short discussion on state-of-the-art and current challenges.
15:30-16:30	Invited talk: Visual recognition: ML methods for handling hidden structure, high dimensionality and large-scale data Fei-Fei Li
16:30-17:00	Contributed talk: Toward Artificial Synesthesia Linking Images and Sounds via Words Han Xiao, Thomas Stibor
17:00-17:30	Break & Poster / demo session
17:00-17:30	Coffee Break
17:30-18:00	Invited talk: No Hype, All Hallelujah: Structured Models in Computer Vision Sebastian Nowozin
18:00-18:30	Close Panel discussion / thoughts / feedback : Challenges in Computer Vision, opportunities for ML to make a ‘quantum leap’ impact

INVITED SPEAKERS

Scale Image Annotation: Learning to Rank with Joint Word-Image

Samy Bengio, Google

Image annotation datasets are becoming larger and larger, with tens of millions of images and tens of thousands of possible annotations. We propose a strongly performing method that scales to such datasets by simultaneously learning to optimize precision at k of the ranked list of annotations for a given image and learning a low-dimensional joint embedding space for both images and annotations. Our method both outperforms several baseline methods and, in comparison to them, is faster and consumes less memory. We also demonstrate how our method learns an interpretable model, where annotations with alternate spellings or even languages are close in the embedding space. Hence, even when our model does not predict the exact annotation given by a human labeler, it often predicts similar annotations, a fact that we try to quantify by measuring the newly introduced ‘‘sibling’’ precision metric, where our method also obtains excellent results (this is joint work with Jason Weston and Nicolas Usunier). The talk will finish with my own views of the upcoming challenges for the machine learning for computer vision community.

Visual recognition: ML methods for handling hidden structure, high dimensionality and large-scale data

Fei-Fei Li, Stanford

Vision Lab Understanding the meaning and structure of images is a central topic in computer vision. In this talk, I present a number of recent projects from the Stanford Vision Lab on tackling the problem of scene understanding, object recognition and human-object interactions. Each of these visual recognition tasks highlights an aspect of the challenges in vision: inferring hidden structure, handling high dimensionality, and classifying tens of thousands of objects in a data ontology of tens of millions of samples. We show how we use structural learning, sparsification and efficient metric learning methods in both generative and discriminative frameworks to approach these problems. If time permits, I will also discuss challenges and opportunities of ML research in vision now and in the future.

No Hype, All Hallelujah: Structured Models in Computer Vision

Sebastian Nowozin, Microsoft Research

Rich statistical models have revolutionized computer vision research: graphical models and structured prediction in particular are now commonly used tools to address hard computer vision problems. I discuss what distinguishes these computer vision problems from other machine learning problems and how this poses unique challenges. One current line of research to address these is in enriching the model structure: the use of latent variables, hierarchical and deep architectures, higher-order interactions, and structure learning. All suffer from the limitations of todays estimation and learning methods and make clear the need for alternatives.Towards this goal I will discuss current parameter estimation methods for discrete graphical models, exposing conceptual flaws of popularly used methods and advantages of less well-known estimators. I will conclude with a positive outlook on the future.

ACCEPTED CONTRIBUTIONS

Sifting through Images with Multinomial Relevance Feedback

Dorota Glowacka, University College London
Alan Medlar, University College London
John Shawe-Taylor, University College London

This paper presents the theory, design principles, implementation and performance results of a content-based image retrieval system based on multinomial relevance feedback. The system relies on an interactive search paradigm in which at each round a user is presented with a set of k images and is required to selectone that is closest to their target. Performance is measured by the number of rounds needed to identify a specific target image as well as the the average distance from the target of the set of k images presented to the user at each iteration. Results of experiments involving simulations as well as real users are presented. The conjugate prior Dirichlet distribution is used to model the problem motivating an algorithm that trades exploration and exploitation in presenting the images in each round. A sparse data representation makes the algorithm scalable. Experimental results show that the new approach compares favourably with previous work.

Toward Artificial Synesthesia Linking Images and Sounds via Words

Han Xiao and Thomas Stibor
Technische Universitunchen at Műnchen

We tackle a new challenge of modeling a perceptual experience in which a stimulus in one modality gives rise to an experience in a different sensory modality, termed synesthesia. To meet the challenge, we propose a probabilistic framework based on graphical models that enables to link visual modalities and auditory modalities via natural language text. Experimental results indicate usefulness and applicability of the framework, and may lead to a better understanding of human’s synesthesia experiences.

Gesture Recognition Competition

Isabelle Guyon, ClopiNet
Vassilis Athitsos, University of Texas at Arlington
Jitendra Malik, University of California at Berkeley
Ivan Laptev, INRIA

We are organizing a competition on gesture recognition. This challenge is part of a series of challenges on the theme of unsupervised and transfer learning. The goal is to push the state of the art in algorithms capable of learning data representations, which may be re-used from task to task, using unlabeled data and/or labeled data from similar domains. In this challenge, the competitors will obtain a large database of videos of American Sign Language performed by native signers. They will be tested in a live competition on gesture recognition in which the systems will be demonstrated at the site of a conference. The test will be carried on a small but new sign language vocabulary.

Challenges of Data Visualization

http://cseweb.ucsd.edu/~lvdmaaten/workshops/nips2010/

LOCATION
Westin: Alpine DE
Saturday, 07:30 - 10:30 and 3:30 - 6:30

Barbara Hammer hammer@in.tu-clausthal.de
University of Bielefeld

Laurens van der Maaten lvdmaaten@gmail.com
UC San Diego / Delft University of Technology

Fei Sha feisha@usc.edu
University of Southern California

Alexander J. Smola alex@smola.org
Yahoo! Research

ABSTRACT

The increasing amount and complexity of electronic data sets turns visualization into a key technology to provide an intuitive interface to the information. Unsupervised learning has developed powerful techniques for, e.g., manifold learning, dimensionality reduction, collaborative filtering, and topic modeling. However, the field has so far not fully appreciated the problems that data analysts seeking to apply unsupervised learning to information visualization are facing such as heterogeneous and context dependent objectives or streaming and distributed data with different credibility. Moreover, the unsupervised learning field has hitherto failed to develop human-in-the-loop approaches to data visualization, even though such approaches including e.g. user relevance feedback are necessary to arrive at valid and interesting results.

As a consequence, a number of challenges arise in the context of data visualization which cannot be solved by classical methods in the field:

- *Methods have to deal with modern data formats and data sets: How can the technologies be adapted to deal with streaming and probably non i.i.d. data sets?*
  How can specific data formats be visualized appropriately such as spatio-temporal data, spectral data, data characterized by a general probably non-metric dissimilarity measure, etc.? How can we deal with heterogeneous data and different credibility? How can the dissimilarity measure be adapted to emphasize the aspects which are relevant for visualization?

- *Available techniques for specific tasks should be combined in a canonic way:*
  How can unsupervised learning techniques be combined to construct good visualizations? For instance, how can we effectively combine techniques for clustering, collaborative filtering, and topic modeling with dimensionality reduction to construct scatter plots that reveal the similarity between groups of data, movies, or documents? How can we arrive at context dependent visualization?

SCHEDULE

7:30-7:40	Opening remarks
7:40-8:10	Data Triangulation, Information Visualization and Ethnomining: Understanding Everyday Interactions with Technologies E. Churchill
8:10-8:30	Optimal Embedding of Heterogeneous Data with Edge Crossing Constraints A. Shabbeer, C. Ozcaglar, M. Gonzalez, K.P. Bennett
8:30-9:00	Information Visualization and Retrieval S. Kaski
9:00-9:20	Coffee Break
9:20-9:30	Demo: The Topic Browser An Interactive Tool for Browsing Topic Models M. Gardner, J. Lund, D. Walker, J. Lutes, J. Hansen, E. Ringger, K. Seppi
9:30-10:00	Challenges in Data Visualization R. Kosara
10:00-10:30	Bayesian Inference with Kernels A. Gretton
10:30-3:30	Break
3:30-4:00	Visual Analytics for Relational Data L. Getoor
4:00-4:20	A Psychophysical Investigation of Dimensionality Reduction J.M. Lewis, L. van der Maaten, V. R. de Sa
4:20-4:30	Demo: Inside the Selection Box: Visualising active learning selection strategies B. Mac Namee, R. Hu, and S. J. Delany
4:30-4:00	Discussion
5:00-5:40	Coffee Break and Poster Session
5:40-6:00	Probabilistic Spectral Dimensionality Reduction N.D. Lawrence
6:00-6:30	Kernelized Sorting for Similarity Visualization: Globby Search Engine N. Quadrianto, A. Smola, K. Kersting, K. Gawande, W. Buntine, L. Song, T. Tuytelaars
6:30	Closing

- *Visualization techniques should be accompanied by theoretical guarantees:*
  What are reasonable mathematical specifications of data visualization to shape this inherently ill-posed problem? Can this be controlled by the user in an efficient way? How can visualization be evaluated? What are reasonable benchmarks? What are reasonable evaluation measures?

- *Visualization techniques should be ready to use for users outside the field:*
  Which methods are suited to users outside the field? How can the necessity be avoided to set specific technical parameters by hand or choose from different possible mathematical algorithms by hand? Can this necessity be substituted by intuitive interactive mechanisms which can be used by non-experts?

The goal of the workshop is to identify the state-of-the-art with respect to these challenges and to discuss possibilities to meet these demands with modern techniques.

INVITED SPEAKERS

Data Triangulation, Information Visualization and Ethnomining: Understanding Everyday Interactions with Technologies

Elizabeth Churchill, Yahoo! Research

As networked technologies increasingly become part of our everyday lives, more and more data about their usage are being collected. Privacy issues notwithstanding, these data provide an excellent way to develop models of how, when, where and why people use networked technologies, services and applications. By taking what has been dubbed an ethnomining" approach -that is datamining with an ethnographic lens -I suggest there a excellent opportunities for a highly productive collaboration between datamining, information visualization and social science. In this talk, I will walk through the kinds of research questions that are posed by social scientists and ethnographers, the kinds of research methods that are used, the kinds of data that are generated. I will conclude with a vision for a fruitful collaboration between dataminers, information visualization experts and social scientists interested in building socio technical models of human behaviour. I will illustrate with case studies and examples.
Visual Analytics for Relational Data Lise Getoor, University of Maryland, College Park

Bayesian Inference with Kernels

Arthur Gretton, University College London

An embedding of probability distributions into a reproducing kernel Hilbert space (RKHS) has been introduced: like the characteristic function, this provides a unique representation of a probability distribution in a high dimensional feature space. This representation forms the basis of an inference procedure on graphical models, where the likelihoods are represented as RKHS functions. The resulting algorithm is completely nonparametric: all aspects of the model are represented implicitly, and learned from a training sample. Both exact inference on trees and loopy BP on pairwise Markov random fields are demonstrated. Kernel message passing can be applied to general domains where kernels are defined, handling challenging cases such as discrete variables with huge domains, or very complex, non-Gaussian continuous distributions. We apply kernel message passing and competing approaches to cross-language document retrieval, depth prediction from still images, protein configuration prediction, and paper topic inference from citation networks: these are all large-scale problems, with continuous-valued or structured random variables having complex underlying probability distributions. In all cases, kernel BP performs outstandingly, being orders of magnitude faster than state-of-the-art nonparametric alternatives, and returning more accurate results.

Joint work with Danny Bickson, Kenji Fukumizu, Carlos Guestrin, Yucheng Low, Le Song

Information Visualization and Retrieval

Samuel Kaski, Aalto University School of Science and Technology

Challenges in Information Visualization

Robert Kosara, UNC Charlotte

While information visualization (infovis) is developing a lot of useful applications, there is an increasing awareness of its lack of a foundational theory and overarching principles. While we can draw from the theory gathered in statistics, perception, cognition, computing, etc., we need to develop our own set of principles to build future visualization systems on. In this talk, I will describe some recent research results that highlight gaps in our knowledge, and discuss our first steps towards building a new body of theory for infovis.

Kernelized Sorting for Similarity Visualization: Globby Search Engine

Novi Quadrianto, Australian National University

Bridging the gap between keyword-based and content-based search engines is the next big thing in enhancing the user experience and serving up more systematic search results. Globby achieves this by returning keyword-based query relevant images in a set of pages, where each page contains several images with content alike images are placed at proximal locations. Globby is based on Kernelized Sorting, a provably locally convergent statistical method, to arrange similar images close to each other in predefined but arbitrary structures with an additional ranking constraint that the most relevant image to the query is placed at, for example, top left corner.

Joint work with Alex Smola, Kristian Kersting, Kishor Gawande, Wray Buntine, Le Song, Tinne Tuytelaars

ACCEPTED CONTRIBUTIONS

Optimal Embedding of Heterogeneous Graph Data with Edge Crossing Constraints

Amina Shabbeer
Cagri Ozcaglar
M. Gonzalez
K. P. Bennett

We propose a novel approach to visualization of heterogeneous data characterized by both a relationship graph structure and intrinsic features. Each data point is a node in a graph with a given structure. Each data point is also associated with a set of features that have a corresponding distance or similarity measure. A successful visualization accurately captures the desired proximity structure as measured by some embedding objective while simultaneously optimizing an aesthetic criterion, no edge crossings. The edge-crossing constraint is expressed as a nonlinear constraint which has an intuitive geometric interpretation closely related to support vector machine classification. The approach can be generalized to remove intersections of general convex polygons including node-edge and node-node intersections. We demonstrate the approach on multi-dimensional scaling or equivalently Kamada-Kawai force-directed graph layout, by modifying the stress majorization algorithm to include penalized edge crossings. The resulting Expectation-Maximization-like algorithm can be readily adapted to other supervised and unsupervised optimization-based embedding or dimensionality reduction methods. The method is demonstrated on a problem in tuberculosis molecular epidemiology -creating spoligo forests for visualizing genetic relatedness between strains of the Mycobacterium tuberculosis complex characterized by a phylogenetic forest, and multiple biomarkers with a corresponding non-metric genetic distance.

Visualization in Low Dimensional Space by Tessellation of Linear Subspaces

Pooyan Khajehpour Tadavani, Babak Alipanahi, Ali Ghodsi

The problem of nonlinear dimensionality reduction is most often formulated as an instance of semidefinite programming (SDP). However, the effectiveness of the SDP-based algorithms is limited by the computational complexity of the SDP solvers. Due to this limitation, some large-scale variations of the SDP-based methods have been proposed, which are relatively fast and scalable; however, they reduce the size of the SDP by some heuristic approximation, which results in sub-optimal solutions. In this paper, we propose a novel method, which significantly reduces the size and the number of constraints of the SDP problem. We show the search space of the SDP problem can be restricted to a small face of the semidefinite cone. Crucially, the proposed formulation is not an approximation and very large SDP problems can be solved exactly and efficiently by this technique. In addition, unlike most methods, this algorithm provides a parametric mapping that allows out-of-sample test points to be sensibly mapped into the embedded space.

Adaptive Visualization of Text Documents Incorporating Domain Knowledge

Axel Soto

We present a method for visualizing text corpora that are assumed to contain labeled and unlabeled documents. Our method aims at learning data mappings of labeled documents including the terms that are most relevant for label discrimination. We can use this information to visualize mapped unlabeled documents as well. We also show how this method allows the inclusion of user's feedback. This feedback is supplied in an iterative process, so that the user can use the output of the method to provide its domain knowledge of the data. At the same time, this technique is well suited for providing a new low-dimensional space where traditional clustering or classification methods can be applied. Even though our approach is able to deal with document labels that are discrete classes, continuous values, or associated vectors, we confine the experiments of this article to labels that represent non-overlapped topics. This approach is evaluated using a set of short and noisy documents, which is considered as a challenging task in the text mining literature.

A Psychophysical Investigation of Dimensionality Reduction

Joshua M. Lewis
Laurens van der Maaten
Virginia R. de Sa

A cornucopia of dimensionality reduction techniques have emerged over the past decade, leaving data analysts with a wide variety of choices for reducing their data. Means of evaluating and comparing low-dimensional embeddings useful for visualization, however, are very limited. When proposing a new technique it is common to simply show rival embeddings side-by-side and let human judgment determine which embedding is superior. This study investigates whether such human embedding evaluations are reliable, i.e., whether humans tend to agree on the quality of an embedding. Our results reveal that, although experts are reasonably consistent in their evaluation of embeddings, novices generally disagree on the quality of an embedding. We discuss the impact of this result on the way dimensionality reduction researchers should present their results, and on applicability of dimensionality reduction outside of machine learning.

The Nystrom approximation for relational generative topographic mappings

Andrej Gisbrecht
Bassam Mokbel
Barbara Hammer

Relational generative topographic mappings (RGTM) provide a statistically motivated data inspection and visualization tool for pairwise dissimilarities by fitting a constraint Gaussian mixture model to the data. Since it is based on pairwise dissimilarities of data, it scales quadratically with the number of training samples, making the method infeasible for large data sets. In this contribution, we transfer the Nystrom approximation to RGTM and we investigate its effect on the method. This leads to a linear method which reliability depends on the intrinsic dimensionality of the dissimilarity matrix.

Inside the Selection Box: Visualising active learning selection strategies
Brian Mac Namee
Rong Hu
Sarah Jane Delany

Visualisations can be used to provide developers with insights into the inner workings of interactive machine learning techniques. In active learning, an inherently interactive machine learning technique, the design of selection strategies is the key research question and this paper demonstrates how spring model based visualisations can be used to provide insight into the precise operation of various selection strategies. Using sample datasets, this paper provides detailed examples of the differences between a range of selection strategies.

Probabilistic Spectral Dimensionality Reduction
Neil D. Lawrence

We introduce a new perspective on spectral dimensionality reduction which views these methods as Gaussian random fields (GRFs). Our unifying perspective is based on the maximum entropy principle which is in turn inspired by maximum variance unfolding. The resulting probabilistic models are based on GRFs. The resulting model is a nonlinear generalization of principal component analysis. We show that parameter fitting in the locally linear embedding is approximate maximum likelihood in these models. We develop new algorithms that directly maximize the likelihood and show that these new algorithms are competitive with the leading spectral approaches on a robot navigation visualization and a human motion capture data set. Finally the maximum likelihood perspective allows us to introduce a new approach to dimensionality reduction based on L1 regularization of the GRF via the graphical lasso.

Fast Optimization for t-SNE
Laurens van der Maaten,

The paper presents an alternative optimization technique for t-SNE that is orders of magnitude faster than the original optimization technique, and that produces results that are at least as good.

The Topic Browser An Interactive Tool for Browsing Topic Models
Matthew J. Gardner
Jeff Lund
Dan Walker
Joshua Lutes
Josh Hansen
Eric Ringger
Kevin Seppi

Topic models have been shown to reveal the semantic content in large corpora. Many individualized visualizations of topic models have been reported in the literature, showing the potential of topic models to give valuable insight into a corpus. However, good, general tools for browsing the entire output of a topic model along with the analyzed corpus have been lacking. We present an interactive tool that incorporates both prior work in displaying topic models as well as some novel ideas that greatly enhance the visualization of these models.

Kick-starting GPLVM Optimization via a Connection to Metric MDS
Sebastian Bitzer
Christopher K. I. Williams

The Gaussian Process Latent Variable Model (GPLVM) is an attractive model for dimensionality reduction, but the optimization of the GPLVM likelihood with respect to the latent point locations is difficult, and prone to local optima. Here we start from the insight that in the GPLVM, we should have that $k(x_i, x_j) \simeq$ sij, where $k(x_i, x_j)$ is the kernel function evaluated at latent points x_i and x_j , and sij is the corresponding estimate from the data. For an isotropic covariance function this relationship can be inverted to yield an estimate of the interpoint distances $\{d_{ij}\}$ in the latent space, and these can be fed into a multidimensional scaling (MDS) algorithm. This yields an initial estimate of the latent locations, which can be subsequently optimized in the usual GPLVM fashion We compare two variants of this approach to the standard PCA initialization and to the ISOMAP algorithm, and show that our initialization converges to the best GPLVM likelihoods on all six tested motion capture data sets.

On the Effect of Clustering on Quality Assessment Measures for Dimensionality Reduction
Bassam Mokbel
Andrej Gisbrecht
Barbara Hammer

Visualization techniques constitute important interfaces between the rapidly increasing digital information available today and the human user. In this context, numerous dimensionality reduction models have been proposed in the literature. This fact recently motivated the development of frameworks for a reliable and automatic quality assessment of the produced results. These specialized measurements represent a means to objectively assess an overall qualitative change under spatial transformation. The rapidly increasing size of the data sets, however, causes the need to not only map given data points to low dimensionality, but to priorily compress the available information, e.g., by means of clustering techniques. While a standard evaluation measure for clustering is given by the quantization error, its counterpart, the reconstruction error, is usually infeasible for dimensionality reduction, such that one has to rely on alternatives. In this paper, we investigate in how far two existing quality measures used for dimensionality reduction are appropriate to evaluate the quality of clustering outputs, such that only one quality measure can be used for the full visualization process in the context of large data sets. In this contribution, we present empirical results from different basic clustering scenarios as a first study to analyze how structural characteristics of the data are reflected in the quality assessment.

Charting Chemical Space: Challenges and Opportunities for AI and Machine Learning

LOCATION
Westin: Glacier
Saturday, 7:30 - 10:30 and 3:30 - 6:30

Pierre F. Baldi pfbaldi@ics.uci.edu
University of California, Irvine

Klaus-Robert Muller Klaus-Robert.Mueller@tu-berlin.de
TU Berlin

Gisbert Schneider gisbert.schneider@pharma.ethz.ch
ETH

ABSTRACT

In spite of its central role and position between physics and biology, chemistry has remained in a somewhat backward state of informatics development compared to its two close relatives, primarily for historical reasons. Computers, open public databases, and large collaborative projects have become the pervasive hallmark of research in physics and biology, but are still at an early stage of development in chemistry. Recently, however, large repositories with millions of small molecules have become freely available, and equally large repositories of chemical reactions have also become available, albeit not freely. These data create a wealth of interesting informatics and machine learning challenges to efficiently store, search, and predict the physical, chemical, and biological properties of small molecules and reactions and chart ‘‘chemical space’’, with significant scientific and technological impacts. Small organic molecules, in particular, with at most a few dozen atoms play a fundamental role in chemistry, biology, biotechnology, and pharmacology. They can be used, for instance, as combinatorial building blocks for chemical synthesis, as molecular probes for perturbing and analyzing biological systems in chemical genomics and systems biology, and for the screening, design, and discovery of new drugs and other useful compounds. Huge arrays of new small molecules can be produced in a relatively short time. Chemoinformatics methods must be able to cope with the inherently graphical, non-vectorial, nature of raw chemical information on small organic molecules and organic reactions, and the vast combinatorial nature of chemical space, containing over 1060 possible small organic molecules. Recently described grand challenges for chemoinformatics include:

(1) overcoming stalled drug discovery;
(2) helping to develop green chemistry and address global warming;
(3) understanding life from a chemical perspective; and
(4) enabling the network of the world´s chemical and biological information to be accessible and interpretable.

This one day workshop will provide a forum to brainstorm about these issues, explore the role and contributions machine learning methods can make to chemistry and chemoinformatics, and hopefully foster new ideas and collaborations.

SCHEDULE

8:35-9:00	The Statistical Significance of Chemical Similarity Scores Ramzi Nasr, UCI, Pierre Baldi, UCI
9:00-9:10	Coffee Break
9:10-9:35	Single versus Multiple Sorting in All Pairs Similarity Search Koji Tsuda, NIAIST, YasuoTabei, JST
9:35-10:00	Predictive Variance of Gaussian Process and Confidence Estimation Katja Hansen, TU Berlin
10:00-10:30	Discussion
15:30-15:35	Welcome back
15:35-16:00	Bottlenecks in Drug Design and Virtual Screening Gisbert Schneider, ETH
16:00-16:25	Analyzing Large-Scale Cheminformatics and Chemogenomics Datasets through Dimension Reduction Judy Qiu, Indiana University [joint work with Geoffrey Fox and David Wild, Indiana University]
16:25-16:50	Machine Learning for In Silico Chemogenomics Jean-Philippe Vert, Mines ParisTech
16:50-17:00	Coffee Break
17:00-17:25	From Machine Learning to Novel Agonists of the Peroxisome Proliferator-Activated Receptor Matthias Rupp, TU Berlin
17:25-17:50	Machine Learning in Drug Design: Lessons Learned Jorg Wichard, Bayer Schering
17:50-18:10	Progress, Challenges, and Opportunities Pierre Baldi, UCI
18:10-18:30	Panel Discussion (moderated by the organizers)
19:00:-22:00	Social Dinner (sponsored by PASCAL)

Machine Learning for Computational Chemistry

Kristin Bennett, Rensselaer Exploratory Center for Cheminformatics Research, Rensselaer Polytechnic Institute Troy, NY

Successful in silico prediction of the bioactivities of small molecule has the potential to greatly increase the rate of discovery of new drugs while decreasing the approximately one billion dollar cost to develop a successful drug. In this talk, we will examine how machine learning has been used in cheminformatics to date. Specifically, we will explore the underlying learning tasks such as predicting the IC50, review the descriptors used to represent molecules in the cheminformatics community, and discuss the issues introduced by high-throughput screening.

Invariant Representations of Molecules via Harmonic Analysis on the Symmetric Group

Risi Kondor, Caltech
Nino Shervashidze and Karsten Borgwardt
Max Planck Inst for Biological Cybernetics, Tubingen

One of the fundamental problems in representing molecules as graphs in learning algorithms is that the representation must be invariant to the numbering of the atoms. In this talk we present a new approach to this age old problem based on the theory of harmonic analysis on the permutation group. We find that harmonic analysis leads to a notion of ‘‘generalized bispectrum’’ of the adjacency matrix that gives us 49 invariants computable in O(n3) time (where n is the number of atoms). A somewhat more involved approach based on finding biologically relevant subgraphs increases the number of invariants at some computational expense. Overall, we find that in several classification tasks of organic molecules, our new approach outperforms standard graph kernels that have been used in the machine learning community.

The Statistical Distribution of Chemical Similarity Scores and Its Extreme Values

Ramzi Nasr and Pierre F. Baldi
University of California, Irvine

As repositories of molecules continue to expand and become more open, it becomes increasingly important to develop tools to search them efficiently and assess the statistical significance of chemical similarity scores. In this talk, we will discuss a general framework for understanding, modeling, predicting, and approximating the distribution of chemical similarity scores and its extreme values in large databases. The framework can be applied to different chemical representations and similarity measures but is demonstrated here using the most common binary fingerprints with the Tanimoto similarity measure. We show that the distribution of Tanimoto scores can be approximated by the ratio of two correlated Normal random variables associated with the corresponding unions and intersections. This remains true also when the distribution of similarity scores is conditioned on the size of the query molecules in order to derive more fine-grained results and improve chemical retrieval. The corresponding extreme value distributions for the maximum scores are approximated by Weibull distributions. From these various distributions and their analytical forms, Z-scores, E-values, and p-values are derived to assess the significance of similarity scores. In addition, the framework allows one to predict the value of standard chemical retrieval metrics, such as Sensitivity and Specificity at fixed thresholds, or ROC (Receiver Operating Characteristic) curves at multiple thresholds, and to detect outliers in the form of atypical molecules. Numerous and diverse experiments carried with large sets of molecules from the ChemDB show remarkable agreement between theory and empirical results.

All Pairs Similarity Search for 25 Million Chemical Fingerprints

Koji Tsuda, AIST/JST Yasuo Tabei,

JST Public chemical compound databases such as PubChem are growing very rapidly to contain tens of millions of compounds. Machine learning algorithms such as ISOMAP, semi-supervised learning and spectral clustering require a neighborhood graph which takes $O(n^2)$ time to construct naively. We address the following problem of all pairs similarity search for chemical fingerprints: Given n fingerprints, report all pairs whose Tanimoto coefficient is at least $1-\epsilon$. We propose a fast algorithm termed SketchSort that employs min-wise independent permutation and multiple sorting. Empirical results show that the computational time of our algorithm is much faster than existing Tanimoto-based similarity search methods.

Predictive Variance of Gaussian Process and Confidence Estimation

Katja Hansen, TU Berlin, ETH Zurich

Gaussian processes have been successfully applied in chemoinformatics to predict relevant properties of new compounds, based on previous experimental results (supervised learning). In contrast to many other machine learning approaches, they provide an estimate of variance, or error bar, along with the prediction value itself. This error bar may be used as a confidence estimate in order to discard individual uncertain predictions. The performance gain that arises from the use of these error bars will be illustrated in application studies on lipophilicity and solubility prediction. However, this positive impact is not observed in all applications. We discuss the limits of error bars and compare different methods to assess the quality of confidence estimates, illustrating selected aspects with results from simulation studies.

Bottlenecks in Drug Design and Virtual Screening

Gisbert Schneider, Swiss Federal Institute of Technology (ETH) Zurich, Switzerland

Computer-assisted molecular design and virtual screening have become valuable tools for hit and lead identification in early-phase drug discovery projects. Despite their appeal and sustained success, there are several critical issues that must be properly addressed before further progress can be made. We will present success stories and failures, and discuss which information might be essential for advances in method development, including the accuracy and sparsity of available data, the appropriateness of molecular representation, and consideration of entropic contributions to receptor-ligand interaction. Unquestionably many algorithmic solutions may be adapted from engineering and computer sciences and taken on for virtual screening purposes not to construct 'just another virtual screening tool' but to provide an appropriate mathematical framework that is actually able to capture advanced chemical knowledge. Smart combinations of innovative machine learning approaches and innovative molecular modeling concepts might be suited to help overcome some of the current limitations of computer-assisted drug design.

Analyzing Large-Scale Cheminformatics and Chemogenomics Datasets through Dimension Reduction

Judy Qiu, Geoffrey Fox and David Wild
Indiana University

We have developed a set of dimension reduction algorithms based on MDS (Multidimensional Scaling) and GTM (Generative Topographic Mapping) that can be used to create 2D or 3D projections of large scientific datasets (several hundred thousand points), along with a visualization tool called PlotViz for analyzing these projections. We previously described the application of these methods for visualizing chemical space using PubChem data and structural key fingerprints, with extensions to the Plotviz tool for visualizing chemical structure information [1]. We have now used these methods to create 3D projections of not just the PubChem dataset (representing chemical space), but also PubChem Bioassay and ChEMBL (representing biologically-related space) and DrugBank (representing drug space). Using the Plotviz tool, we are able to compare these datasets and identify regions of chemical space associated with biological activities and drugability. Further, we have used the method to identify activity cliffs, areas of space associated with large changes in biological activities, and have linked the visualizations with our Chem2Bio2RDF system [2] to allow more complex biological relationships (such as gene associations) to be analyzed. Finally, we are developing parallel algorithms and cloud computing platforms to enable these methods to be used on even bigger datasets.

[1] Choi, J.Y. , Bae, S.H., Qiu, J., Fox, G., Chen, B., Wild. D.J. Browsing Large Scale Cheminformatics Data with Dimension Reduction. Emerging Computational Methods for the Life Sciences Workshop, ACM Symposium for High Performance Distributed Computing Jun 21-25, 2010, Chicago IL. Accepted.

[2] Chen, B., Dong. X., Jiao, D., Wang, H., Zhu, Q., Ding, Y., Wild, D.J. Chem2Bio2RDF: a semantic framework for linking and data mining chemogenomic and systems chemical biology data. BMC Bioinformatics 2010, 11, 255.

Machine Learning for In Silico Chemogenomics

Jean-Philippe Vert, Mines ParisTech

From Machine Learning to Novel Agonists of the Peroxisome Proliferator-Activated Receptor

Matthias Rupp, TU, Berlin

Ligand-based virtual screening is an application in cheminformatics that is well-suited for empirical inference algorithms. It is commonly formulated as a supervised regression problem, where the samples are small molecules (ligands), and target labels quantify a desired interaction with a given macromolecular target, such as a receptor, channel, or gene. Typically, i) only few samples are available, ii) samples are not i.i.d. distributed due to the presence of chemical series in the training data, iii) test data often come from a different distribution, iv) false negatives and false positives may have different costs, and v) structured representations of the samples are available. We present a successful inter-disciplinary case study, where we employed machine learning methods adapted to the problem domain to identify novel ligands of the peroxisome proliferator-activated receptor gamma, a nuclear receptor that plays a crucial role in type 2 diabetes and dyslipidemia. Employed techniques included Gaussian processes, leave k clusters out cross-validation, a modified loss function, a graph kernel designed for chemical structure graphs, and multiple kernel learning. The study delivered a selective ligand with a novel cyclobutane scaffold, de-orphanized a natural bioactive product, and hints at the natural product origins of pharmacophore patterns in synthetic ligands.

Machine Learning in Drug Design: Lessons Learned

Jorg Wichard, Bayer Schering

Statistical methods and computational intelligence have become an integral part of the drug design process. Established technologies, such as high throughput techniques for lead structure identification based on the advances in assay miniaturization provide huge amounts of biological data which need to be transformed into valuable information and knowledge by appropriate means. This is where machine learning techniques come into play and this is where we meet our challenges: How to treat unbalanced data sets and biological measurements associated with noise? How accurate can a model be, at all, but how accurate does it need to be ? Can we define a 'domain of applicability' and can we predict the prediction error? We give several examples of the successfull application of machine learning methods along the early stages of drug discovery thereby focusing on particular issues in compound profiling: Metabolic stability, mutagenicity, solubility, and others.

Progress, Challenges, and Opportunities

Pierre F. Baldi and Ramzi Nasr, University of California, Irvine

We will briefly describe a remarkable convergence between chemoinformatics and information retrieval problems and solutions, and recent progress in search algorithms for very large databases of documents or molecules with applications to drug discovery. Beyond drug discovery, we will describe several problems and recent progress in the area of chemical reactions. Finally, we will highlight four grand challenge areas that could provide material for the Discussion session of the workshop.

Coarse-to-Fine Learning and Inference

http://learning.cis.upenn.edu/coarse2fine/

LOCATION
Westin: Alpine DE
Friday, 7:30 - 10:30 and 3:30 - 6:30

Ben Taskar taskar@cis.upenn.edu
University of Pennsylvania

David Weiss djweiss@cis.upenn.edu
University of Pennsylvania

Benjamin J. Sapp bensapp@cis.upenn.edu
University of Pennsylvania

Slav Petrov slav@google.com
Google

ABSTRACT

The bottleneck in many complex prediction problems is the prohibitive cost of inference or search at test time. Examples include structured problems such as object detection and segmentation, natural language parsing and translation, as well as standard classification with kernelized or costly features or a very large number of classes. These problems present a fundamental trade-off between approximation error (bias) and inference or search error due to computational constraints as we consider models of increasing complexity. This trade-off is much less understood than the traditional approximation/estimation (bias/variance) trade-off but is constantly encountered in machine learning applications. The primary aim of this workshop is to formally explore this trade-off and to unify a variety of recent approaches, which can be broadly described as ‘‘coarse-to-fine’’ methods, that explicitly learn to control this trade-off. Unlike approximate inference algorithms, coarse-to-fine methods typically involve exact inference in a coarsened or reduced output space that is then iteratively refined. They have been used with great success in specific applications in computer vision (e.g., face detection) and natural language processing (e.g., parsing, machine translation). However, coarse-to-fine methods have not been studied and formalized as a general machine learning problem. Thus many natural theoretical and empirical questions have remained un-posed; e.g., when will such methods succeed, what is the fundamental theory linking these applications, and what formal guarantees can be found?

In order to begin asking and answering these questions, our workshop will bring together researchers from machine learning, computer vision, and natural language processing who are addressing large-scale prediction problems where inference cost is a major bottleneck. To this end, a significant portion of the workshop will be given over to discussion, in the form of two organized panel discussions and a small poster session. We have taken care to invite speakers who come from each of the research areas mentioned above, and we intend to similarly ensure that the panels are comprised of speakers from multiple communities. Furthermore, because the ‘‘coarse-to-fine’’ label is broadly interpreted across many different fields, we also invite any submission that involves learning to address the bias/computation trade-off or that provides new theoretical insight into this problem. We anticipate that this workshop will lead to concrete new research directions in the analysis and development of coarse-to-fine and other methods that address the bias/computation trade-off, including the establishment of several benchmark problems to allow easier entry by researchers who are not domain experts into this area.

SCHEDULE

7:30-7:40	Opening remarks
7:45-8:25	Cats and Dogs Donald Geman
8:30-9:10	A Cascade Architecture for Object Detection Grammars David McAllester
9:15-9:30	Coffee Break
9:30-10:10	Deep hierarchical inference machines Drew Bagnell
10:15-10:25	Spotlight: Towards a General Framework for Coarse-to-Fine Inference and Learning Chloe Kiddon
10:30-4:00	Steepest Descent
4:00-4:40	Coarse-to-fine Decoding for Parsing and Machine Translation Slav Petrov
4:45-4:55	Spotlight: Coarse to Fine, Cost-Sensitive Classification of Email Jay Pujara
5:00-5:15	Coffee Break
5:15-5:55	Structured Prediction Cascades Ben Taskar
6:00-6:30	Challenge Problems & Conclusions Panel discussion

INVITED SPEAKERS

Cats and Dogs

Donald Geman, Johns Hopkins University

A challenging problem in semantic scene interpretation is the detection and description of deformable objects, such as animals. Common strategies, like bags of features or looping over a partition of pose space with dedicated binary classifiers, fail to provide a rich description, e.g., a complex pose. It is more efficient to design a coarse-to-fine search based on pose-indexed features, nested partitions of the pose space and a hierarchy of classifiers. The hierarchy is then ''folded'' for one-shot learning and ''unfolded'' during scene parsing. These principles, and the utility of rich descriptions, will be illustrated by attacking the Microsoft Assira CAPTCHA, which requires determining which of twelve images contains cats and which contains dogs. Initially searching for cats and dogs together, then applying a cat vs dog detector to the estimated pose, results in high accuracy. This is joint work with Kan Jiang and Francois Fleuret.

A Cascade Architecture for Object Detection Grammars

David McAllester, Toyota Technological Institute

This talk will describe a cascade architecture for the UoCTTI object detector. This detector achieved the highest average precision (averaged over all 20 classes) in each of the 2007, 2008, and 2009 PASCAL VOC object detection challenges (and 4th highest in 2010). The detector can be viewed as a parser for a general form of visual grammar, although the grammars in current use are very shallow. We will describe a general cascade architecture for parsing such visual grammars in which multiple parsing passes are performed with increasingly expensive visual features. Within each pass a placement of a nonterminal may be pruned (eliminate at this stage of the cascade) based on the scores of earlier nonterminals at the same level of feature sophistication as well as nonterminal scores derived from simpler features. This achieves roughly an order of magnitude improvement in detection time with negligible reduction in performance. We also give a generalization theorem bounding the performance reduction due to the cascade pruning.

Deep hierarchical inference machines

J. Andrew Bagnell, Carnegie Mellon University TBA

Coarse-to-fine Decoding for Parsing and Machine Translation Slav Petrov, Google Research State-of-the-art NLP models are anything but compact. Syntactic parsers have huge grammars, machine translation systems have huge transfer tables, and so on across a range of tasks. Exhaustive inference becomes prohibitive with such complexity, requiring efficient approximations to infer optimal structures.

Hierarchical coarse-to-fine methods address this challenge, by exploiting a sequence of models which introduce complexity gradually. At the top of the sequence is a trivial model in which learning and inference are both cheap. Each subsequent model refines the previous one, until a final, full-complexity model is reached. Because each refinement introduces only limited complexity, both learning and inference can be done in an incremental fashion. In this talk, I will describe two latent variable coarse-to-fine systems for syntactic parsing and machine translation.

Structured Prediction Cascades

Ben Taskar, University of Pennsylvania

Structured prediction tasks pose a fundamental bias-computation trade-off: The need for complex models to increase predictive power on the one hand and the limited computational resources for inference in the exponentially-sized output spaces on the other. We formulate and develop structured prediction cascades to address this trade-off: a sequence of increasingly complex models that progressively filter the space of possible outputs. We represent an exponentially large set of filtered outputs using max marginals and propose a novel convex loss for learning cascades that balances filtering error with filtering efficiency. We derive generalization bounds for error and efficiency losses and evaluate our approach on several natural language and vision problems: handwriting recognition, part-of-speech tagging and articulated pose estimation in images and videos. We find that the learned cascades are capable of reducing the complexity of inference by up to several orders of magnitude, and yield significantly higher accuracy by enabling the use of models with higher order dependencies and richer features.

ACCEPTED CONTRIBUTIONS

Towards a General Framework for Coarse-to-Fine Inference and Learning

Jay Pujara, Lise Getoor, University of Maryland

While the number of coarse-to-fine approaches and applications has been steadily growing, no unifying framework has been developed for general coarse-to-fine inference and learning. We present our Coarse-to-Fine Probabilistic Inference (CFPI) framework, a first step towards a general coarse-to-fine inference method, that leverages a given or induced type hierarchy over objects in the domain for efficient, scalable inference. Starting by considering the inference problem at the coarsest type level, our approach performs inference at successively finer grains, pruning low-probability variables before refining. CFPI can be applied with any probabilistic inference method, used in both propositional and relational domains, and applied to many problems in vision, NLP, etc. In this paper, we apply CFPI with the relational lifted belief propagation algorithm to social network link pre-diction and biomolecular event prediction tasks. Experiments in these domains show utilizing our inference framework can greatly improve the scalability of probabilistic inference without sacrificing accuracy. We also show how to learn parameters in a coarse-to-fine manner to maximize the efficiency of CFPI.

Coarse-to-Fine, Cost-Sensitive Classification of E-Mail

Chloe Kiddon, Pedro Domingos, University of Washington

In many real-world scenarios, it is necessary to make judgments at differing levels of granularity due to computational constraints. Particularly when there are a large number of classifications that must be done in a real-time streaming setting and there is a significant difference in the time required to acquire different subsets of features, it is important to have an intelligent strategy for optimizing classification accuracy versus computational costs. Accurate and timely email classification requires trading off the classification granularity with the feature acquisition costs. To solve this problem, we introduce a Granular Cost-Sensitive Classifier (GCSC) which modulates the cost of feature acquisition with the granularity of the clas-sification, allowing inexpensive classification at a coarse level and more costly classification at finer levels of granularity. Our approach can classify messages with greater accuracy while incurring a lower feature acquisition cost relative to baseline classifiers that do not make use of cost information.

Computational Social Science and the Wisdom of Crowds

http://www.cs.umass.edu/~wallach/workshops/nips2010css/

LOCATION
Westin: Westin: Callaghan
Friday, 7:30 - 10:30 and 3:30 - 6:30

Jennifer Wortman Vaughan jenn@cs.ucla.edu
UCLA

Hanna M. Wallach wallach@cs.umass.edu
University of Massachusetts Amherst

ABSTRACT

Computational social science is an emerging academic research area at the intersection of computer science, statistics, and the social sciences, in which quantitative methods and computational tools are used to identify and answer social science questions. The field is driven by new sources of data from the Internet, sensor networks, government databases, crowdsourcing systems, and more, as well as by recent advances in computational modeling, machine learning, statistics, and social network analysis.

The related area of social computing deals with the mechanisms through which people interact with computational systems, examining how and why people contribute to crowdsourcing sites, and the Internet more generally. Examples of social computing systems include prediction markets, reputation systems, and collaborative filtering systems, all designed with the intent of capturing the wisdom of crowds.

Machine learning plays in important role in both of these research areas, but to make truly groundbreaking advances, collaboration is necessary: social scientists and economists are uniquely positioned to identify the most pertinent and vital questions and problems, as well as to provide insight into data generation, while computer scientists contribute significant expertise in developing novel, quantitative methods and tools. To date there have been few in-person venues for researchers in these traditionally disparate areas to interact. This workshop will address this need, with an emphasis on the role of machine learning, making NIPS an ideal venue. We hope to attract a mix of established members of the NIPS community and researchers who have never attended NIPS and will provide an entirely new perspective.

The primary goals of the workshop are to provide an opportunity for attendees to meet, interact, share ideas, establish new collaborations, and to inform the wider NIPS community about current research in computational social science and social computing.

Program Committee: Lars Backstrom (Cornell University), Jordan Boyd-Graber (University of Maryland), Jonathan Chang (Facebook), Sanmay Das (Rensselaer Polytechnic Institute), Ofer Dekel (Microsoft Research), Laura Dietz (Max Planck Institute for Computer Science), Arpita Ghosh (Yahoo! Research), John Horton (Harvard University), Shaili Jain (Yale University), David Jensen (University of Massachusetts, Amherst), Lian Jian (Annenberg School of Communications, University of Southern California), Edith Law (Carnegie Mellon University), David Lazer (Political Science and Computer Science, Northeastern University & Kennedy School of Government, Harvard University), Winter Mason (Yahoo! Research), Andrew McCallum (University of Massachusetts, Amherst), Mary McGlohon (Google), Daniel Ramage (Stanford University), Noah Smith (Carnegie Mellon University), Victoria Stodden (Yale Law School), and Sid Suri (Yahoo! Research).

SCHEDULE

7:30--7:35	Opening remarks
7:35--8:15	Invited talk: Inferring Political Slant from Digging Patterns Paul Resnick
8:15--8:55	Invited talk: Synchronous Experiments on Mechanical Turk Winter Mason
8:55--9:35	First poster session and coffee break
9:35--9:50	Contributed talk: The Ideal Point Topic Model: Predicting Legislative Roll Calls from Text Sean Gerrish and David Blei
9:50--10:30	Invited talk: Computer-Assisted Clustering Methods for Conceptualization Justin Grimmer
10:30--3:30	Ski break and discussion
3:30--3:45	Contributed talk: Preferences in College Applications -- a Nonparametric Bayesian Analysis of Top-10 Rankings Alnur Ali, Thomas Brendan Murphy, Marina Meila, and Harr Chen
3:45--4:25	Invited talk: Markets as a Forecasting Tool Yiling Chen
4:25--4:40	Contributed talk: An Optimization-Based Framework for Combinatorial Prediction Market Design Jacob Abernethy, Yiling Chen, and Jennifer Wortman Vaughan
4:40--5:55	Contributed talk: Combining Human and Machine Intelligence for Making Predictions Yiftach Nagar and Thomas W. Malone
6:55--5:35	Second poster session and coffee break
5:35--6:15	Invited talk: An Economic View of Crowdsourcing and Online Labor Markets John Horton
6:15--6:30	Contributed talk: Time Critical Social Mobilization: The DARPA Network Challenge Winning Strategy Galen Pickard, Iyad Rahwan, Wei Pan, Manuel Cebrian, Riley Crane, Anmol Madan, and Alex (Sandy) Pentland

Please visit the workshop website for up-to-date information about the schedule, including the schedule of posters.

INVITED SPEAKERS

Inferring Political Slant from Digging Patterns

Paul Resnick, University of Michigan School of Information

I will describe a graph random walk algorithm that propagates liberal/conservative classifications of political news articles and users. Using data from a large news aggregator, Digg.com, we view a user's ''digg'' on a news article as an approval vote, and assume that liberal users will digg liberal articles more often, and similarly for conservative users and articles. Starting from a few labeled articles and users, the random walk algorithm propagates political leaning labels to the entire graph. I will compare results of the algorithm to an alternative semi-supervised learning algorithm that also uses the diggs as inputs, and a traditional SVM text classifier. Joint work with Daniel Xiaodan Zhou and Qiaozhu Mei.

Synchronous Experiments on Mechanical Turk

Winter Mason, Yahoo! Research

In this talk I will describe two studies that enabled multiple workers on Mechanical Turk to participate simultaneously. In one study by Siddharth Suri, participants played a public goods game, akin to an n-player ''Prisoner's Dilemma'', in which the individually rational behavior and socially optimal behavior are in conflict. By placing the players in a network, he was able to study the network effects of cooperation and defection. In another study, I had participants play a round-based game in which they are searching a grid for a ''hidden treasure'', sharing their discoveries with their neighbors in a network. I find a similar tradeoff between the socially optimal behavior, exploring for the treasure, and the individually optimal behavior, exploiting previously found locations. I will talk about the results of these studies, as well as how we were able to get large groups of workers---up to 48 players at a time---to participate simultaneously.

Computer-Assisted Clustering Methods for Conceptualization

Justin Grimmer, Political Science Department, Stanford University

I describe two computer-assisted clustering method for the discovery of insightful conceptualizations, in the form of clusterings of input objects. Each of the numerous fully automated methods of cluster analysis proposed in statistics, computer science, and biology optimize a different objective function. Almost all are well defined, but how to determine ex ante which one, if any, will partition a given set of objects in an ''insightful'' or ''useful'' way for a given user is unknown and difficult, if not logically impossible. I introduce two different methods for discovering new concepts. The first method creates a metric space of partitions from a large collection of existing cluster analysis methods applied to a given data set (along with millions of other solutions we add based on combinations of existing clusterings). The second method provides a coarse summary of the space of all possible partitions. Interactive software enables a user to explore and interact with both spaces, and quickly reveal or prompt useful or insightful conceptualizations. In addition, although uncommon in unsupervised learning problems, I offer and implement evaluation designs that make computer-assisted approach vulnerable to being proven suboptimal in specific data types. I demonstrate that our approach facilitates more efficient and insightful discovery of useful information than either expert human coders or many existing fully automated methods. I (will) make available an easy-to-use software package that implements all our suggestions.

Markets as a Forecasting Tool

Yiling Chen, Harvard University School of Engineering and Applied Sciences TBA

An Economic View of Crowdsourcing and Online Labor Markets

John Horton, Harvard University

The emergence of online labor markets promises to make other parts of economics---namely labor economics and behavioral economics---relevant to computer science. In this talk, I will describe the key questions, research methods and results from labor and behavioral economics and the ways I think they apply to online work settings. I will then present some of my own work on labor supply, job search and the determinants of productivity. I will also discuss the role that online labor markets could play as a tool for economic development. Online work can help people in developing countries gain access to first world labor markets, but there are number of challenges---technical, managerial, ethical and political---that could hinder wide¬spread adoption. I believe that computer scientists are well-suited to solve some of these challenges. I will conclude by suggesting a research agenda for computer scientists interested in the economic aspects of crowdsourcing/human computation/Games-with-a-purpose.

CONTRIBUTED TALKS

The Ideal Point Topic Model: Predicting Legislative Roll Calls from Text

Sean Gerrish and David Blei
Computer Science Department, Princeton University

We develop the ideal point topic model, a probabilistic model of legislative text. Our model --drawing on ideas from ideal point estimation and topic modeling --predicts voting patterns based on the contents of bills and the inferred political leanings of legislators. It also provides an exploratory window into how legislative language is correlated with political support. Across 14 years of legislative data, we predict specific voting patterns with high accuracy.

Preferences in college applications --a nonparametric Bayesian analysis of top-10 rankings

Alnur Ali, Microsoft Seattle, WA
Thomas Brendan Murphy, School of Mathematical Sciences, University College Dublin
Marina Meila, Department of Statistics, University of Washington
Harr Chen, CSAIL, MIT

Applicants to degree courses in Irish colleges and universities rank up to ten degree courses from a list of over five hundred. These data provide a wealth of information concerning applicant degree choices. A Dirichlet process mixture of generalized Mallows models are used to explore data from a cohort of applicants. We find strong and diverse clusters, which in turn gains us important insights into the workings of the system. No previously tried models or analysis technique are able to model the data with comparable accuracy.

An Optimization-Based Framework for Combinatorial Prediction Market Design

Jacob Abernethy, EECS, UC Berkeley
Yiling Chen, Harvard University School of Engineering and Applied Sciences
Jennifer Wortman Vaughan, Computer Science Department, UCLA

We build on ideas from online convex optimization to create a general framework for the design of efficient prediction markets over very large outcome spaces.

Combining Human and Machine Intelligence for Making Predictions

Yiftach Nagar and Thomas Malone
Center for Collective Intelligence,
Massachusetts Institute of Technology

Statistical models almost always yield predictions that are more accurate than those of human experts. However, humans are better at data acquisition and at recognizing atypical circumstances. We use prediction markets to combine predictions from groups of humans and artificial-intelligence agents and show that they are more robust than those from groups of humans or agents alone.

Time Critical Social Mobilization: The DARPA Network Challenge Winning Strategy

Galen Pickard, MIT Media Lab and MIT Lincoln Lab
Iyad Rahwan, MIT Media Lab and Masadar Inst. of Sci. & Tech.
Wei Pan, Manuel Cebrian, Riley Crane, Anmoi Madan and Alex (Sandy) Pentland
MIT Media Lab

It is now commonplace to see the Web as a platform that can harness the collective abilities of large numbers of people to accomplish tasks with unprecedented speed, accuracy and scale. To push this idea to its limit, DARPA launched its Network Challenge, which aimed to ‘‘explore the roles the Internet and social networking play in the timely communication, wide-area team-building, and urgent mobilization required to solve broad-scope, time-critical problems’’. The challenge required teams to provide coordinates of ten red weather balloons placed at different locations in the continental United States. This large-scale mobilization required the ability to spread information about the tasks widely and quickly, and to incentivize individuals to act. We report on the winning team's strategy, which utilized a novel recursive incentive mechanism to find all balloons in under nine hours. We analyze the theoretical properties of the mechanism, and present data about its performance in the challenge.

CONTRIBUTED POSTERS

Skill Specialization and the Formation of Collaboration Networks
Katharine Anderson, Tepper School of Business, Carnegie Mellon University

The Interaction Between Supervised Learning and Crowdsourcing
Anthony Brew, Dublin Derek and Padraig Cunningham, School of Computer Science & Informatics, University College Dublin

What Leads to Innovation? An Analysis of Collaborative Problem-Solving
Randy M. Casstevens, Department of Computational Social Science, George Mason University

Seeing a home away from the home: Distilling proto-neighborhoods from incidental data with Latent Topic Modeling
Justin Cranshaw and Tae Yano, Department of Computer Science, Carnegie Mellon University

Consistent Confidence Intervals for Maximum Pseudolikelihood Estimators
Bruce A. Desmarais, Department of Political Science, University of Massachusetts Amherst
Skyler J. Cranmer, Department of Political Science, University of North Carolina at Chapel Hill

Inferring Shared Interests from Social Networks
Laura Dietz, Max-Planck Institute for Computer Science, Saarbrucken

Four factors influencing effectiveness in email communication networks
Ofer Engel, Department of Management, London School of Economics and Political Science

With a Little Help from the Computer: Hybrid Human-Machine Systems on Bandit Problems
Bryan R. Gibson, Kwang-Sung Jun and Xiaojin Zhu, Department of Computer Science, University of Wisconsin-Madison

Identifying Focus, Techniques and Domain of Scientific Papers
Sonal Gupta and Christopher D. Manning, Department of Computer Science, Stanford University

Environment Design in Human Computation
Chien-Ju Ho, UCLA
Yen-Ling Kuo and Jane Yung-Jen Hsu
National Taiwan University

Structure Discovery from Partial Rankings
Jonathan Huang, Carnegie Mellon University Ashish Kapoor, Microsoft Research

Ranking Images on Semantic Attributes using Human Computation
Jeroen H.M. Janssens, Eric O. Postma, Tilburg Center for Cognition and Communication, Tilburg University

Aspect and Sentiment Unification Model
Yohan Jo and Alice Oh
Department of Computer Science, KAIST

A Prediction Market Approach to Learning with Sequential Advice
Sindhu Kutty, Department of Computer Science and Engineering, University of Michigan
Rahul Sami, School of Information, University of Michigan

Measuring Confidence in Temporal Topic Models with Posterior Predictive Checks
David Mimno and David Blei
Department of Computer Science, Princeton University

Measuring Access to Electricity Around the World by Satellite
Brian Min, Department of Political Science, University of Michigan

Learning Navigational Maps by Observing the Movement of Crowds
Simon T. O'Callaghan and Fabio T. Ramos
Australian Centre for Field Robotics, University of Sydney

Crowdsourcing for closed loop control
Sarah Osentoski, Christopher Crick, Grayin Jay and Odest Chadwicke Jenkins
Department of Computer Science, Brown University

Modeling Dynamical Influence in Human Interaction Patterns
Wei Pan, Manuel Cebrian, Wen Dong, Taemie Kim and Alex (Sandy) Pentland
MIT Media Lab

Vuvuzelas & Active Learning for Online Classification
Ulrich Paquet, Jurgen Van Gael, David Stern, Gjergji Kasneci, Ralf Herbrich and Thore Graepel
Microsoft Research

What do you know? A topic-model approach to authority identification
Alexandre Passos and Jacques Wainer, RECOD Lab, University of Campinas, Brazil
Aria Haghighi, Department of Computer Science, University of Massachusetts, Amherst

Which universities lead and lag? Toward university rankings based on scholarly output
Daniel Ramage and Christopher D. Manning
Computer Science Department, Stanford University
Dan McFarland, School of Education, Stanford University

Who Leads Whom: Topical Lead-Lag Analysis across Corpora
Xiaolin Shi, Ramesh Nallapati, Jure Leskovec, Dan McFarland and Dan Jurafsky, Stanford University

Gatekeeping Promotes Cooperation in an Ecology of Games
Paul E. Smaldino, Department of Psychology, University of California Davis,
Mark Lubell, Department of Environmental Science and Policy, University of California Davis

Rigging Tournament Brackets for Weaker Players
Isabelle Stanton and Virginia Vassilevska Williams
UC Berkeley

Data Sharing in Social Science Repositories: Facilitating Reproducible Computational Research
Victoria Stodden, Department of Statistics, Columbia University

Proactive Crowd-Sourcing for Emergency Management
Anu Vaidyanathan, Department of Computer Science and Engineering, Indian Institute of Technology Ropar

Towards Building a High-Quality Workforce with Mechanical Turk
Paul Wais, Shivaram Lingamneni, Duncan Cook, Jason Fennell, Benjamin Goldenberg, Daniel Lubarov, David Martin and Hari Khalsa, Yelp, Inc.

Crowdfunding the next hit: Microfunding online experience goods
Chris Ward and Vandana Ramachandran, Department of Operations and Information Systems, University of Utah

Lawmaking 2.0
Boris Yarmakhov, Department of Social Science,
Nizhny Novgorod, State University Vasiliy Burov, Higher School of Economics, Moscow, Russia
Evgeny Patarakin, State Pedagogic University

Predicting Information Spreading in Twitter
Tauhid R. Zaman, Department of Electrical Engineering and Computer Science, MIT
Ralf Herbrich, Microsoft Research Ltd.
Jurgen Van Gael, Microsoft Research Ltd.
David Stern, Microsoft Research Ltd.

Decision Making with Multiple Imperfect Decision Makers

http://www.utia.cz/NIPSHome

LOCATION
Westin: Hilton: Black Tusk
Friday, 7:30 - 10:30 and 3:30 - 6:30

Miroslav Karny school@utia.cas.cz
Institute of Information Theory and Automation

Tatiana V. Guy guy@ieee.org
Institute of Information Theory and Automation

David H. Wolpert david.h.wolpert@nasa.gov
NASA

ABSTRACT

Prescriptive Bayesian decision making has reached a high level of maturity supported by efficient, theoretically well-founded algorithms. While the long-standing problem of participant's rationality is addressed repeatedly, limited cognitive, acting and evaluative abilities/resources of participants involved have not been considered systematically. This problem of so-called imperfect decision makers emerges repeatedly, for instance, i) consistent theory of incomplete Bayesian games cannot be applied by them; ii) a desirable incorporation of deliberation effort into the design of decision strategies remains unsolved. Societal, biological, engineered systems exhibit paradigms that can extend the scope of existing knowledge in prescriptive decision making. Societal and natural sciences and partially technology have considered imperfection aspects at the descriptive level. In particular, a broadly studied emerging behaviour resulting from descriptive properties of interacting imperfect decision makers can be exploited at prescriptive decision making. The goal of this workshop is to explore such connections between descriptive and prescriptive decision making and stimulate an exchange the results and ideas. The workshop will foster discussion of bounded-rationality and imperfection of decision-makers in light of Nature. We believe that in long-term perspective, the workshop will contribute to solution of the problems:

A. How to formalise rational decision making of an imperfect participant?
B. How to create a feasible prescriptive theory, which respects an imperfect participant?
C. How to extend/modify existing feasible prescriptive theories to imperfect decision-makers?

This topic spans both theoretical issues and the development of effective algorithms and it is closely related to the problem of control under varying/uncertain resources' constraints and to the problem of decision-making cost. The workshop aims to bring together different scientific communities, to brainstorm on possible research directions, and to encourage collaboration among researchers with complementary ideas and expertise. The workshop will be based on invited talks, contributed talks and posters. Extensive moderated and informal discussions ensure targeted exchange.

SCHEDULE

7:30-7:40	Opening remarks
7:40-8:10	Bounded Rationality in Multiagent Systems Using Decentralized Metareasoning Shlomo Zilberstein, Alan Carlin
8:10-8:40	Perfection and Bounded Rationality in the Study Cognition Henry Brighton
8:40-9:10	Invariant neuronal activity associated to decision making in a rewarded choice reaction time task Alessandro E. P. Villa
9:10-9:40	Coffee Break
9:40-10:10	Ambiguity aversion and the role of information in decision making and planning Naftali Tishby
10:10-10:40	Sharing of knowledge and preferences among imperfect Bayesian decision makers Miroslav Karny and Tatiana V.Guy
10:40-10:50	Poster spotlights
10:50-11:30	Poster session
3:30-4:00	Using Game Theoretic Concepts to Improve Human Pilot Modeling During Mid-Air Encounters Ritchie Lee and David H. Wolpert
4:00-4:30	The Social Ultimatum Game Yu-Han Chang, Tomer Levinboim, Rajiv Maheswaran
4:30-5:00	Inverse Correlated Equilibrium for Matrix Games Kevin Waugh, Brian D. Ziebart and J. Andrew Bagnell
5:00-5:30	Coffee Break
5:30-6:00	Dynamic Model Averaging for Combining Inferences from Competing Statistical Models Adrian E. Raftery
6:00-6:30	Panel discussion

INVITED SPEAKERS

Bounded Rationality in Multiagent Systems Using Decentralized Metareasoning

Shlomo Zilberstein, University of Massachusetts
Alan Carlin, University of Massachusetts

Metareasoning has been used as a means for achieving bounded rationality by optimizing the tradeoff between the cost and value of the decision making process. Effective monitoring techniques have been developed to allow agents to stop their computation at the “right” time so as to optimize the overall time-dependent utility of the decision. However, these methods were designed for a single decision maker. In this paper, we analyze the problems that arise when several agents solve components of a larger problem, each using an anytime algorithm. Metareasoning is more challenging in this case because each agent is uncertain about the progress made so far by the others. We develop a formal framework for decentralized monitoring of decision making, establish the complexity of several interesting variants of the problem, and propose solution techniques for each case.

Invariant Neuronal Activity Associated to Decision Making in a Rewarded Choice Reaction Time Task

Alessandro E. P. Villa, University of Lausanne

Freely-moving rats were trained in a decision-making reaction time task to provide data that can be compared with noise-compatibility paradigms previously obtained in humans. A group of subjects was trained at first to positively discriminate an auditory pitch in a rewarded Go/Nogo response choice task. In a subsequent phase the same tones were simultaneously presented in different combinations from two locations, such that only the correct tone presented at the correct location is triggering a reward. Other subjects were trained either to discriminate at first the location cue, whereas the pitch cue was introduced in the subsequent phase or to categorize human vowels. At the end of the second phase the rats were chronically recorded with multiple electrodes located in the auditory and inferolimbic cerebral cortical areas. Invariant preferred firing sequences both within, and across cell spike trains tended to appear in association with the response predicted by the subject, as suggested by faster reaction times, or in association with specific errors of decision.

Ambiguity Aversion and the Role of Information in Decision Making and Planning

Naftali Tishby, The Hebrew University of Jerusalem

One of the most striking characterizations of life is the ability to efficiently extract information -through sensory perception, and exploit it -through behavior. There is a growing empirical evidence that information seeking is as important for optimal behavior as reward seeking. Yet our basic algorithms for describing planning and behavior, in particular reinforcement learning (RL), so far ignored this component. In this talk I will describe new extensions of reinforcement learning that combine information seeking and reward seeking behaviors in a principle optimal way. I will argue that Shannon’s information measures provide the only consistent way for trading information with expected future reward and show how the two can be naturally combined in the frameworks of Markov-Decision-Processes (MDP) and Dynamic Programming (DP). This new framework unifies techniques from information theory (like the Huffman source coding algorithm) with methods of optimal control (like the Bellman equation). We show that the resulting optimization problem has a unique global minimum and convergence (even that it lacks convexity). Moreover, the tradeoff between information and value is shown to be robust to fluctuations in the reward values by using the PAC-Bayes generalization bound, providing another interesting justification to its biological relevance.

Sharing of Knowledge and Preferences among Imperfect Bayesian Decision Makers

Miroslav Karny and Tatiana V. Guy
Institute of Information Theory & Automation, Prague

Bayesian decision theory provides a strong theoretical basis for a single-participant decision making under uncertainty, that can be extended to multi-participant problems. However Bayesian decision theory assumes unlimited abilities of a participant to probabilistically model participant’s environment and to optimise decision-making strategy. The paper proposes a novel methodology for sharing of knowledge and strategies among participants, that helps to overcome the non-realistic assumption on participants’ unlimited abilities.

Using Game Theoretic Concepts to Improve Human Pilot Modeling During Mid-Air Encounters

Ritchie Lee, Carnegie Mellon University
David H. Wolpert, NASA

This paper proposes a novel modeling methodology called the “Network-Form Game” that provides a framework for modeling human behavior in a probabilistic manner. We apply this methodology to explicitly and probabilistically model pilots’ responses during mid-air encounters – something that is not considered in current systems. In this framework, pilots are modeled as nodes in a Bayesian Network that defines their interaction with the environment in the context of the problem. Pilot behavior is modeled using a novel game theoretic concept combining Level-k Thinking and Satisficing, with the pilot’s response ultimately decided by his/her utility function. Statistical methods for sampling the Network-Form Game were explored to predict outcome distributions of the system. Finally, we demonstrate the power of this methodology by performing trade analyses on various parameters.

Dynamic Model Averaging for Combining Inferences from Competing Statistical Models
Adrian E. Raftery, University of Washington

Bayesian model averaging (BMA) is the principled Bayesian approach to combining inferences from competing statistical models. We extend BMA to the dynamic situation where predictions are updated as observations arrive and there is model uncertainty. In Dynamic Model Averaging (DMA) a state space model for the parameters of each model is combined with a Markov chain for the generating model, which changes over time. The state space and Markov chain models are both specified in terms of forgetting, giving parsimony. We extend the method to binary outcomes by changing the observation equation from a normal linear regression model to a logistic regression model. We will describe applications to online prediction of the output of a cold rolling mill, intended as part of an adaptive control system, and to data from children with appendicitis who receive either a traditional (open) appendectomy or a laparoscopic procedure. This is based on joint work with Miroslav Karny, Pavel Ettler, Tyler H. McCormick, David Madigan and Randall S. Burd.

ACCEPTED CONTRIBUTIONS

Perfection and Bounded Rationality in the Study Cognition
Henry Brighton, Max Planck Institute for Human Development

Cognition rests on an ability to make accurate inferences from limited observations of an uncertain and potentially changing environment. The cognitive system is often regarded as an existence proof of a computational device which is remarkably well adapted to the problem of dealing with uncertainty. In this talk I will contrast two approaches used to examine, and characterize, cognition from a functional perspective: the rational analysis of cognition, and the study of bounded rationality. Although these two approaches are allied in important respects, they can differ in the extent to which the very idea of perfection, and notions of rational action, should be informed by the constraints of computation. The rational analysis of cognition develops rational, ideal observer models used to judge human responses. To the extent the ideal observer model describes human responses, the rational model can constrain which process level theories are capable of explaining how humans compute. In addition, such findings are typically used to label cognition as more or less rational. The study of bounded rationality has several, often confusing definitions. Here, I will discuss the study of simple cognitive heuristics, which focuses on the essence of Herbert Simon's original conception of the idea. This approach asks how humans and animals function when optimization is not possible. Rather than maximize, minimize, or optimize, boundedly rational actors seek good enough solutions which can nevertheless perform extremely well under conditions of uncertainty. Differences between these two approaches bring into focus fundamental questions. Should resource-bounded organisms and unbounded fictional observers be judged in the same terms? Is the notion of optimality required to discuss rational action? More concretely, in what ways might boundedly rational agents differ from rational agents?

The Social Ultimatum Game
Yu-Han Chang, Tomer Levinboim and Rajiv Maheswaran
University of Southern California

The Ultimatum Game is a key exemplar that shows how human play often deviates from "rational" strategies suggested by game-theoretic analysis. One explanation is that humans cannot put aside the assumption of being in a multi-player multiround environment that they are accustomed to in the real world. We introduce the Social Ultimatum Game, a multi-player multi-round version of the classical Ultimatum Game. We develop mathematical models of human play that include "irrational" concepts such as fairness and adaptation to societal expectations. We also investigate the stability of maintaining a society of "fair" agents under these conditions. This work is a first step towards building a general theory of imperfect, but reasonable, human-like strategic play in repeated, multi-agent games.

Inverse Correlated Equilibrium for Matrix Games
Kevin Waugh, Brian D. Ziebart and J. Andrew Bagnell
Carnegie Mellon University

Modeling the joint behavior of multiple imperfect agents from a small number of observations is a difficult, but important task. In the single-agent, decision-theoretic setting, inverse optimal control has been successfully employed. It views observed behavior as an approximately optimal solution to an unknown decision problem, and learns the decision problem's parameters that best explains the observed behavior. In this work, we introduce the Inverse Correlated Equilibrium problem, the multi-agent extension of inverse optimal control to normal-form games. We describe two approaches for this problem and discuss how they enable generalization and prediction of behavior without knowledge of the game's reward function. The first approach solves a simple convex optimization problem, but, unfortunately, often restricts the ability to generalize more than we desire. The second approach is more computationally intensive, but allows for richer extensions of the demonstrated behavior.

Modeling Bounded Rationality of Agents During Interactions

Qing Guo and Piotr Gmytrasiewicz
University of Illinois at Chicago

Frequently, it is advantageous for an agent to model other agents in order to predict their behavior during an interaction. Modeling others as rational has a long tradition in AI and game theory, but modeling other agents' departures from rationality is difficult and controversial. This paper proposes that bounded rationality be modeled as errors the agent being modeled is making while deciding on action. We are motivated by the work on quantal response equilibria in behavioral game theory which uses Nash equilibria as the solution concept. In contrast, we use decision-theoretic maximization of expected utility and apply the models within the framework of interactive POMDPs. Quantal response assumes that a decision maker is rational, i.e., is maximizing his expected utility, but only approximately so, with an error rate characterized by a single error parameter. Another agent's error rate may be unknown and needs to be estimated during an interaction. We show that the error rate of the quantal response can be estimated using Bayesian update of a suitable conjugate prior, and that it has a sufficient statistic under strong simplifying assumptions. However, if the simplifying assumptions are relaxed, the quantal response does not admit a finite sufficient statistic and a more complex update is needed. This confirms the difficulty of using simple models of bounded rationality in general settings

Scalable Negotiation Protocol based on Issue-Grouping for Highly Nonlinear Situation

Katsuhide Fujita, Nagoya Institute of Technology
Takayuki Ito, Nagoya Institute of Technology
Mark Klein, MIT Sloan School of Management

Most real-world negotiation involves multiple interdependent issues, which makes an agent's utility functions nonlinear. Traditional negotiation mechanisms, which were designed for linear utilities, do not fare well in nonlinear contexts. One of the main challenges in developing effective nonlinear negotiation protocols is scalability; they can't find a high-quality solution when there are many issues, due to computational intractability. One reasonable approach to reducing computational cost, while maintaining good quality outcomes, is to decompose the utility space into several largely independent sub-spaces. In this paper, we propose a method for decomposing a utility space based on every agent's utility space. By employ-ing the simulated annealing technique based on agents' votes, it is not necessary for the proposed method to reveal private utility information. This method allows good outcomes with greater scalability than the method without issue-grouping.

Supra-Bayesian Approach to Merging of Incomplete and Incompatible Data

Vladimira Seckarova, Institute of Information Theory and Automation, Prague

In practice we often need to take every available information into account. Unfortunately the pieces of information given by different sources are often incomplete (with respect to what we are interested in) and have different forms. In this work we try to solve the problem of treating such data in order to get an optimal merger of them. We present a systematic and unified way how to combine the pieces of information by using a Supra-Bayesian approach and other mathematical tools, e.g. Kerridge inaccuracy, maximum entropy principle. To show how the proposed method works a simple example is given at the end of the work.

Multi-Dimensional Trading Problem in Multi-Participant Settings

Jan Zeman, Institute of Information Theory and Automation, Prague

The futures trading is a challenging problem for both mathematicians and economists. The task is based on price speculation, when a speculator buys a commodity contract, waits for price increase and then earns money by reselling the contract. The paper formulates futures trading on multiple markets as a multi-participant decision-making task.

Deep Learning and Unsupervised Feature Learning

LOCATION
Hilton: Cheakmus
Friday, 7:30 - 10:30 and 3:30 - 6:30

Honglak Lee honglak@eecs.umich.edu
University of Michigan

Marc'Aurelio Ranzato ranzato@cs.toronto.edu
University of Toronto

Yoshua Bengio bengioy@iro.umontreal.ca
University of Montreal

Geoffrey E. Hinton hinton@cs.toronto.edu
University of Toronto

Yann LeCun yann@cs.nyu.edu
New York U

Andrew Y. Ng ang@cs.stanford.edu
Stanford University

ABSTRACT

In recent years, there has been a lot of interest in algorithms that learn feature hierarchies from unlabeled data. Deep learning methods such as deep belief networks, sparse coding-based methods, convolutional networks, and deep Boltzmann machines, have shown promise and have already been successfully applied to a variety of tasks in computer vision, audio processing, natural language processing, information retrieval, and robotics. In this workshop, we will bring together researchers who are interested in deep learning and unsupervised feature learning, review the recent technical progress, discuss the challenges, and identify promising future research directions. Through invited talks, panel discussions and presentations by attendees we will attempt to address some of the most important topics in deep learning today. We will discuss whether and why hierarchical systems are beneficial, what principles should guide the design of objective functions used to train these models, what are the advantages and disadvantages of bottom-up versus top-down approaches, how to design scalable systems, and how deep models can relate to biological systems. Finally, we will try to identify some of the major milestones and goals we would like to achieve during the next 5 or 10 years of research in deep learning.

SCHEDULE

7:30-8:30	Tutorial on deep learning and unsupervised feature learning Workshop organizers
8:30-9:00	Invited talk: Large-Scale Image Recognition via Unsupervised Feature Learning Kai Yu
9:00-9:20	Coffee break
9:20-9:40	Spotlights of poster presentations
9:40-10:10	Invited Talk: Learning Recursive Hierarchical Models of Objects Alan Yuille
10:10-10:30	Morning poster session and group discussions
3:30-4:00	Invited talk: Rich Representations for Learning Visual Recognition Jitendra Malik
4:00-4:50	Panel discussion Drew Bagnell, Jitendra Malik, Tomaso Poggio, Kai Yu, Alan Yuille, Yoshua Bengio, Yann LeCun, Andrew Ng
4:50-5:10	Coffee break
5:10-5:30	Contributed talk 1 TBD
5:30-5:50	Contributed talk 2 TBD
5:50-6:30	Afternoon poster session and group discussions

INVITED SPEAKERS

Large-scale Image Recognition via Unsupervised Feature Learning
Kai Yu, NEC Cupertino

In this talk I will share our experience about large-scale image recognition by using feature learning. We worked on extending sparse coding to a broader family of nonlinear coding methods that explore the geometrical structure of sensory image data. The coding of image local features gives rise to significantly better features, which enables simple linear classifiers to produce stronger results, and also scale much better than nonlinear SVMs using Chi-square or intersection kernels. The methods achieved state-of-the-art results on a range of challenging image classification tasks, including Caltech 101, Caltech 256, PASCAL VOC, and ImageNet.

Learning Recursive Hierarchical Models of Objects
Alan Yuille, UCLA

This talk describes work on learning object models. Objects are represented by recursive compositional models (RCMs) which are constructed from hierarchical dictionaries of more elementary RCMs. These dictionaries are learnt in an unsupervised manner using principles such as suspicious concidence and competitive exclusion. For multiple objects, we learn hierarchical dictionaries which encourage part-sharing (i.e. sharing of elementary RCMs). This gives an efficient representation of multiple objects and yields efficient inference and learning. We deomonstrate results on benchmarked real images. We will discuss similarities and differences between this work and more traditional deep belief networks.

Rich Representations for Learning Visual Recognition
Jitendra Malik, UC Berkeley

In recent years, the problem of visual object recognition has been modeled in the framework of statistical pattern classification, resulting in some striking progress. However for realistic versions of the task, such as object detection in natural scenes as in the PASCAL benchmark, performance numbers of the best systems are still in the the 30-40% range. I believe that if our goal is to model the human visual recognition system, or to design more practically effective computer recognition systems, we need a richer formalism. Just as we should not formulate the child language acquisition problem as one of starting from a set of transcribed sentences with no access to cues such as from phonetics or social communicative context, so also in vision, we need to consider the rich input which children can exploit to acquire their visual vocabulary. In particular, perceptual organization, object tracking, and functional interaction provide very useful scaffolding for the acquisition of visual object categories. I will present some specific results in this general philosophy. Having access to a notion of corresponding keypoints across different exemplars enables us to derive a notion of part, ‘‘poselet’’ which deals with issues such as varying 3d pose, articulation, and occlusion. Combining this with the use of bottom-up grouping gives us a powerful attack on the grand challenge of visual object recognition and segmentation.

Discrete Optimization in Machine Learning: Structures, Algorithms and Applications

http://www.discml.cc/

LOCATION
Hilton: Cheakmus
Saturday, 7:30 - 10:30 and 3:30 - 6:30

Andreas Krause krausea@caltech.edu
California Institute of Technology

Pradeep K. Ravikumar pradeepr@cs.utexas.edu
University of Texas, Austin

Jeff A. Bilmes bilmes@ee.washington.edu
University of Washington

Stefanie Jegelka jegelka@tuebingen.mpg.de
Max Planck Institute for Biological Cybernetics

ABSTRACT

Solving optimization problems with ultimately discrete solutions is becoming increasingly important in machine learning: At the core of statistical machine learning is to infer conclusions from data, and when the variables underlying the data are discrete, both the tasks of inferring the model from data, as well as performing predictions using the estimated model are discrete optimization problems. Many of the resulting optimization problems are NP-hard, and typically, as the problem size increases, standard off-the-shelf optimization procedures become intractable. Fortunately, most discrete optimization problems that arise in machine learning have specific structure, which can be leveraged in order to develop tractable exact or approximate optimization procedures. For example, consider the case of a discrete graphical model over a set of random variables. For the task of prediction, a key structural object is the ''marginal polytope,'' a convex bounded set characterized by the underlying graph of the graphical model. Properties of this polytope, as well as its approximations, have been successfully used to develop efficient algorithms for inference. For the task of model selection, a key structural object is the discrete graph itself.

Another problem structure is sparsity: While estimating a high-dimensional model for regression from a limited amount of data is typically an ill-posed problem, it becomes solvable if it is known that many of the coefficients are zero. Another problem structure, submodularity, a discrete analog of convexity, has been shown to arise in many machine learning problems, including structure learning of probabilistic models, variable selection and clustering. One of the primary goals of this workshop is to investigate how to leverage such structures. There are two major classes of approaches towards solving such discrete optimization problems machine learning: Combinatorial algorithms and continuous relaxations. In the first, the discrete optimization problems are solved directly in the discrete constraint space of the variables.

Typically these take the form of search based procedures, where the discrete structure is exploited to limit the search space. In the other, the discrete problems are transformed into continuous, often tractable convex problems by relaxing the integrality constraints. The exact fractional solutions are then ''rounded'' back to the discrete domain. Another goal of this workshop is to bring researchers in these two communities together in order to discuss (a) tradeoffs and respective benefits of the existing approaches, and (b) problem structures suited to the respective approaches. For instance submodular problems can be tractably solved using combinatorial algorithms; similarly, in certain cases, the continuous relaxations yield discrete solutions that are either exact or with objective within a multiplicative factor of the true solution. In addition to studying discrete structures and algorithms, the workshop will put a particular emphasis on novel applications of discrete optimization in machine learning

Please see workshop webpage (http://www.discml.cc) for the final schedule.

SCHEDULE

7.30-7.40	Opening remarks
7.40-8.15	TBA Tamir Hazan
8.15-9.05	TBA Satoru Iwata
9.05-9.20	Coffee Break
9.20-10.10	Multilinear relaxation: a tool for maximization of submodular functions Jan Vondrak
10.10-10.40	Spotlight presentations
3.10-3.35	Online submodular minimization with combinatorial constraints Stefanie Jegelka
3.35-4.10	Energy Minimization with Label costs and Applications in Multi-Model Fitting Yuri Boykov
4.10-4.45	Dual decomposition for inference in natural language processing Michael Collins
4.45-5.00	Coffee break
5.00-5.35	Information Theoretic Model Validation by Approximate Optimization Joachim Buhmann
5.35-6.10	Taming Information Overload Carlos Guestrin
6.10-6.35	Adaptive Submodularity: A New Approach to Active Learning and Stochastic Optimization Daniel Golovin

INVITED SPEAKERS

TBA
Tamir Hazan, Toyota Technological Institute Chicago

Multilinear relaxation: a tool for maximization of submodular functions
Jan Vondrak, IBM Almaden Research Center

Problems involving maximization of submodular functions arise in many applications, such as combinatorial auctions and coverage optimization in wireless networks. Submodular maximization can be also thought of as a unifying framefork for several classical problems including Max Cut, Max k-Cover and broadcast scheduling. The traditional approaches to maximization of submodular functions are combinatorial, using either greedy or local search techniques. I will describe a new approach, which is analogous to linear programming in the sense that a discrete problem is replaced by a continuous one. In the case of submodular functions, the objective function is replaced by a multilinear polynomial. This objective function is neither convex nor concave, and new techniques are required to handle it. Still, we show that this ‘‘multilinear relaxation’’ provides improved results for a wide range of problems and in several cases leads to an optimal approximation. A particular result I will discuss is an optimal (1-1/e)-approximation for welfare maximization in combinatorial auctions.

TBA
Satoru Iwata, Kyoto University

Energy Minimization with Label costs and Applications in Multi-Model Fitting
Yuri Boykov, University of Western Ontario

The a-expansion algorithm has had a significant impact in computer vision due to its generality, effectiveness, and speed. Until recently, it could only minimize energies that involve unary, pairwise, and specialized higher-order terms. We propose an extension of a-expansion that can simultaneously optimize ‘‘label costs’’ with certain optimality guarantees. An energy with label costs can penalize a solution based on the set of labels that appear in it. The simplest special case is to penalize the number of labels in the solution, but the proposed energy is significantly more general than this. Usefulness of label costs is demonstrated by a number of specific applications in vision (e.g. in object recognition) that appeared in the last year. Our work (see CVPR 2010, IJCV submission) studies label costs from a general perspective, including discussion of multiple algorithms, optimality bounds, extensions, and fast special cases (e.g. UFL heuristics). In this talk we focus on natural generic applications of label costs is multi-model fitting and demonstrate several examples: homography detection, motion segmentation, unsupervised image segmentation, compression, and FMM. We also discuss a method (PEARL) for effective exploration of the continuum of labels -an important practical obstacle for a-expansion in model fitting. We discuss why our optimization-based approach to multi-model fitting is significantly more robust than standard extensions of RANSAC (e.g. sequential RANSAC) currently dominant in vision.

Taming Information Overload
Carlos Guestrin, Carnegie Mellon University

The internet has allowed us to democratize information, but has also brought us a new challenge: Information Overload. With the huge amounts of information out there, it can be very difficult to keep track of what’s most important to us. This problem is not limited to the web: The democratic process can potential benefit from open debate, but it is very difficult for an individual to understand all the facets of an issue. And, in science, with the proliferation of research publications, it is difficult for anyone to stay on top of the developments in their own fields, let alone make the types of connections between fields that can be transformative. In this talk, I will cover three conceptual efforts that together could help tame information overload:

- Richer query models -going beyond simple keyword search.
- Personalization -helping the user adaptively express their information needs.
- Structured outputs -expressing relationships between documents, rather than simply returning a list of documents, allowing the user, for example, to connect the dots between news articles.

Dual decomposition for inference in natural language processing
Michael Collins, Massachusetts Institute of Technology

There has been a long history in combinatorial optimization of methods that exploit structure in complex problems, using methods such as dual decomposition or Lagrangian relaxation. These methods leverage the observation that complex inference problems can often be decomposed into efficiently solvable sub-problems. In this talk I’ll describe work on inference algorithms for NLP based on dual decomposition. I’ll describe work on two problems: 1) Non-projective dependency parsing, where the inference problem is NP-hard for all but the simplest models. 2) Problems that involve ‘‘intersections’’ of two or more dynamic programming problems. These problems are solvable in polynomial time, but in practice the intersected dynamic programs are far too large to be practical. The algorithms are simple and efficient, building on standard combinatorial algorithms (e.g., dynamic programming, minimum spanning tree) as oracles; they provably solve a linear programming relaxation of the original problem; and empirically they very often lead to an exact solution to the original problem. The work is related to recent methods that use dual decomposition as an alternative to belief propagation for inference in Markov random fields. This is joint work with Tommi Jaakkola, Terry Koo, Sasha Rush, and David Sontag.

Information Theoretic Model Validation by Approximate Optimization
Joachim Buhmann, Swiss Federal Institute of Technology Zurich

Model selection in pattern recognition requires (i) to specify a suitable cost function for the data interpretation and (ii) to control the degrees of freedom depending on the noise level in the data. We advocate an information theoretic perspective where the uncertainty in the measurements quantizes the solution space of the underlying optimization problem, thereby adaptively regularizing the cost function. A pattern recognition model, which can tolerate a higher level of fluctuations in the measurements than alternative models, is considered to be superior provided that the solution is equally informative. The optimal tradeoff between ‘‘informativeness’’ and ‘‘robustness’’ is quantified by the approximation capacity of the selected cost function. Empirical evidence for this model selection concept is provided by cluster validation in computer security, i.e., multilabel clustering of Boolean data for role based access control, but also in high dimensional Gaussian mixture models and the analysis of microarray data. Furthermore, the approximation capacity of the SVD cost function suggests a cutoff value for the SVD spectrum.

Online submodular minimization with combinatorial constraints
Stefanie Jegelka, Max Planck Institute for Biological Cybernetics

Recently, there has been rising interest in replacing the linear (modular) cost function in combinatorial problems by a submodular cost function. This enhancement opens the door for powerful models, but at the same time leads to very hard optimization problems. Building on recent progress in submodular minimization with combinatorial constraints, and in online submodular minimization, this talk outlines routes for approximate online submodular minimization with combinatorial constraints. We address this scenario using the example of Minimum Cooperative Cut", whose modular analogue, Minimum Cut, has been important in many applications.

Adaptive Submodularity: A New Approach to Active Learning and Stochastic Optimization
Daniel Golovin, Caltech

Solving stochastic optimization problems under partial observability, where one needs to adaptively make decisions with uncertain outcomes, is a fundamental but notoriously difficult challenge. In this talk, I will introduce a new concept called adaptive submodularity, which generalizes submodular set functions to adaptive policies. In many respects adaptive submodularity plays the same role for adaptive problems as submodularity plays for nonadaptive problems. Specifically, just as many nonadaptive problems with submodular objectives have efficient algorithms with good approximation guarantees, so too do adaptive problems with adaptive submodular objectives. We use this fact to recover and generalize several previous results in adaptive optimization, including results for active learning and adaptive variants of maximum coverage and set cover. Applications include machine diagnosis, observation selection and sensor placement problems, and adaptive viral marketing. Joint work with Andreas Krause.

ACCEPTED CONTRIBUTIONS

Multi-Objective 3-SAT with Survey-Propagation
Cyril Furtlehner, Marc Schoenauer,

Convex message passing and a 3-approximation decoding algorithm for metric facility location
Nevena Lazic, Brendan J. Frey, Parham Aarabi,

Simultaneous Learning and Covering with Adversarial Noise
Andrew Guillory, Jeff A. Bilmes,

Finding modal states in a discrete, sparsely connected, loopy graph by Iterated Cutset Conditioning
Gabi Teodoru, Maneesh Sahani,

A flat histogram method for inference with probabilistic and deterministic constraints
Stefano Ermon, Carla Gomes, Ashish Sabharwal, Bart Selman,

Online algorithms for submodular minimization with combinatorial constraints
Stefanie Jegelka, Jeff A. Bilmes,

An Application of the Submodular Principal Partition to Training Data Subset Selection
Hui Lin, Jeff A. Bilmes,

Deriving Height-Based Bounds by Relaxing the Maximum Satisfiability Problem
Eric Hsu

Learning and Planning from Batch Time Series Data

http://www-personal.umich.edu/~danjl/nips10_workshop/

LOCATION
Hilton: Sutcliffe A
Saturday, 7:30 - 10:30 and 3:30 - 6:30

Daniel J. Lizotte danjl@umich.edu
University of Michigan

Michael Bowling bowling@cs.ualberta.ca
University of Alberta

Susan Murphy samurphy@umich.edu
University of Michigan

Joelle Pineau jpineau@cs.mcgill.ca
McGill University

Sandeep Vijan svijan@umich.edu
University of Michigan

Abstract

Intended Audience: Researchers interested in models and algorithms for learning and planning from batches of time series, including those interested in batch reinforcement learning, dynamic Bayes nets, dynamical systems, and similar topics. Also, researchers interested in any applications where such algorithms and models can be of use, for example in medicine and robotics. Overview: Consider the problem of learning a model or control policy from a batch of trajectories collected a priori that record observations over time. This scenario presents an array of practical challenges. For example, batch data are often noisy and/or partially missing. The data may be high-dimensional because the data collector may not know a priori which observations are useful for decision making. In fact, a data collector may not even have a clear idea of which observations should be used to measure the quality of a policy. Finally, even given low-noise data with a few useful state features and a well-defined objective, the performance of the learner can only be evaluated using the same batch of data that was available for learning. The above challenges encountered in batch learning and planning from time series data are beginning to be addressed by adapting techniques that have proven useful in regression and classification. Careful modelling, filtering, or smoothing could mitigate noisy or missing observations. Appropriate regularization could be used for feature selection. Methods from multi-criterion optimization could be useful for choosing a performance measure. Specialized data re-sampling methods could yield valid assessments of policy performance when gathering new on-policy data is not possible. As applications of reinforcement learning and related methods have become more widespread, practitioners have encountered the above challenges along with many others, and they have begun to develop and adapt a variety of methods from other areas of machine learning and statistics to address these challenges. The goal of our workshop is to further this development by bringing together researchers who are interested in learning and planning methods for batch time series data and researchers who are interested in applying these methods in medicine, robotics, and other relevant domains. Longer term we hope to jump-start synergistic collaborations aimed at improving the quality of learning and planning from training sets of time series for use in medical applications.

SCHEDULE

7:30-7:35	Welcome
7:35-8:15	Beyond function approximators for batch mode reinforcement learning: rebuilding trajectories Damien Ernst
8:20-9:00	Mean-Variance Optimization in Markov Decision Processes Shie Mannor
9:00-9:15	Coffee Break
9:15-10:30	Data Brainstorming: Structure of Health Data at Ann Arbor Veterans Affairs
3:30-3:35	Welcome back
3:35-4:15	Sample complexity of LSTD and Bellman Residual Minimization Remi Munos
4:20-5:00	Learning closed-loop policies from batch data Csaba Szepesvari
5:00-5:15	Coffee Break
5:15-6:30	Poster Session

INVITED SPEAKERS

Beyond function approximators for batch mode reinforcement learning: rebuilding trajectories
Damien Ernst, University of Li`ege

In this talk, I will address the problem of learning a high-performance policy from a set of trajectories for optimal control problems with large state-action spaces. The dominant approach for solving this problem is to use these trajectories to train ‘‘function approximators’’ that either represent the dynamics of the underlying control problem, a (state-action) value function or a policy. These techniques have matured a lot over the last ten years and are now credible means for tackling many real life problems.

In this first part of this talk, I will however argue that a technology based on function approximators may mask relevant information that could be extracted from a set of trajectories. This can have several negative consequences: It may lead to potentially unsafe policies, make performance guarantees difficult to obtain, fail to effectively exploit the information contained in trajectories which are optimal or near-optimal, and offer few clues about the areas of the state space that should be sampled to generate better policies.

Afterwards, I will present a new line of research which is articulated around the rebuilding of ‘‘artificial trajectories’’ from the set of data and show that it opens avenues for designing new algorithms that avoid some of the weaknesses of those relying on function approximators.

Mean-Variance Optimization in Markov Decision Processes
Shie Mannor, Technion, McGill University

We consider Markov decision processes under performance measures that involve both the mean and the variance of the cumulative reward. We show that either randomized or history-based policies can improve performance. It turns out that the complexity of computing a policy that maximizes the mean reward under a variance constraint is NP-hard for some cases, and strongly NP-hard for others. We offer pseudopolynomial exact and approximation algorithms and comment on Monte-Carlo simulation and policy gradient methods for these performance measures.

Sample complexity of LSTD and Bellman Residual Minimization
Remi Munos, INRIA Lille

How many data do we actually need to accurately evaluate the performance of a policy? And how should the data be generated? I will address those questions in light of recent works on two well-known algorithms, namely Least Squares Temporal Differences and Bellman Residual Minimization, and provide approximation error bounds in terms of properties of the function space that is used to approximate the value function and the number of available samples. I will discuss two possible assumptions about the way data are generated: either we possess a generative model and we can obtain samples anywhere in space, or we are constrained to follow trajectories. I will also discuss performance bounds for the two approaches in a policy iteration scheme.

Learning closed-loop policies from batch data
Csaba Szepesvari, University of Alberta

A very common situation in learning to make decisions is when learning must happen given a batch of data. This, obviously puts limits on what can be learned. However, in supervised learning this problem has been dealt with very successfully in the past (in fact, supervised learning researchers are just about to discover the power of interactive learning). The question is: what and what cannot be learned if we want to learn policies? What is the difference to supervised learning? How fast can learning be? What methods to use? Can we use model selection? How? Can we predict the performance of a learnt policy? In this talk I will review previous work on these problems and will enumerate a number of challenges that remain open.

ACCEPTED CONTRIBUTIONS

Batch reinforcement learning for optimizing driving assistance strategies
Olivier Pietquin, Supelec Fabio Tango, Centro Ricerche Fiat

Batch Reinforcement Learning for Spoken Dialogue Systems with Sparse Value Function Approximation
Olivier Pietquin, Supelec

Extracting Motifs from Time Series Generated by Concurrent Activities
Jagannadan Varadarajan, Rémi Emonet and Jean-Marc Odobez
Idiap Research Institute

Kernel-Based Stochastic Factorization for Batch Reinforcement Learning
André M. S. Barreto, Laboratório Nacional de Computacao
Doina Precup, McGill University

Learning Dynamic Models with Latent Confounding Processes
Mark Voortman, Decision Systems Lab, University of Pittsburgh
Denver Dash, Intel Labs Pittsburgh

Learning from a Single Demonstration: Motion Planning with Skill Segmentation
Scott Kuindersma, George Konidaris, Roderic Groupen and Andrew Barto,
University of Massachusetts Amherst, Computer Science Department

Learning RDBNs for Activity Recognition
Cristina Manfredotti, Howard Hamilton and Sandra Zilles
University of Regina

Modeling Latent Pathophysiologic States
Rohit Joshi, Kanak Kshetri, Choong-Hyun Lee and Peter Szolovits
MIT CSAIL

A Perspective on Understanding Infantile Colic
Ansaf Salleb-Aouissi, Axinia Radeva, Rebecca Passonneau and Ashish Tomar, Columbia University Center for Computational Learning Systems
Mary McCord and Harriet McGurk
Columbia UniversityCollege of Physicians and Surgeons
Noémie Elhadad, Columbia University Biomedical Informatics

Reducing Model Bias in Reinforcement Learning
Marc Peter Deisenroth, University of Washington Carl Edward Rasmussen, University of Cambridge

Similarities in resting state and feature-driven activity: Non-parametric evaluation of human fMRI
Jacquelyn A. Shelton, Elvira Fischer and Andreas Bartels
Max Planck Institute for Biological Cybernetics
Matthew B. Blaschko, University of Oxford Arthur Gretton, University College London
Jan Műller, Technische University at Berlin

Learning on Cores, Clusters, and Clouds

http://lccc.eecs.berkeley.edu/

LOCATION
Hilton: Mt Currie South
Saturday, 7:30 - 10:30 and 3:30 - 6:30

Alekh Agarwal alekhagarwal@gmail.com
University of California Berkeley

Lawrence Cayton lcayton@tuebingen.mpg.de
Max Planck Institute for Biological Cybernetics

Ofer Dekel oferd@microsoft.com
Microsoft Research

John Duchi jduchi@cs.berkeley.edu
University of California Berkeley

John Langford jl@hunch.net
Yahoo

Abstract

In the current era of web-scale datasets, high throughput biology and astrophysics, and multilanguage machine translation, modern datasets no longer fit on a single computer and traditional machine learning algorithms often have prohibitively long running times. Parallelized and distributed machine learning is no longer a luxury; it has become a necessity. Moreover, industry leaders have already declared that clouds are the future of computing, and new computing platforms such as Microsoft's Azure and Amazon's EC2 are bringing distributed computing to the masses. The machine learning community has been slow to react to these important trends in computing, and it is time for us to step up to the challenge. While some parallel and distributed machine learning algorithms already exist, many relevant issues are yet to be addressed. Distributed learning algorithms should be robust to node failures and network latencies, and they should be able to exploit the power of asynchronous updates. Some of these issues have been tackled in other fields where distributed computation is more mature, such as convex optimization and numerical linear algebra, and we can learn from their successes and their failures. The workshop aims to draw the attention of machine learning researchers to this rich and emerging area of problems and to establish a community of researchers that are interested in distributed learning. We would like to define a number of common problems for distributed learning (online/batch, synchronous/asynchronous, cloud/cluster/multicore) and to encourage future research that is comparable and compatible. We also hope to expose the learning community to relevant work in fields such as distributed optimization and distributed linear algebra. The day-long workshop aims to identify research problems that are unique to distributed learning. The target audience includes leading researchers from academia and industry that are interested in distributed and large-scale learning.

SCHEDULE

7:30-8:00	Opening remarks
8:00-9:00	Keynote 1 John Tsitsiklis
9:00-9:20	Coffee Break
9:20-9:45	Optimal Distributed Online Prediction Using Mini-Batches Lin Xiao
9:45-10:10	MapReduce/Bigtable for Distributed Optimization Keith Hall
10:10-10:30	Mini talks
3:30-4:30	Keynote: Machine Learning in the Cloud with GraphLab Carlos Guestrin
4:30-4:55	Distributed MAP Inference for Undirected Graphical Models Sameer Singh
4:55-5:15	Coffee Break
5:15-5:40	Gradient Boosted Decision Trees on Hadoop Jerry Ye
5:40-6:00	Mini talks
6:00-6:30	Panel Discussion and Summary

INVITED SPEAKERS

Keynote 1

John N. Tsitsiklis, Massachusetts Institute of Technology

Keynote: Machine Learning in the Cloud with GraphLab

Carlos Guestrin, Carnegie Mellon

Exponentially increasing dataset sizes have driven Machine Learning experts to explore using parallel and distributed computing for their research. Furthermore, cloud computing resources such as Amazon EC2 have become increasingly available, providing cheap and scalable platforms for large scale computation. However, due to the complexities involved in distributed design, it can be difficult for ML researchers to take full advantage of cloud resources. Existing high-level parallel abstractions like MapReduce are insufficiently expressive while low-level tools like MPI and Pthreads leave ML experts repeatedly solving the same design challenges. By targeting common patterns in ML, we developed GraphLab, which compactly expresses asynchronous iterative algorithms with sparse computational dependencies common in ML, while ensuring data consistency and achieving a high degree of parallel performance. We demonstrate the expressiveness of the GraphLab framework by designing and implementing parallel versions for a variety of ML tasks, including learning graphical models with approximate inference, Gibbs sampling, tensor factorization, Co-EM, Lasso and Compressed Sensing. We show that using GraphLab we can achieve excellent parallel performance on large-scale real-world problems and demonstrate their scalability on Amazon EC2, using up to 256 processors.

Optimal Distributed Online Prediction using Mini-Batches

Ofer Dekel, Ran Gilad-Bachrach, Ohad Shamir and Lin Xiao, Microsoft Research

MapReduce/Bigtable for Distributed Optimization

Keith Hall, Google Scott Gilpin, Google Gideon Mann, Google

Distributed MAP Inference for Undirected Graphical Models

Sameer Singh and Andrew McCallum,
University of Massachusetts
Amar Subramanyaz and Fernando Pereira,
Google

Gradient Boosted Decision Trees on Hadoop

Jerry Ye, Jyh-Herng Chow, Jiang Chen and Zhaohui Zheng, Yahoo

Low-rank Methods for Large-scale Machine Learning

http://www.eecs.berkeley.edu/~ameet/low-rank-nips10/

LOCATION
Westin: Alpine BC
Saturday, 7:30 - 10:30 and 3:30 - 6:30

Arthur Gretton arthur.gretton@gmail.com
Gatsby Unit, UCL

Michael Mahoney mahoneymw@gmail.com
Stanford

Mehryar Mohri mohri@google.com
Google Research

Ameet S. Talwalkar ameet@eecs.berkeley.edu
UC Berkeley

Abstract

Abstract
Today's data-driven society is full of large-scale datasets. In the context of machine learning, these datasets are often represented by large matrices representing either a set of real-valued features for each point or pairwise similarities between points. Hence, modern learning problems in computer vision, natural language processing, computational biology, and other areas often face the daunting task of storing and operating on matrices with thousands to millions of entries. An attractive solution to this problem involves working with low-rank approximations of the original matrix. Low-rank approximation is at the core of widely used algorithms such as Principle Component Analysis, Multidimensional Scaling, Latent Semantic Indexing, and manifold learning. Furthermore, low-rank matrices appear in a wide variety of applications including lossy data compression, collaborative filtering, image processing, text analysis, matrix completion and metric learning. In this workshop, we aim to survey recent work on matrix approximation with an emphasis on usefulness for practical large-scale machine learning problems. We aim to provide a forum for researchers to discuss several important questions associated with low-rank approximation techniques.

SCHEDULE

7:30-7:55	Introduction Michael Mahoney
7:55-8:20	Greedy Nystrom Approximation Ahmed K. Farahat, Ali Ghodsi, Mohamed S. Kamel
8:20-9:00	Sublinear Optimization for Machine Learning, and its Relatives Ken Clarkson
9:00-9:40	Coffee Break and Posters
9:40-10:05	Low-rank Methods for Learning Quantum States Stephen Becker, Brielin Brown, Jens Eisert, Steve Flammia, David, Gross, Yi-Kai Liu
10:05-10:30	Online Learning in the Manifold of Low-Rank Matrices Uri Shalit, Daphna Weinshall, Gal Checkik
15:30-16:10	Randomized Algorithms for Low-Rank Approximations and Data Applications Petros Drineas
16:10-16:35	Normalized power iterations for the computation of SVD Per-Gunnar Martinsson, Arthur Szlam, Mark Tygert
16:35-17:00	Spotlight Talks
17:00-17:50	Coffee Break and Posters
17:50-18:30	TBD Emmanuel Candes

INVITED SPEAKERS

TBD

Emmanuel Candes, Stanford University TBD

Sublinear Optimization for Machine Learning, and Its Relatives

Ken Clarkson, IBM Almaden Research Center

In recent work, Elad Hazan, David Woodruff, and I have shown that some classical algorithmic problems of machine learning can be solved approximately, with provable bounds that hold with high probability. Moreover, our algorithms are sublinear, that is, they do not need to touch all the data. Specifically, for a set of points $a_1 \ldots a_n$ in d dimensions, or equivalently an n by d matrix A, we show that finding a d-vector x that approximately maximizes the margin $\min_i a_i \text{ dot } x$ can be done in $O(n + d)/\epsilon^2$ time, up to logarithmic factors, where $\epsilon > 0$ is an additive approximation parameter. We have a similar result for the MEB problem, of finding the Minimum Enclosing Ball containing the input. Our algorithm is a primal-dual version of the classical perceptron training algorithm, in which both the primal and the dual variables are updated using randomization. We also show that our approach extends to kernelized versions of these problems, for some popular kernels. In the course of its operations our algorithm finds a subset of $O\sim(1/\epsilon^2)$ rows of A, and the same number of columns, that determine the solution we find. The rows of A are a *coreset* of the data points, while the column subset amounts to a kind a special-purpose feature selection. Thus a kind of 'low rank' subset of the data matrix determines our solution. I will try to relate these results to prior work on sketching, sampling, and sparsification for data analysis.

Randomized Algorithms for Low-Rank Approximations and Data Applications

Petros Drineas, Rensselear Polytechnic Institute

The introduction of randomization into the design and analysis of algorithms for common matrix problems (such as matrix multiplication, least-squares regression, the Singular Value Decomposition (SVD), etc.) over the last decade has provided a novel paradigm and complementary perspective to traditional numerical linear algebra approaches to matrix computations. In this talk, we will provide an overview of this approach, including how it can be used to approximate problems ranging from matrix multiplication and the SVD of matrices to approximately solving least-squares problems and systems of linear equations. In addition, application of these algorithms to large-scale data analysis will also be discussed.

CONTRIBUTING SPEAKERS

Greedy Nystrom Approximation

Ahmed K. Farahat, University of Waterloo Ali Ghodsi, University of Waterloo Mohamed S. Kamel, University of Waterloo

Low-rank Methods for Learning Quantum States

Stephen Becker, Caltech Brielin Brown, University of Virginia
Jens Eisert, University of Potsdam Steve Flammia, Caltech
David Gross, ETH Zurich
Yi-Kai Liu, UC Berkeley

Online Learning in the Manifold of Low-Rank Matrices

Uri Shalit and Daphna Weinshall,
Hebrew University of Jerusalem
Gal Checkik, Google Research and Bar Ilan University

Normalized power iterations for the computation of SVD

Per-Gunnar Martinsson, University of Colorado Boulder
Arthur Szlamand Mark Tygert,
Courant Institute of Mathematical Sciences

SPOTLIGHT POSTERS

Restricted strong convexity and weighted matrix completion: Optimal bounds with noise

Sahand Negahban, UC Berkeley
Martin Wainwright, UC Berkeley

Column Subset Selection with Missing Data

Laura Balzano, University of Wisconsin Madison
Robert Nowak, University of Wisconsin Madison
Waheed U. Bajwa, Duke University

Grappling with Gigantic Matrices: Fast Approximations based on Distance Geometry

Christian Thurau, Kristian Kersting and Christian Bauckhage, Fraunhofer IAIS

Latent Factor Topic Models with Rank-Reducing Beta Process Priors

John Paisley, Princeton University David Blei, Princeton University

Large Scale GPU Based Inference for the Infinite Relational Model

Toke Jansen Hansen, Morten Mørup and Lars Kai Hansen, Technical University of Denmark

Machine Learning for Assistive Technologies

http://www.cs.uwaterloo.ca/~jhoey/mlat-nips2010/index.html

LOCATION
Westin: Glacier
Friday, 7:30 - 10:30 and 3:30 - 6:30

Jesse Hoey jhoey@cs.uwaterloo.ca
University of Waterloo

Pascal Poupart ppoupart@cs.uwaterloo.ca
University of Waterloo

Thomas Ploetz t.ploetz@newcastle.ac.uk
Newcastle University

Abstract

Abstract
An aging demographic has been identified as a challenge for healthcare provision, with technology tipped to play an increasingly significant role. Already, assistive technologies for cognitive and physical disabilities are being developed at an increasingly rapid rate. However, the use of complex technological solutions by specific and diverse user groups is a significant challenge for universal design. For example, 'smart homes' that recognise inhabitant activities for assessment and assistance have not seen significant uptake by target user groups. The reason for this is primarily that user requirements for this type of technology are very diverse, making a single universal design extremely challenging. Machine learning techniques are therefore playing an increasing role in allowing assistive technologies to be adaptive to persons with diverse needs. However, the ability to adapt to these needs carries a number of theoretical challenges and research directions, including but not limited to decision making under uncertainty, sequence modeling, activity recognition, active learning, hierarchical models, sensor networks, computer vision, preference elicitation, interface design and game theory. This workshop will expose the research area of assistive technology to machine learning specialists, will provide a forum for machine learning researchers and medical/industrial practitioners to brainstorm about the main challenges, and will lead to developments of new research ideas and directions in which machine learning approaches are applied to complex assistive technology problems.

SCHEDULE

7:30-8:15	Keynote Talk: A Decision-Theoretic Framework for Assistive Technologies Prasad Tadepalli
8:15-8:30	On the design and validation of an intelligent powered wheelchair: Lessons from the SmartWheeler project Joelle Pineau
8:30-8:45	Learning the Reward Model of Dialogue POMDPs From Data Hamid R. Chinaei
8:45-9:00	Transfer Learning for the Labelling Problem Gwenn Englebienne
9:00-9:15	Coffee Break
9:15-9:45	Posters I
9:45-10:00	Adaptive Prompting for Intelligent Wheelchairs Pooja Viswanathan
10:00-10:15	Considerations for Developing Intelligent Assistive Technologies for Older Adults with Dementia: Lessons Learned Through Real-world Applications Jennifer Boger
10:15-10:30	The Research Challenges of ClassmateAssist: A Personal and Physical Math Coin Tutoring System Georgios Theocharous
10:30-3:30	Break
3:30-4:15	Keynote Talk: Building Machines for Care Matthai Philipose
4:15-4:30	A Utility and Context Based Framework for Addressing the Machine Learning Challenges of Detecting Rare but Important Clinical Events Holly B. Jimison
4:30-4:45	Robust activity recognition for assistive technologies: Benchmarking ML techniques Ricardo Chavarriaga
4:45-5:00	Automating the Calibration of a Neonatal Condition Monitoring System Chris Williams
5:00-5:15	Coffee Break
5:15-5:45	Posters II
5:45-6:30	group discussion/panel
6:30	close

INVITED SPEAKERS

A Decision-Theoretic Framework for Assistive Technologies

Prasad Tadepalli, School of EECS, Oregon State University

The potential of assistive technologies to transform the lives of both able and disabled people cannot be overestimated. In this talk, we describe a decision-theoretic framework that captures the general problem of optimally assisting a goal-directed user. Since the goals of the users are typically unobserved, a key problem is to infer them from their actions, and balance the uncertainty of the goal with the usefulness of the help offered. We study several instances of this problem as special cases of more general Partially Observable Markov Decision Processes (POMDPs). We apply this framework to a number of domains including the real-world task of folder prediction in Windows, and show that, in spite of the bad worst-case complexity, the performance of myopic heuristics is quite good. We develop a formal model that explains the effectiveness of the myopic heuristics and derive a simple bound on the worst case number of mistakes made relative to an assistant who knows the goals of the user. We suggest open problems and future directions in this line of research to advance the state of the art in assistive technologies. Joint work with Alan Fern.

Building Machines for Care

Matthai Philipose, Intel Corporation

Assisting the old and infirm with day-to-day tasks is currently exhausting, expensive and manual. Many proposals over the last decade have sought to build machines that could help. In this talk, I will use insights from three generations of sensor-based eldercare systems built and deployed by Intel to understand such machines. The core technology is that of measuring and reporting human behavior based on statistical processing of sensor data. A close look at the details of caregiving reveals ‘‘three R’s’’ that such technology must deliver. Recognition must not only be highly reliable, it should preferably apply to a rich variety of behaviors, and do so while providing reassurance of its correctness. I will discuss our efforts and challenges in meeting these goals, with particular focus on sensors (ranging from novel high-density sensing to the use of 3-D cameras), reasoning (ranging across the full spectrum of supervision in learners) and information delivery (ranging from uninterpreted ‘‘witness sequences’’ to providing justification for inference). I will end with a short list of challenge problems that either stand in the way of immediate widespread adoption of these systems or gate dramatic advances in functionality.

ACCEPTED CONTRIBUTIONS

On the design and validation of an intelligent powered wheelchair: Lessons from the SmartWheeler project

Joelle Pineau, McGill University Amin Atrash, McGill University Robert Kaplow, McGill University Julien Villemure, McGill University

New-generation, intelligent, powered wheelchairs promise to increase the mobility and freedom of individuals with serious chronic mobility impairments. And while rapid progress continues to be made in terms of the engineering capabilities of robotic wheelchairs, many projects fall short of the target in terms of ease of use, conviviality, and robustness. This paper describes the SmartWheeler, a multi-functional intelligent wheelchair, which leverages state-of-the-art probabilistic techniques for both autonomous navigation and user interaction modeling, to provide a novel robust solution to the problem of assistive mobility. We also discuss the use of standardized evaluation in the development and testing of such technology.

Learning the Reward Model of Dialogue POMDPs from Data

Abdeslam Boularias, Laval University Quebec Hamid R. Chinaei, Laval University Quebec Brahim Chaib-draa, Laval University Quebec

Spoken language communication between human and machines has become a challenge in research and technology. In particular, enabling the health care robots with spoken language interface is of great attention. Recently, there has been interest for modelling the dialogue manager of spoken dialogue systems using Partially Observable Markov Decision Processes (POMDPs). With the goal of modelling the dialogue manager of health care robots as dialogue POMDPs, we would like to learn the reward model of dialogue POMDPs from expert’s data. In a previous paper work, we used an unsupervised learning method for learning the states, as well as the transition and observation functions of the dialogue POMDPs based on human human dialogues. Continuing our objective of learning the components of dialogue POMDPs from data, we introduce a novel inverse reinforcement learning algorithm for learning the reward function of the dialogue POMDP model. Based on the introduced method, and from an available corpus of data we construct a dialogue POMDP. Then, the learned dialogue policies, based on the learned POMDP, are evaluated. The empirical evaluation shows that the performance of the learned POMDP is higher than expert performance in non, low, and medium noise levels, but the high noise level. At the end, current limita-tions and future directions are addressed

Transfer Learning for the Labelling Problem

G. Englebienne, T. L. M. van Kasteren and B. J. A. Krose, University of Amsterdam

The collection of real-world data is a continuous problem for researchers investigating healthcare systems. In particular, although it is often relatively straight-forward to collect sets of measurements, the corresponding annotation is typically difficult and expensive to obtain. In this paper, we explore how transfer learning can be used to create person-specific probabilistic models for activity recognition when no annotations are available for that person. We evaluate our approach on three large real-world datasets and show that we achieve good classification performance even when little or no labelled data is available.

Adaptive Prompting for Intelligent Wheelchairs

Pooja Viswanathan, University of British Columbia
James Little, University of British Columbia
Alan Mackworth, University of British Columbia

Cognitive impairments prevent older adults from using powered wheelchairs due to safety concerns, thus reducing their mobility and independence. An intelligent powered wheelchair is proposed to help restore mobility, while ensuring safe navigation. Machine vision and learning techniques are used to help prevent collisions with obstacles, as well as provide navigation assistance through adaptive prompts.

Considerations for Developing Intelligent Assistive Technologies for Older Adults with Dementia: Lessons Learned Through Real-world Applications

Jennifer N. Boger, University of Toronto
Tim Adlam, Royal United Hospital Bath
Jasper Snoek, University of Toronto
Alex Mihailidis, University of Toronto

Older adults with dementia represent a rapidly growing demographic that requires much supervision and care. Intelligent assistive technologies hold great promise as a way to support both a person with dementia and his or her caregivers. If these technologies are to be effective, it is imperative that developers have a good understanding of the needs and abilities of this population to ensure interventions are appropriate. This paper presents lessons learned through real-world applications of prototype devices and outlines considerations for developers who are looking to create supportive technologies for people with dementia.

The Research Challenges of ClassmateAssist: A Personal and Physical Math Coin Tutoring System

Georgios Theocharous, Intel Labs
Nicholas Butko, University of California San Diego
Matthai Philipose, Intel Labs

Many elementary mathematics teachers believe that learning improves significantly when students are instructed with physical objects such as coins, called manipulatives. Unfortunately, teaching with manipulatives is a time consuming process that is best with personalized 1-to-1 tutoring. In this paper, we explore the research challenges and solutions of an automated physical and personal tutoring solution.

A Utility and Context Based Framework for Addressing the Machine Learning Challenges of Detecting Rare but Important Clinical Events

Holly Jimison, Oregon Health & Science University
Misha Pavel, Oregon Health & Science University

Designing systems for home use by older adults and those with chronic conditions that impair cognitive or physical function can be especially challenging. Machine learning techniques are critical for categorizing patient state and understanding how best to intervene in an autonomous setting. One of the primary challenges that arises with many clinical applications is that the important events to detect can be quite rare, for example, falls in the home. This means that we often have minimal or no examples at all of the events to use in the training of standard classification approaches. This paper describes the challenges of applying machine learning techniques to home monitoring clinical data, as well as a framework for integrating environmental context and utilities associated with event classes to address the issue of detecting important but rare events.

Robust activity recognition for assistive technologies: Benchmarking ML techniques

Ricardo Chavarriaga, Jose del R. Millan, Hesam Sagha and Hamidreza Bayati,
Ecole Polytechnique Federale de Lausanne
Paul Lukowicz and David Bannach,
University of Passau
Daniel Roggen, Zurich Kilian Forster, Iberto Calatroni and Gerhard Tröster, ETH Zurich
Gerold Hölzl, Alois Ferscha and Marc Kurz,
Johanes Kepler University Linz

An increasing need for healthcare provision and assistive technologies (AT) calls for the development of machine learning techniques able to cope with the variability inherent to real-world deployments. In the particular case of activity recognition applications sensor networks may be prone to changes at different levels ranging from sensor data variability to network reconfiguration. Robust methods are required to deal with those changes providing graceful degradation upon failure or self-configuration and adaptation capabilities that ensure their proper operation for long periods of time. Currently there is a lack of common tools and datasets that allow for replicable and fair comparison of different recognition approaches. We introduce a large database of human daily activities recorded in a sensor-rich environment. The database provides large amount of instances of the recorded activities using a significant number of sensors. In addition, we reviewed some of the techniques that have been proposed to cope with changes in the system, including missing data, sensor location/orientation change, as well as the possibility to exploit data from unknown discovered sensors. These techniques have been tested in the aforementioned datasets showing its suitability to emulate different sensor network configurations and recognition goals.

Automating the Calibration of a Neonatal Condition Monitoring System

Christopher K. I. Williams and Ioan Stanculescu,
University of Edinburgh

Condition monitoring of premature babies in intensive care can be carried out using a Factorial Switching Linear Dynamical System (FSLDS) [19]. A crucial part of training the FSLDS is the manual calibration stage, where an interval of normality must be identified for each baby that is monitored. In this paper we replace this manual step by using a classifier to predict whether an interval is normal or not. We show that the monitoring results obtained using automated calibration are almost as good as those using manual calibration.

Learning mixed acoustic/articulatory models for disabled speech

Frank Rudzicz, University of Toronto

This paper argues that automatic speech recognition (ASR) should accommodate dysarthric speech by incorporating knowledge of the production characteristics of these speakers. We describe the acquisition of a new database of dysarthric speech that includes aligned acoustics and articulatory data obtained by electromagnetic articulography. This database is used to train theoretical and empirical models of the vocal tract within ASR which are compared against discriminative models such as neural networks, support vector machines, and conditional random fields.Results show significant improvements in accuracy over the baseline through the use of production knowledge.

An Uncued Brain-Computer Interface Using Reservoir Computing

Pieter-Jan Kindermans, Pieter Buteneers, David Verstraeten and Benjamin Schrauwen,
Ghent University

Brain-Computer Interfaces are an important and promising avenue for possible next-generation assistive devices. In this article, we show how Reservoir Computing --a computationally efficient way of training recurrent neural networks --combined with a novel feature selection algorithm based on Common Spatial Patterns can be used to drastically improve performance in an uncued motor imagery based Brain-Computer Interface (BCI). The objective of this BCI is to label each sample of EEG data as either motor imagery class 1 (e.g. left hand), motor imagery class 2 (e.g. right hand) or a rest state (i.e., no motor imagery). When comparing the results of the proposed method with the results from the BCI Competition IV (where this dataset was introduced), it turns out that the proposed method outperforms the winner of the competition.

A new model based on task recognition and monitoring for the development of sensory substitution assistive systems for the visually impaired

Pantelis Elinas, Yi Li and Lochana Perera,
University of Sydney

According to the World Health Organization more than 314 million people world-wide suffer from some form of visual impairment with 87% of them living in the developing world. Clearly, there exists a need for the development of cost-effective assistive technologies for improving the quality of life for the visually impaired. Sensory substitution systems aim to replace one sensory signal, e.g., vision, with another, e.g., haptic, delivered via vibrotactile or electrotactile stimulation. Although much progress has been made towards the development of such systems some of which have been made available commercially their capabilities and usability are still far from desired. In this paper, we survey recent advances in sensory substitution with a focus on developing assistive devices for the visually impaired, identify a number of roadblocks to the development of more advanced systems and propose a new development model utilizing state-of-the-art machine learning techniques. We believe that the proposed model will allow us to lift some of the obstacles preventing the mass adaptation of such assistive devices.

Task Assistance for Persons with cognitive Disabilities (TAPeD)

Christian Peters, Thomas Hermann and Sven Wachsmuth,
Bielefeld University

TAPeD is a project at the Cognitive Interaction Technology, Center of Excellence (CITEC) at Bielefeld University, with the aim to develop an automatic prompting system in the healthcare domain. In comparison to systems applied in individual user's homes to prolong the user's independence in everyday life, we aim to develop a system for a residential home where persons with different cognitive disabilities live together and share the same system. The cognitive disabilities include learning disabilities, obsessiveness, epilepsy, behavioral and autistic spectrum disorders. We cooperate with Haus Bersaba, a residential home belonging to v. Bodelschwinghsche Stiftungen Bethel which is a care facility in Bielefeld, Germany. Our user group has problems fulfilling Activities of Daily Living (ADLs), in particular in brushing teeth as a basic ADL. Brushing teeth has a high relevance in everyday life with regard to personal welfare of the inhabitants since disregarding oral hygiene can lead to severe medical problems.

Activity Recognition for Users of Rolling Walker Mobility Aids

Mathieu Sinn and Pascal Poupart
University of Waterloo

We present Smart Walkers, a comprehensive approach to enhancing independent and safe mobility of elderly people. The idea of the Smart Walkers project is to equip rolling walker mobility aids with sensors and actuators. The goal is to assist users, caregivers and clinicians, e.g., by monitoring the physical and mental conditions of the user, detecting risks of falling, assessing the effectiveness of therapeutic interventions, and providing active navigation assistance. The key problem in building the Smart Walkers technology is the ability to recognize the user activity from the stream of sensor measurements. In this paper we present supervised and unsupervised machine learning algorithms for this purpose and discuss their performance on real user data. We find that the best results are obtained for Conditional Random Fields with feature functions based on thresholding, achieving an accuracy of 85-90%.

Probabilistic Cursor Trajectory Prediction via Inverse Optimal Control

Brian D. Ziebart, Anind Dey and J. Andrew Bagnell
Carnegie Mellon University

Many tasks in people's everyday lives can be viewed as control problems where assistive technologies could intervene in various ways to improve task performance. Understanding people's goals and predicting the actions they intend to employ to achieve those goals is important for choosing appropriate interventions. In this paper, we present a novel inverse optimal control approach for learning and predicting those intentions and future actions. We are particularly motivated by assistive technologies for cursor-based input in this work. Under our approach, cursor trajectories are assumed to be stochastically generated from a continuous control process and the parameters of that process that best explain a user's goal-directed cursor motions are learned. The resulting predictions are then efficient for real-time intervention selection and personalized to the individual.

Toward a system for stroke rehabilitation user centred

Roger Luis Velazquez and Enrique Sucar
National Institute for Astrophysics, Optics and Electronics

This paper describes a design of a system rehabilitation games-based therapy, whose goal is monitor the user, provide recommendations on the parameters of activities and adaptation to user performance. We believe that a rehabilitation system coupled to the human behavior is necessary to improve the potential for rehabilitation. To accomplish this, we should consider at least two essential aspects in the design of a recommendation for a robotic rehabilitation system -physical and emotional. We propose a model that takes into account physical aspects and emotions of the user, and can be adapted to the abilities of the patient with the aim of measuring their learning in his rehabilitation tasks.

Enhancing Social Interactions of Individuals with Visual Impairments: A Case Study for Assistive Machine Learning

Vineeth Balasubramanian, Shayok Chakraborty, Sreekar Krishna and Sethuraman Panchanathan,
Arizona State University

Individuals with visual impairments face serious challenges in experiencing the fundamental privileges of social interactions. The realization of a Social Interaction Assistant (SIA) device for such individuals involves solving several challenging problems in pattern analysis and machine intelligence such as person recognition/tracking, head/body pose estimation, posture/ gesture recognition, expression recognition, and human-object interaction recognition on a combination of wearable and ubiquitous computing platforms. This work presents sample machine learning contributions that have been made as part of the development of such a SIA device, including integrated face localization and detection, user-conformal confidence measures and online active learning.

Self Organizing Maps for Affective State Detection

Bert Arnrich, Cornelia Kappeler-Setz, Roberto La Marca, Gerhard Tröster and Ulrike Ehlert,
University Zurich

In this contribution we present two experimental scenarios in which we employed Self Organizing Maps (SOMs) to detect affective states. The first scenario is related towards designing a ''Personal Stress Prevention Assistant'': we summarize our efforts to detect affective information related to stress in the posture channel. We show that a person-independent discrimination of stress from cognitive load is feasible when using data from a pressure mat mounted on a seat. The second scenario is embedded towards assisting patients with manic depression: we present preliminary results in detecting emotions from voice data. Our findings illustrate that a person-dependent discrimination of emotions from voice data seems feasible and that a general model might be appropriate to discriminate high and low levels of arousal.

Robust Local Video Event Detection for Action Recognition

Amir-Hossein Shabani, John Zelek and David A. Clausi,
University of Waterloo

Human action recognition is an important component of elderly activity analysis for assisted living. Actions can be represented using a set of local video events. Using Poisson filtering, a novel robust video event detection approach is devel-oped. This approach is consistent with the human's biological vision and motion perception models. The extracted video events show high precision rate and high reproducibility score under different view perspectives and scale changes. In a standard bag-of-words framework, these events are shown to improve the average accuracy in the recognition of ten different actions in the Weizmann data set.

SNAP: Syndetic Assistance Processes

Jesse Hoey, University of Waterloo
Thomas Ploetz, Dan Jackson, Cuong Pham and Patrick Olivier
Newcastle University
Andrew Monk, York University

Activity recognition in intelligent environments could play a key role for support-ing people in their activities of daily life. Partially observable Markov decision process (POMDP) models have been used successfully, for example, to assist people with dementia when carrying out small multi-step tasks such as hand washing. POMDP models are a powerful, yet flexible framework for modeling assistance that can deal with uncertainty and utility in a theoretically well-justified manner. Unfortunately, POMDPs usually require a very labor intensive, manual setup procedure. This paper describes a knowledge driven method for automatically generating POMDP activity recognition and context sensitive prompting systems for a complex tasks.We call the resulting POMDP a S NAP (SyNdetic Assistance Process). The method starts with a psychologically justified (syndetic) description of the task and the particular environment in which it is to be carried out that can be generated from empirical data. This is then combined with a specification of the available sensors and effectors to build a working prompting system that tracks a person's activities and learns their abilities by using sensor data as evidence in the context of the S NAP POMDP. The method is illustrated by building a system that prompts through the task of making a cup of tea in a real-world kitchen.

Machine Learning for Social Computing

http://mlg.cs.purdue.edu/doku.php?id=mlsc2010

LOCATION
Westin: Alpine A
Saturday, 7:30 - 10:30 and 3:30 - 6:30

Zenglin Xu xu218@purdue.edu
Purdue University

Irwin King king@cse.cuhk.edu.hk
The Chinese University of Hong Kong

Shenghuo Zhu zsh@sv.nec-labs.com
NEC Laboratories America

Yuan (Alan) Qi alanqi@cs.purdue.edu
Purdue university

Rong Yan yanrong@gmail.com
Facebook

John Yen jyen@ist.psu.edu
Penn State University

Abstract

An aging demographic has been identified as a challenge for healthcare provision, with technology tipped to play an increasingly significant role. Already, assistive technologies for cognitive and physical disabilities are being developed at an increasingly rapid rate. However, the use of complex technological solutions by specific and diverse user groups is a significant challenge for universal design. For example, 'smart homes' that recognise inhabitant activities for assessment and assistance have not seen significant uptake by target user groups. The reason for this is primarily that user requirements for this type of technology are very diverse, making a single universal design extremely challenging. Machine learning techniques are therefore playing an increasing role in allowing assistive technologies to be adaptive to persons with diverse needs. However, the ability to adapt to these needs carries a number of theoretical challenges and research directions, including but not limited to decision making under uncertainty, sequence modeling, activity recognition, active learning, hierarchical models, sensor networks, computer vision, preference elicitation, interface design and game theory. This workshop will expose the research area of assistive technology to machine learning specialists, will provide a forum for machine learning researchers and medical/industrial practitioners to brainstorm about the main challenges, and will lead to developments of new research ideas and directions in which machine learning approaches are applied to complex assistive technology problems.

SCHEDULE

7:30-7:35	Opening remarks
7:35-8:20	[Invited talk] Multiscale Community Blockmodel for Network Exploration Eric P. Xing
8:20-8:40	TopicFlow Model: Unsupervised Learning of Topic-specific Influences of Hyperlinked Documents
8:40-9:00	Discovering Demographic Language Variation
9:00-9:25	Coffee Break and poster session
9:25-10:10	[Invited talk] People You May Know: Friend Suggestions on Facebook Lars Backstrom
10:10-10:30	Detecting Latent User Properties in Social Media
15:30-16:15	[Invited talk] Entity Disambiguation and Social Computing Lee Giles
16:15-16:35	Integrating Specialist and Folk Knowledge with Affinity Propagation
16:35-16:55	Context Sensitive Topic Models for Author Influence in a Linked Corpus
16:55-17:30	Coffee Break and poster session
17:30-17:50	Multiview Clustering with Incomplete Views
17:50-18:10	Dynamic NMFs and Temporal Regularization for Online Analysis of Streaming Text
18:10-18:30	General Discussion

INVITED SPEAKERS

Multiscale Community Blockmodel for Network Exploration
Eric P. Xing, Carnegie Mellon University

Real world networks exhibit a complex set of phenomena such as underlying hierarchical organization, multiscale interaction, and varying topologies of communities. Most existing methods do not adequately capture the intrinsic interplay among such phenomena. We propose a nonparametric Multiscale Community Blockmodel (MSCB) to model the generation of hierarchies in social communities, selective membership of actors to subsets of these communities, and the resultant networks due to within- and cross-community interactions. By using the nested Chinese Restaurant Process, our model automatically infers the hierarchy structure from the data. We develop a collapsed gibbs sampling algorithm for posterior inference, conduct extensive validation using synthetic networks, and demonstrate the utility of our model in real-world datasets such as predator-prey networks and citation networks.

Pepole You May Know: Friend Suggestions on Facebook
Lars Backstrom, Facebook

Facebook's friend recommendation system helps people connect with their friends. Our system, called People You May Know, uses a combination of results from sociology and machine learning to make the best suggestions possible. We will look at some of the challenges involved in building a system that can handle the scale of Facebook and provide high quality recommendations. In this talk I will discuss both the algorithmic and machine learning challenges that we have faced and overcome in building this system.

Entity Disambiguation and Social Computing
Lee Giles, Pennsylvania State University TBD.

ACCEPTED CONTRIBUTIONS

TopicFlow Model: Unsupervised Learning of Topic-specific Influences of Hyperlinked Documents
Ramesh Nallapati, Stanford University
Christopher D. Manning, Stanford University

Modeling influence of entities in networked data is an important problem in information retrieval and data mining. Popular algorithms such as PageRank capture this notion of authority by analyzing the hyperlink structure, but they ignore the topical content of the document. However, often times, authority is topic dependent, e.g., a web page of high authority in politics may be an unknown entity in sports. In this work, we describe a new model called TopicFlow, that combines ideas from network flow and topic modeling, and captures the notion of topic specific influences of hyperlinked documents in a completely unsupervised fashion. We show that on the task of citation recommendation, the TopicFlow model, when combined with TF-IDF based cosine similarity, outperforms several competitive baselines by as much as 6.4%. We also present some qualitative visualizations to demonstrate the expressive power of the new model.

Discovering Demographic Language Variation
Brendan O'Connor, Jacob Eisenstein, Eric P. Xing and Noah A. Smith, Carnegie Mellon University

Even within a single language community, speakers from different backgrounds demonstrate sub-stantial linguistic variation. Salient speaker characteristics include geography [9, 6], race [12], and socioeconomic status [8, 4]; they impact language at the phonological, lexical, and morphosyntactic levels [14]. Sociolinguistics and dialectology feature a strong quantitative tradition of studying the relationship between language and social and geographical identity. In general, these approaches begin by identifying both the communities of interest and the relevant linguistic dimensions of vari-ability; for example, a researcher might identify the term "yinz" as characteristic of Pittsburgh di-alect [3], and then model its relationship to the socioeconomic status of the speaker. Thus, while this approach has a quantitative foundation in modeling the relationship between linguistic and extra-linguistic data, it requires extensive fieldwork and linguistic expertise to identify the ‘‘inputs’’ that are to be analyzed. In this paper, we propose a new exploratory methodology for discovering demographic and geo-graphic language variation from text and metadata. We unite these information sources in a Bayesian generative model, which explains both linguistic variation and demographic features through a set of generative distributions, each of which is associated with a (latent) community of speakers. Thus, our model is capable of discovering both the relevant sociolinguistic communities, as well as the key dimensions of linguistic variation.

Detecting Latent User Properties in Social Media
Delip Rao, Johns Hopkins University
David Yarowsky, Johns Hopkins University

The ability to identify user attributes such as gender, age, regional origin, and political orientation solely from user language in social media such as Twitter or similar highly informal content has important applications in advertising, personalization, and recommendation. This paper includes a novel investigation of stacked-SVM-based classification algorithms over a rich set of original features, applied to classifying these four user attributes. We propose new sociolinguistics-based features for classifying user attributes in Twitter-style informal written genres, as distinct from the other primarily spoken genres previously studied in the user-property classification literature. Our models, singly and in ensemble, significantly outperform baseline models in all cases.

Integrating Specialist and Folk Knowledge with Affinity Propagation
Jeon Hyung Kang, USC Information Sciences Institute
Kristina Lerman, USC Information Sciences Institute

Knowledge on the social Web grows each time a user annotates a resource, for example, a Web page, a scientific article, a photo, or a video. While attaching descriptive labels, known as tags, to resources is still the most popular form of annotation, some social Web sites also allow users to create structured annotations.

For example, social bookmarking sites Delicious (http://del.icio.us) and Bibsonomy (http://bibsonomy.org) allow users to specify broader--narrower relations between tags, and the social photosharing site Flickr (http://flickr.com) allows users to organize photos within folder-like hierarchies. While such annotations reflect individual users' needs and requirements for organizing the content they create, collectively social annotations provide valuable evidence for harvesting social knowledge. Folksonomies, or taxonomies of concepts automatically extracted from social annotations of many users, will eventually help us better search for, browse, organize, manage, and integrate information on the Web.

Context Sensitive Topic Models for Author Influence in a Linked Corpus

Saurabh Kataria, Penn State University
Prasenjit Mitra, Penn State University
Lee Giles, Pennsylvania State University

In a document network such as citation network of scientific documents, web-logs, etc., the content produced by authors exhibit their interest in certain topics whereas some authors tend to influence other authors' interests. In this work, we propose to model the influence of cited authors along with the interests of citing authors. Moreover, we hypothesize that apart from the citations present in a documents, the context surrounding the citation mention provides extra topical information about the cited authors. However, associating terms in the context to the cited authors remain an open problem. We propose a novel document generation schemes that incorporate the context while modeling the interests of citing authors and influence of the cited authors simultaneously. Our experiments show significant improvements over baseline models for various evaluation criteria such as link prediction between document and cited author, log-likelihood estimation on unseen text.

Multiview Clustering with Incomplete Views

Anusua Trivedi, University of Utah
Piyush Rai, University of Utah
Hal Daume III, University of Utah
Scott DuVall, University of Utah

Multiview clustering algorithms allow leveraging information from multiple views of the data and therefore lead to improved clustering. Canonical Correlation Analysis (CCA) is one such approach that extracts shared features from multiple views of the data by computing their low-dimensional projections in a shared subspace. These features can then be used with any off-the-shelf clustering algorithm. The CCA based approach however suffers from a shortcoming since it assumes availability of features across all views of each example. Using the kernel variant of CCA, we present an approach that can learn in the setting when features in some view are present only for a small fraction of all the examples. As an example, we apply our method on webpage clustering with multiple views of the data where one view is the page-text and other view is the social tags assigned to the webpage. We consider the case when the tags are available only for a small subset of the webpages which means that the tag view is incomplete. Experimental results establish the effectiveness of the proposed method.

Dynamic NMFs and Temporal Regularization for Online Analysis of Streaming Text

Vikas Sindhwani, IBM Watson Research Center
Ankan Saha, University of Chicago

Learning a dictionary of basis elements with the objective of building compact data representations is a problem of fundamental importance in statistics, machine learning and signal processing. In many settings, data points appear as a stream of high dimensional feature vectors. Streaming datasets present new twists to the problem. On one hand, basis elements need to be dynamically adapted to the statistics of incoming datapoints, while on the other hand, many applications require early detection of rising new trends. The analysis of social media streams formed by tweets and blog posts is a prime example of such a setting, where topics of social discussions need to be continuously tracked and new emerging themes need to be rapidly detected. We formalize such problems in terms of online learning of dynamic non-negative matrix factorizations (NMF) with novel forms of temporal regularization. We describe a scalable optimization framework for our algorithms and report empirical results on topic detection problems in simulated document streams and real-world news stories.

Machine Learning in Computational Biology

http://mlcb.org

LOCATION
Hilton: Sutcliffe B
Friday, 7:30 - 10:30 and 3:30 - 6:30

Gunnar Raetsch Gunnar.Raetsch@tuebingen.mpg.de
Max Planck Society

Jean-Philippe Vert Jean-Philippe.Vert@mines.org
ParisTech

Tomer Hertz thertz@fhcrc.org
Fred Hutchnison Cancer Research Center

Yanjun Qi yanjun@nec-labs.com

Abstract

The field of computational biology has seen dramatic growth over the past few years, both in terms of new available data, new scientific questions, and new challenges for learning and inference. In particular, biological data is often relationally structured and highly diverse, well-suited to approaches that combine multiple weak evidence from heterogeneous sources. These data may include sequenced genomes of a variety of organisms, gene expression data from multiple technologies, protein expression data, protein sequence and 3D structural data, protein interactions, gene ontology and pathway databases, genetic variation data (such as SNPs), and an enormous amount of textual data in the biological and medical literature. New types of scientific and clinical problems require the development of novel supervised and unsupervised learning methods that can use these growing resources. The goal of this workshop is to present emerging problems and machine learning techniques in computational biology. We invited several speakers from the biology/bioinformatics community who will present current research problems in bioinformatics, and we invite contributed talks on novel learning approaches in computational biology. We encourage contributions describing either progress on new bioinformatics problems or work on established problems using methods that are substantially different from standard approaches. Kernel methods, graphical models, feature selection and other techniques applied to relevant bioinformatics problems would all be appropriate for the workshop.

SCHEDULE

7:25-7:50	HIV-Haplotype Inference using a Constraint-based Dirichlet Process Mixture Model Sandhya Prabhakaran
7:50-8:15	De novo RNA-seq-based Genome Annotation Jonas Behr
8:15-8:40	Inferring exon junction expression from RNA-Seq data Boyko Kakaradov
8:40-09:05	Predicting Tissue-Dependent Alternative Splicing Using Bayesian Neural Networks Hui Xiong
09:05-09:20	Coffee
9:20-10:05	Invited Talk: Orienting physical networks Roded Sharan
10:05-10:35	Link transfer for improving protein-protein interaction prediction using multiple species Adriana Birlutiu
15:45-16:10	Inferring Multiple Regulation Networks Julien Chiquet
16:10-16:35	A probabilistic topic model for the analysis of blood disease disorders Gerald Quon
16:35-17:00	JigPheno: Semantic Feature Extraction from biological images Theofanis Karaletsos
17:00-17:15	Coffee
17:15-17:40	A joint genotypic/phenotypic mixture model for GWAS Anna Goldenberg
17:40-18:00	Correction for Hidden Confounders in the Genetic Analysis of Gene Expression Jennifer Listgarten
18:00-18:20	Explaining Confounding Factors in eQTL Studies using a Dictionary of Latent Variables Nicolo Fusi
18:20-18:30	Discussion

INVITED SPEAKERS

Orienting physical networks

Roded Sharan, Tel-Aviv University

In a network orientation problem one is given a mixed graph, consisting of directed and undirected edges, and a set of source-target vertex pairs. The goal is to orient the undirected edges so that a maximum number of pairs admit a directed path from the source to the target. This problem arises in the context of analyzing physical networks of protein-protein and protein-DNA interactions. While the latter are naturally directed from a transcription factor to a gene, the direction of signal flow in protein-protein interactions is often unknown or cannot even be measured en masse. One then tries to infer this information by using causality data on pairs of genes such that the perturbation of one gene changes the expression level of the other gene. In my talk I will discuss the complexity of the problem, show approximation algorithms for several variants of it and present an efficient ILP solution for it. I will then describe the application of this algorithm to orient protein-protein interactions in yeast, improving our understanding of the structure and function of the network.

ACCEPTED CONTRIBUTIONS

HIV-Haplotype Inference using a Constraint-based Dirichlet Process Mixture Model

Sandhya Prabhakaran, Melanie Rey, Osvaldo Zagordi, Niko Beerenwinkel

De novo RNA-seq-based Genome Annotation

Jonas Behr, Regina Bohnert, Georg Zeller, Gabriele Schweikert, Lisa Hartmann, Lisa Smith, Gunnar Raetsch

Inferring exon junction expression from RNA-Seq data

Boyko Kakaradov, Brendan J. Frey

Predicting Tissue-Dependent Alternative Splicing Using Bayesian Neural Networks

Hui Xiong, Yoseph Barash, Brendan J. Frey

Link transfer for improving protein-protein interaction prediction using multiple species

Adriana Birlutiu, Florence d'Alche-Buc, Celine Brouard, Tom Heskes

Inferring Multiple Regulation Networks

Julien Chiquet, Yves Grandvalet, Christophe Ambroise,

A probabilistic topic model for the analysis of blood disease disorders

Gerald Quon, Ang Cui, Wenlian Qiao, Peter Zandstra, Rae Yeung, Alan Rosenberg, Quaid Morris,

JigPheno: Semantic Feature Extraction from biological images

Theofanis Karaletsos, Oliver Stegle, John Winn, Karsten Borgwardt,

A joint genotypic/phenotypic mixture model for GWAS

Anna Goldenberg, David Warde-Farley, Michael Brudno, Quaid Morris,

Correction for Hidden Confounders in the Genetic Analysis of Gene Expression

Jennifer Listgarten, Carl Cadie, Erich S. Schadt, David Heckerman

Explaining Confounding Factors in eQTL Studies using a Dictionary of Latent Variables

Nicolo Fusi, Oliver Stegle, Neil D. Lawrence, WS16

Machine Learning in Online Advertising

http://research.microsoft.com/~mload-2010/

LOCATION
Hilton: Diamond Head
Friday, 7:30 - 10:30 and 3:30 - 6:30

James G. Shanahan james.shanahan@gmail.com
Independent Consultant

Deepak Agarwal dagarwal@yahoo-inc.com
Yahoo! Research Labs

Tao Qin taoqin@microsoft.com
Microsoft Reserach Asia

Tie-Yan Liu tyliu@microsoft.com
Microsoft Research Asia

Abstract

Over the past 15 years online advertising, a $65 billion industry worldwide in 2009, has been pivotal to the success of the world wide web. This success has arisen largely from the transformation of the advertising industry from a low-tech, human intensive, ''Mad Men'' (ref AMC TV Series) way of doing work (that were common place for much of the 20th century and the early days of online advertising) to highly optimized, mathematical, machine learning-centric processes (some of which have been adapted from Wall Street) that form the backbone of many current online advertising systems. The dramatic growth of online advertising poses great challenges to the machine learning research community and calls for new technologies to be developed. Online advertising is a complex problem, especially from machine learning point of view. It contains multiple parties (i.e., advertisers, users, publishers, and ad platforms), which interact with each other and also have conflict of interests. It is highly dynamic in terms of the rapid change of user information needs, non-stationary bids of advertisers, and the frequent occurrences of ads campaigns. It is of very large scale, with billions of keywords, tens of millions of ads, billions of users, millions of advertisers where events such as clicks and actions can be extremely rare. In addition, the field lies at intersection of machine learning, economics, optimization, distributed systems and information science. For such a complex problem, conventional machine learning technologies and evaluation methodologies might not be sufficient, and the development of new algorithms and theories is sorely needed. The goal of this workshop is to overview the state of the art in online advertising, and to discuss future directions and challenges in research and development, from a machine learning point of view. We expect the workshop to help develop a community of researchers who are interested in this area, and yield future collaboration and exchanges. Possible topics include:

1) Dynamic/non-stationary/online learning algorithms for online advertising
2) Large scale machine learning for online advertising
3) Learning theory for online advertising
4) Learning to rank for ads display
5) Auction mechanism design for paid search, social network advertising and microblog advertising
6) System modeling for ad platform
7) Traffic and click through rate prediction
8) Bids optimization
9) Metrics and evaluation
10) Yield optimisation
11) Behavioral targeting modeling
12) Click fraud detection
13) Privacy in advertising
14) Crowd sourcing and inference
15) Mobile advertising and social advertising
16) Public datasets creation for research on online advertising

SCHEDULE

7:30-7:45	Opening remarks
7:45-8:30	Keynote: Machine Learning for Display Advertising Foster Provost
8:30-9:00	Invited talk: AdPredictor { Large Scale Bayesian Click-Through Rate Prediction in Microsoft's Bing Search Engine hore Graepel, Joaquin Qui~nonero Candela
9:00-9:30	Coffee Break
9:30-10:00	Invited talk: Hybrid Bidding for Keyword Auctions Ashish Goel
10:00-10:30	Poster Boaster and Discussion Sessions
3:30-4:15	Keynote: Visualization and Modeling of the Joint Behavior of Two Long Tailed Random Variables Art Owen
3:15-4:45	Invited talk: Click Modeling in Search Advertising: Challenges and Solutions Jianchang Mao
4:45-5:05	Invited talk: Digital Advertising: Going from Broadcast to Personalized Advertising James G. Shanahan
5:05-5:30	Coffee Break
5:30-5:50	Invited talk: Machine Learning for Advertiser Engagement Tao Qin
5:50-6:30	Wrap up and discussions

KEYNOTE SPEAKERS

Machine Learning for Display Advertising
Foster Provost, New York University

Most on-line advertisements are display ads, yet as compared to sponsored search, display advertising has received relatively little attention in the research literature. Nonetheless, display advertising is a hotbed of application for machine learning technologies. In this talk, I will discuss some of the relevant differences between online display advertising and traditional advertising, such as the ability to profile and target individuals and the associated privacy concerns, as well as differences from search advertising, such as the relative irrelevance of clicks on ads and the concerns over the content next to which brands' ads appear. Then I will dig down and discuss how these issues can be addressed with machine learning. I will focus on two main results based on work with the successful machine-learning based firm Media6degrees. (i) Privacy-friendly ''social targeting'' can be quite effective, based on identifying browsers that share fine-grained interests with a brand's existing customers--as exhibited through their browsing behavior. (ii) Clicks often are a poor surrogate for conversions for training targeting models, but there are effective alternatives.
This work was done in collaboration with Brian Dalessandro, Rod Hook, Alan Murray, Claudia Perlich, and Xiaohan Zhang.

Visualization and Modeling of the Joint Behavior of Two Long Tailed Random Variables
Art Owen, Stanford University

Many of the variables relevant to online advertising have heavy tails. Keywords range from very frequent to obscure. Advertisers span a great size range. Host web sites range from very popular to rarely visited. Much is known about the statistical properties of heavy tailed random variables. The Zipf distribution and Zipf-Mandelbrot distribution are frequently good approximations. Much less attention has been paid to the joint distribution of two or more such quantities. In this work, we present a graphical display that shows the joint behavior of two long tailed random variables. For ratings data (Netflix movies, Yahoo songs) we often see a strong head to tail affinity where the major players of one type are over-represented with the minor players of the other. We look at several examples which reveal properties of the mechanism underlying the data. Then we present some mathematical models based on bipartite preferential attachment mechanisms and a Zipf-Poisson ensemble. This is joint work with Justin Dyer.

INVITED SPEAKERS

Hybrid Bidding for Keyword Auctions
Ashish Goel, Stanford University

Search auctions have become a dominant source of revenue generation on the Internet. Such auctions have typically used per-click bidding and pricing. We propose the use of hybrid auctions where an advertiser can make a per-impression as well as a per-click bid, and the auctioneer then chooses one of the two as the pricing mechanism. We assume that the advertiser and the auctioneer both have separate beliefs (called priors) on the click-probability of an advertisement. We first prove that the hybrid auction is truthful, assuming that the advertisers are risk-neutral. We then show that this auction is different from the existing per-click auction in multiple ways: 1) It takes into account the risk characteristics of the advertisers. 2) For obscure keywords, the auctioneer is unlikely to have a very sharp prior on the click-probabilities. In such situations, the hybrid auction can result in significantly higher revenue. 3) An advertiser who believes that its click-probability is much higher than the auctioneer's estimate can use per-impression bids to correct the auctioneer's prior without incurring any extra cost. 4) The hybrid auction can allow the advertiser and auctioneer to implement complex dynamic programming strategies. As Internet commerce matures, we need more sophisticated pricing models to exploit all the information held by each of the participants. We believe that hybrid auctions could be an important step in this direction.

AdPredictor – Large Scale Bayesian Click-Through Rate Prediction in Microsoft's Bing Search Engine
Thore Graepel, Microsoft Research Cambridge
Joaquin Quinonero Candela, Microsoft

In the past years online advertising has grown at least an order of magnitude faster than advertising on all other media. Bing and Yahoo! have recently joined forces: all ads on both search engines are now served by Microsoft adCenter and all search results on Yahoo! are powered by Bing. Accurate predictions of the probability that a user clicks on an advertisement for a given query increase the efficiency of the ads network and benefit all three parties involved: the user, the advertiser, and the search engine. This talk presents the core machine learning model used by Microsoft adCenter for click prediction: an online Bayesian probabilistic classification model that has the ability to learn efficiently from terabytes of web usage data. The model explicitly represents uncertainty allowing for fully probabilistic predictions: 2 positives out of 10 instances or 200 out of 1000 both give an average of 20%, but in the first case the uncertainty about the prediction is larger. We discuss some challenges in machine learning for online systems, such as valid metrics, causal loops and biases in the training data.

Click Modeling in Search Advertising: Challenges and Solutions

Jianchang Mao, Yahoo! Labs

Sponsored search is an important form of online advertising that serves ads that match user's query on search result page. The goal is to select an optimal placement of eligible ads to maximize a total utility function that captures the expected revenue, user experience and advertiser return on investment. Most search engines use a pay-per-click model where advertisers pay the search engine a cost determined by an auction mechanism (e.g., generalized second price) only when users click on their ad. In this case, the expected revenue is directly tied to the probability of click on ads. Click is also often used as a proxy for measuring search user experience, and is a traffic driver for advertisers. Therefore, estimation of the probability of click is the central problem in sponsored search. It affects ranking, placement, quality filtering and price of ads.

Estimating click probability given a query-ad-user tuple is a challenging statistical modeling problem for a large variety of reasons, including click sparsity for the long tail of query-ad-user tuples, noisy clicks, missing data, dynamic and seasonal effects, strong position bias, selection bias, and externalities (context of an ad being displayed). In this talk, I will provide an overview on some of the machine learning techniques recently developed in Advertising Sciences team at Yahoo! Labs to deal with those challenges in click modeling. In specific, I will briefly describe: (i) a temporal click model for estimating positional bias, externalities, and unbiased user-perceived ad quality in a combined model; (ii) techniques for reducing sparsity by aggregating click history for sub-queries extracted with a CRF model and by leveraging data hierarchies; and (iii) use of a generative model for handling missing click history features. The talk is intended to give a flavor of how machine learning techniques can help solve some of the challenging click modeling problems arising in online advertising.

Digital Advertising: Going from Broadcast to Personalized Advertising

James G. Shanahan, Independent Consultant

Online advertising is a form of promotion that uses the Internet and World Wide Web for the expressed purpose of delivering marketing messages to attract customers. Examples of online advertising include text ads that appear on search engine results pages, banner ads, in-text ads, or Rich Media ads that appear on regular web pages, portals or applications. Since it inception over 15 years ago, online advertising has grown rapidly and currently accounts for 10% of the overall advertising spend (which is approximately $600 billion worldwide)). A large part of the more recent success in this field has come from the following key factors:

Personalization: offline advertising (via broadcast TV, radio, newspaper etc.) is largely a broadcast form of communication where as digital advertising is much more targeted and thus enables a personalized, and possibly informative, message to consumers.
Interactivity: internet advertising is becoming increasingly interactive with the advent of new forms of advertising such as social advertising; this is enables advertises and consumers to operate in a more conversant manner.
Engagement: consumers are spending more time online than with any other form of media thereby enabling a broader reach and deeper connection with consumers.
Explainabilty: advertisers are beginning to understand their consumers better.

This shift in focus in digital advertising from location (i.e., publisher web pages) to personalization has brought with it numerous challenges some of which have received a lot of research attention in the data mining and machine learning communities over the past 10-20 years. In this talk I will review, along the dimensions outlined above, some of these key technical problems and challenges that arise when adverting becomes personal. This will be done within the context of the elaborate (and ever-evolving) ecosystems of modern day digital advertising where one has to capture, store, and process petabytes of data within the constraints of a, sometimes, sequential workflow. The ultimate goal to is provide millisecond-based decision-making at each step of this workflow that enables customizable and engaging consumer experiences.

Keywords: Online advertising, personalization, classification, prediction, learning to rank ads, sponsored search, contextual advertising, display advertising, behavioral targeting, ad exchanges.

Machine Learning for Advertiser Engagement

Tao Qin, Microsoft Research Asia

Advertiser engagement, which goal is to attract more advertisers, make them loyal to the ad platform, and make them willing to spend more money on (online) advertising, is very important for an ad platform to boost its long-term revenue. Industry has paid more and more attention to advertiser engagement. For example, many search engines have provided tools to help advertisers, including keyword suggestion, traffic (number of impressions/ clicks) estimation, and bid suggestion. However, from the research point of view, the effort on advertiser engagement is still limited. In this talk, we discuss the challenges in advertiser engagement, especially from the machine learning perspective. Actually machine learning algorithms can be used in many aspects of online advertising, such as CTR prediction. We propose a number of principles that should be considered when using machine learning technologies to help advertiser engagement.

(1) Accurate. The results of learning algorithms should be as accurate as possible. This principle is the same as that in other machine learning tasks.
(2) Socially fair. The learning algorithms should promote diversity and be fair to even tail advertisers. In this way, more advertisers will feel engaged and the entire ads eco-system will become more healthy.
(3) Understandable. The evaluation metrics and learned models should be easy to interpret. In this way, it is easier for advertisers to diagnose their campaigns and identify the key aspects to improve. This will also make the ad platform more transparent to advertisers and increase their trust in the ad platform.
(4) Actionable. The learning algorithms should provide actionable suggestions/feedback to advertisers. In this way, the advertisers can take effective actions to improve their performances, and therefore stick to the ad platform in a more loyal fashion.

We will show several example problems in online advertising (such as effectiveness evaluation and auction mechanism) and discuss possible solutions based the above principles.
This is joint work with Bin Gao and Tie-Yan Liu.

Machine Learning meets Computational Photography

http://people.kyb.tuebingen.mpg.de/mhirsch/mlmcp/mlmcp.html

LOCATION
Hilton: Black Tusk
Saturday, 7:30 - 10:30 and 3:30 - 6:30

Stefan Harmeling stefan.harmeling@tuebingen.mpg.de
Max Planck Institute for Biological Cybernetics

Michael Hirsch michael.hirsch@tuebingen.mpg.de
Max-Planck Institute for Biological Cybernetics

Bill Freeman billf@mit.edu
Massachusetts Institute of Technology

Peyman Milanfar milanfar@ee.ucsc.edu
University of California

Abstract

Computational photography (CP) is a new field that explores and is about to redefine how we take photographs and videos. Applications of CP are not only ‘‘everyday’’ photography but also new methods for scientific imaging, such as microscopy, biomedical imaging, and astronomical imaging, and can thus be expected to have a significant impact in many areas. There is an apparent convergence of methods, what we have traditionally called ‘‘image processing’’, and recently many works in machine vision, all of which seem to be addressing very much the same, if not tightly related problems. These include deblurring, denoising, and enhancement algorithms of various kinds. What do we learn from this convergence and its application to CP? Can we create more contact between the practitioners of these fields, who often do not interact? Does this convergence mean that the fields are intellectually shrinking to the same point, or expanding and hence overlapping with each other more? Besides discussing such questions, the goal of this workshop is two-fold: (i) to present the current approaches, their possible limitations, and open problems of CP to the NIPS community, and (ii) to foster interaction between researchers from machine learning, neuro science and CP to advance the state of the art in CP. The key of the existing CP approaches is to combine (i) creative hardware designs with (ii) sophisticated computations, such as e.g. new approaches to blind deconvolution. This interplay between both hardware and software is what makes CP an ideal real-world domain for the whole NIPS community, who could contribute in various ways to its advancement, be it by enabling new imaging devices that are possible due to the latest machine learning methods or by new camera and processing designs that are inspired by our neurological understanding of natural visual systems. Thus the target group of participants are researchers from the whole NIPS community (machine learning and neuro science) and researchers working on CP and related fields.

SCHEDULE

7:30-7:32	Opening remarks
7:32-8:12	Invited talk: A Tour of Modern ``Image Processing’’ Peyman Milanfar
8:12-8:52	Invited talk: Mask-based Light Field Capture and Display Douglas Lanman
8:52-9:10	Coffee break
9:10-9:50	Invited talk: Denoising of Natural Images: Optimality and Fundamental Lower Bounds Boaz Nadler
9:50-10:30	Invited talk: Optimizing the Blur-Noise Tradeoff with Multiple-Photo Capture Sam Hasinoff
3:30-4:10	Invited talk: Tomography - The Other Inverse Problem Wolfgang Heidrich
4:10-4:50	Invited talk: Computational Photography with Mirrors Amit Agrawal
4:50-5:10	Coffee break
5:10-5:50	Invited talk: Codes for Sampling Images over Time and Space Bill Freeman
5:50-6:30	Panel discussion

INVITED SPEAKERS

A Tour of Modern “Image Processing”

Peyman Milanfar, UCSC

Recent developments in computational imaging and restoration have heralded the arrival and convergence of several powerful methods for adaptive processing of multidimensional data. Examples include Moving Least Square (from Graphics), the Bilateral Filter and Anisotropic Diffusion (from Vision), Boosting and Spectral Methods (from Machine Learning), Non-local Means (from Signal Processing), Bregman Iterations (from Applied Math), Kernel Regression and Iterative Scaling (from Statistics). While these approaches found their inspirations in diverse fields of nascence, they are deeply connected. In this talk, I will present a practical and unified framework for understanding some common underpinnings of these methods. This leads to new insights and a broad understanding of how these diverse methods interrelate. I will also discuss several applications, and the statistical performance of the resulting algorithms. Finally I briefly illustrate connections between these techniques and classical Bayesian approaches.

Mask-based Light Field Capture and Display
Douglas Lanman, MIT Media Lab

This talk describes light-efficient methods for capturing and displaying 3D images using thin, optically-attenuating masks. Light transport is modeled, under geometrical optics, as a 4D function: the light field; this function records the amount of light traveling through any point along any direction. Conventional photographs only record a 2D projection of the incident light field. Each image point is produced by integrating over the full hemisphere of incidence angles. Similarly, conventional displays only approximate a diffuse surface, where the amount of light leaving any point is constant over the full hemisphere of viewing angles. Thus, conventional cameras and displays only support 2D images, for which the perception of scene depth is lost. 3D images can be captured and displayed by including masks in conventional camera and display architectures. Parallax barriers are one example; a mask containing a uniform array of slits is placed slightly in front of a conventional display. This mask only allows certain disjoint display regions to be visible from each viewpoint. 3D image capture is achieved by placing a similar mask close to a sensor. In both cases, 3D images come at the cost of decreased resolution and brightness. This talk will present a first-principles analysis of dual-layer camera and display architectures, wherein the first layer is a conventional sensor or display and the second layer is a mask. Novel masks are developed that facilitate 3D image capture and display, outperforming conventional parallax barriers in terms of total light transmission and light field resolution. For 3D capture, a family of static, periodic, non-adaptive masks is derived from a frequency-domain analysis. For 3D display, a linear algebraic analysis reveals a set of time-multiplexed, aperiodic, adaptive masks. Four motivating applications are presented: digital photography, single-shot visual hull reconstruction, depth-sensing LCDs, and 3D display using dual-stacked LCDs.

Denoising of Natural Images: Optimality and Fundamental Lower Bounds
Boaz Nadler, Weizmann Institute of Science

In natural image denoising, the task is to estimate a clean version of a given noisy image, using prior knowledge on the statistics of natural images. The problem has been studied intensively with impressive progress achieved in recent years. However, it seems that image denoising has reached a plateau, with new algorithms improving over previous ones by only fractional dB values. A key question is thus: How much more can current methods be improved? In this talk we'll discuss optimal natural image denoising and its fundamental lower bounds. In particular, we'll show that at moderate noise levels, current state-of-the-art denoising algorithms, that use a fixed small support window around each denoised pixel, are approaching optimality and cannot be further improved beyond fractional dB values. Joint work with Anat Levin.

Optimizing the Blur-Noise Tradeoff with Multiple-Photo Capture
Sam Hasinoff, MIT CSAIL

Capturing multiple photos at different focus settings is a powerful approach for reducing optical blur, but how many photos should we capture within a fixed time budget? We develop a framework to analyze optimal capture strategies balancing the tradeoff between defocus and sensor noise, incorporating uncertainty in resolving scene depth. We derive analytic formulas for restoration error and use Monte Carlo integration over depth to derive optimal capture strategies for different camera designs, under a wide range of photographic scenarios. We also derive a new upper bound on how well spatial frequencies can be preserved over the depth of field. Our results show that by capturing the optimal number of photos, a standard camera can achieve performance at the level of more complex computational cameras, in all but the most demanding of cases. We also show that computational cameras, although specifically designed to improve one-shot performance, generally benefit from capturing multiple photos as well.

Tomography---the other Inverse Problem
Wolfgang Heidrich, UBC

Much Computational Photography work has recently focussed on deconvolution problems, which are a specific kind of inverse problem. In this presentation, I will talk about tomography, another form of inverse problem that has so far largely been neglected in this community. I will demonstrate how tomographic methods can be used to solve problems in Computational Photography, and I will outline how I believe Machine Learning approaches can help in tomographic reconstruction down the road.

Computational Photography with Mirrors
Amit Agrawal, MERL

The narrow field of view of conventional cameras is a limitation in wide-angle imaging applications. Catadioptric cameras combine a mirror with a camera to obtain wide field of view in a single photo. However, algorithms developed for perspective cameras have difficulties on wide-angle photos due to non-perspective effects and distortions. In this talk, I will describe how to extend computational photography approaches to wide angle imaging using mirrors. I will describe a complete system for wide-angle light field capture, processing and rendering using an array of mirrors for effects such as digital refocusing in wide-angle images. I will present new analytical models for non-central catadioptric imaging, which leads to significant speed up and high accuracy for applications such as sparse and dense 3D reconstruction using wide-angle photos. Finally, I will discuss new challenges in traditional vision applications such as deblurring, image enhancement, feature extraction and recognition for wide-angle photos.

Codes for Sampling Images over Time and Space
Bill Freeman, MIT EECS and CSAIL

The human visual system, and almost every digital camera, achieves trichromacy by using sensors of different spectral sensitivity arranged in some spatial pattern. What pattern of spatial sampling is best? I'll describe work on learning an optimal (ok, locally optimal) color filter array. This learned sampling pattern leads to better reconstructions of color images than the Bayer pattern and many others. If there is time, I'll talk about some issues in sampling over time, and show simulation results for a camera that can both add and subtract from a pixel intensity counter.

Modeling Human Communication Dynamics

http://people.kyb.tuebingen.mpg.de/mhirsch/mlmcp/mlmcp.html

LOCATION
Westin: Alpine A
Friday, 7:30 - 10:30 and 3:30 - 6:30

Louis-Philippe Morency morency@ict.usc.edu
University of Southern California

Daniel Gatica-Perez gatica@idiap.ch
IDIAP Research Institute

Nigel G. Ward nigel@utep.edu
University of Texas, El Paso

Abstract

Modeling human communicative dynamics brings exciting new problems and challenges to the NIPS community. The first goal of this workshop is to raise awareness in the machine learning community of these problems, including some applications needs, the special properties of these input streams, and the modeling challenges. The second goal is to exchange information about methods, techniques, and algorithms suitable for modeling human communication dynamics. Face-to-face communication is a highly interactive process in which the participants mutually exchange and interpret verbal and nonverbal messages. Both the interpersonal dynamics and the dynamic interactions among an individual´s perceptual, cognitive, and motor processes are swift and complex. How people accomplish these feats of coordination is a question of great scientific interest. Models of human communication dynamics also have much potential practical value, for applications including the understanding of communications problems such as autism and the creation of socially intelligent robots able to recognize, predict, and analyze verbal and nonverbal behaviors in real-time interaction with humans.

SCHEDULE

7:30-7:35	Welcome
7:35-8:15	Invited talk: Models for Multiparty Turn Taking in Situated Dialog Dan Bohus
8:15-8:55	Invited talk: Modeling Natural Facial Behavior Marian Stewart Bartlett
8:55-9:15	Coffee break
9:15-9:40	Teaser talks for posters
9:40-10:20	Invited talk Jeff A. Bilmes
10:20-10:30	General discussion
10:30-15:30	Lunch/Poster preview
15:30-16:10	Invited talk: Communicative Inference as Social Cognition Noah D. Goodman
16:10-17:10	Poster session
17:10-17:30	Coffee Break
17:30-18:10	Invited talk Justine Cassell

INVITED SPEAKERS

Models for Multiparty Turn Taking in Situated Dialog
Dan Bohus, Microsoft Research

In this talk I will describe recent work on a computational framework for managing the turn-taking process in situated, multiparty dialog. The approach harnesses component models that leverage audio-visual evidence to track the multiparty conversational dynamics, to make floor control decisions, and to render these decisions into a set of coordinated verbal and non-verbal behaviors. I will review experiments which demonstrate how the approach enables an embodied conversational agent to participate in and shape flow of multiparty dialog. Finally, I will discuss lessons learned from these experiments and highlight future challenges in this space.

Modeling Natural Facial Behavior
Marian S. Bartlett, UCSD

This talk reviews recent research in my lab modeling natural facial expression with automated systems. Automated systems enable new research into expression dynamics that was previously infeasible with manual coding, or which would have required application of electrodes to the face, which can influence facial behavior. The talk first describes projects on measurement of dynamic coupling of facial behavior to measure spontaneous mimicry, as well as detection of deception. We show that facial mimicry correlates with the ability to detect when a person is lying. This had long been hypothesized by embodied theories of cognition but never previously shown. These findings were made possible by the use of novel computer vision techniques that allowed us to obtain rich quantitative information about facial dynamics. The talk next describes development of interventions for children with autism. The interventions employ computer vision systems to train facial expression production, provide practice in facial mimicry, and immediate feedback on the child’s facial expressions. Finally, if time permits, I will review our work on children’s facial behavior during problem solving. Clustering techniques are employed to demonstrate differences in expression dynamics between older and younger children during problem solving.

TBA
Jeff A. Bilmes, University of Washington

Communicative Inference as Social Cognition
Noah D. Goodman, Stanford University
I will describe an approach to capturing inferential aspects of human communication within a probabilistic framework for social cognition. I will apply this framework to a set of empirical results on pragmatics, natural pedagogy, and word learning.

TBA
Justine Cassell, Carnegie Mellon University

ACCEPTED CONTRIBUTIONS

An Instantaneous Correlation Algorithm for Assessing Intra and Inter Subject Coordination During Communicative Behavior
Adriano V. Barbosa, Eric Vatikiotis-Bateson, Martin Oberg and Rose-Marie D´echaine, University of British Columbia

Inferring truth from multiple annotators for multimodal social interaction analysis
Gokul Chittaranjan, IDIAP
Oya Aran, IDIAP
Daniel Gatica-Perez, IDIAP

Multi-modal Analysis of Interactional Rapport in Three Language/Cultural Groups
Susan Duncan, Tiziana Baldenebro and Atoor Lawandaw, University of Chicago
Gina-Anne Levow, University of Washington

Modeling and Tracking Surface Instructional Discourse Dynamics
Juan M. Huerta, IBM Research

Interval-based Modeling of Human Communication Dynamics via Hybrid Dynamical Systems
Hiroaki Kawashima, Kyoto University Takashi Matsuyama, Kyoto University

Modeling Social Cues: Effective Features for Predicting Listener Nods
Faisal Khan, Bilge Mutlu and Xiaojin Zhu, University of Wisconsin -Madison

Enriching Interpersonal Human Interactions towards Effective Personal and Professional Communications
Sreekar Krishna and Sethuraman Panchanathan, Arizona State University
Bhavesh M. Patel, Mayo Clinic Hospital, Phoenix

Preliminaries to an Account of Multi-Party Conversational Turn-Taking as an Antiferromagnetic Spin Glass
Kornel Laskowski, KTH Mattias Heldner, KTH Jens Edlund, KTH

Modeling Individuals and Groups in Face-to-Face Interactions
Bruno Lepri, FBK and MIT Wen Dong, MIT Fabio Pianesi, FBK Alex (Sandy) Pentland, MIT

Optimizing End-of-Turn Detection for Spoken Dialog Systems
Antoine Raux, Honda Research Maxine Eskenazi, Carnegie Mellon University

The Dynamics of Concurrent Gaze and Speech: Order is Important, Synchronization, Not So Much
Maria Staudte, Saarland University
Matthew W. Crocker, Saarland University

The Challenge of Modeling Dialog Dynamics
Nigel G. Ward, University of Texas at El Paso

Monte Carlo Methods for Bayesian Inference in Modern Day Applications

http://montecarlo.wikidot.com/

LOCATION
Hilton: Mt Currie North
Friday, 7:30 - 10:30 and 3:30 - 6:30

Ryan P. Adams rpa@cs.toronto.edu
University of Toronto

Mark A. Girolami girolami@dcs.gla.ac.uk
University of Glasgow

Iain Murray i.murray@ed.ac.uk
University of Edinburgh

Abstract

Monte Carlo methods have been the dominant form of approximate inference for Bayesian statistics over the last couple of decades. Monte Carlo methods are interesting as a technical topic of research in themselves, as well as enjoying widespread practical use. In a diverse number of application areas Monte Carlo methods have enabled Bayesian inference over classes of statistical models which previously would have been infeasible. Despite this broad and sustained attention, it is often still far from clear how best to set up a Monte Carlo method for a given problem, how to diagnose if it is working well, and how to improve under-performing methods. The impact of these issues is even more pronounced with new emerging applications. This workshop is aimed equally at practitioners and core Monte Carlo researchers. For practitioners we hope to identify what properties of applications are important for selecting, running and checking a Monte Carlo algorithm. Monte Carlo methods are applied to a broad variety of problems. The workshop aims to identify and explore what properties of these disparate areas are important to think about when applying Monte Carlo methods.
The workshop wiki contains a more detailed list of discussion topics and recommended background reading. We welcome contributions: anyone can create an account and edit the wiki.

SCHEDULE

7:50--8:00	Welcome
8:00--8:30	Review of open and contributed issues and discussion topics Iain Murray
8:30--9:00	Field Report: Adaptive Particle MCMC for Mixed Effect Models and Stochastic Differential Equations Julien Cornebise
9:00--9:30	Coffee Break
9:30--10:00	Natively probabilistic computation: principles and applications Vikash Mansinghka
10:00--10:30	Panel and general discussion
10:30--11:00	Poster spotlights
11:00--13:00	Posters
15:30--16:00	Gaussian processes and compactly supported correlation functions Derek Bingham
16:00--16:30	MCMC for Bayesian Nonparametric Models: A Bag of Tricks Yee Whye Teh
16:30--17:00	Distributed Gibbs sampling for Topic Models and Bayesian Networks Max Welling
17:00--17:30	Coffee Break
17:30--18:00	MCMC in Probabilistic Databases and Cluster Computing Andrew McCallum
18:00--18:30	Discussion and wrap-up

INVITED SPEAKERS

Field Report: Adaptive Particle MCMC for Mixed Effect Models and Stochastic Differential Equations
Julien Cornebise, University of British Columbia

We report here on our practical experience when combining Particle MCMC with Adaptive Monte-Carlo in nonlinear mixed-effect models for longitudinal studies. Our motivating application is a population pharmacokinetic/pharmacodynamic study for glucose/insulin regulation, but the very same specific conditional independence structure is used way beyond pharmacology: forestry, dairy sciences, waste-water treatment. . . We show how to introduce Adaptive Monte-Carlo to improve the mixing in these intricate models (strongly non-linear dynamics). We exploit the conditional independence structure to reduce considerably the dimension of the parameter space on which the proposal covariance matrix is learned, hence further augmenting the mixing. We also show how this allows for exact proposal on some of the hyper-parameters, alternating Particle Metropolis Hastings and Gibbs sampler. We also warn against a very simple-to-overlook mistake easy to make when using Adaptive MC on a constrained domain (e.g. positivity). Finally, we will study which parts of the inference can be parallelized, especially on a GPU. We outline some practical caveats in terms of random number generation and of ease of implementation.
Joint work with Arnaud Doucet and Gareth W. Peters.

Natively probabilistic computation: principles and applications
Vikash Mansinghka, Navia Systems

Complex probabilistic models and Bayesian inference are becoming increasingly critical across science and industry, especially in large-scale data analysis. They are also central to our best computational accounts of human cognition, perception and action. However, all these efforts struggle with the infamous curse of dimensionality. Rich probabilistic models can seem hard to write and even harder to solve, as specifying and calculating probabilities often appears to require the manipulation of exponentially (and sometimes infinitely) large tables of numbers.

We argue that these difficulties reflect a basic mismatch between the needs of probabilistic reasoning and the deterministic, functional orientation of our current hardware, programming languages and CS theory. To mitigate these issues, we have been developing a stack of abstractions for natively probabilistic computation, based around stochastic simulators (or samplers) for distributions (especially posterior simulation), rather than evaluators for deterministic functions. Ultimately, our aim is to build and deploy machines for which Bayesian inference by stochastic simulation is as natural as logic and arithmetic are for current microprocessors.

In this talk, I will briefly describe two systems we have built using these new tools: 1. Veritable, a probabilistic table, based on a novel nonparametric Bayesian probabilistic program and fast stochastic inference. 2. A prototype software-reprogrammable probabilistic video processor, offering 1000x latency and throughput and 10-30x power improvements over software solutions, making dense vision MRFs tractable in real time. I will also discuss the challenges involved in analyzing the computational complexity of probabilistic programs and in reliably synthesizing probably tractable programs. I will touch on both new theoretical results and some reinterpretations of recent theory that is, to the best of our knowledge, new to the Bayesian statistics community.
This talk includes joint work with Eric Jonas, Cameron Freer, Daniel Roy, Keith Bonawitz, Beau Cronin, and Joshua Tenenbaum.

Gaussian processes and compactly supported correlation functions
Derek Bingham, Simon Fraser University

Computer simulators are often used to study real-world processes that are too difficult to observe directly. Experimenter's are often interested in building a statistical surrogate for the computer model to avoid constantly running the simulator at different input settings. Building an emulator for a computer simulator using standard Gaussian process models can be computationally infeasible when the number of evaluated input values is large. As an alternative, we propose using compactly supported correlation functions, which produce sparse correlation matrices that can be more easily manipulated. Following the usual approach of taking the correlation to be a product of correlations in each input dimension, we show how to impose restrictions on the correlation range for each input, giving sparsity, while also allowing the ranges to trade-off against one another, thereby giving good predictive performance when the data are anisotropic. Issues related to Bayesian inference and implementation of MCMC for large data-sets will be discussed. As an illustration, the method is used to construct an emulator of photometric red-shifts of cosmological objects.
This is joint work with Salman Habib, Katrin Heitman (Los Alamos National Lab) and Cari Kaufman (UC Berkeley).

MCMC for Bayesian Nonparametric Models: A Bag of Tricks
Yee Whye Teh, University College London

Markov chain Monte Carlo (MCMC) sampling forms the most popular class of inference methods for Bayesian nonparametric models based on Dirichlet processes, Indian buffet processes etc. And there are good reasons why: there is a huge variety of very powerful techniques, they can be used together, they often do not require solving complex equations and integrals, and they are guaranteed to converge to the distribution of interest. In this talk, I will describe the bag of tricks that have helped me design MCMC samplers, and discuss some open questions related to inference in Bayesian nonparametric models.

Distributed Gibbs sampling for Topic Models and Bayesian Networks
Max Welling, University of California, Irvine

The exponential growth of data has prompted the need for scalable machine learning. We have studied the possibility of performing efficient Gibbs sampling in the face of very large datasets. To achieve that goal we have explored two orthogonal ideas: 1) distributed Gibbs sampling --both synchronous and asynchronous, 2) adaptive Gibbs updates that avoid the need to look at every possible assignment. These techniques are tested on topic models and more general Bayesian networks, both parametric and non-parametric. We show perplexity and speedup results on a number of datasets: Nips, Newsgroups, NYT, Wikipedia and Medline, the latter containing over 700M tokens. The main conclusions are that 1) distributed inference, while no longer exact, has no noticeable negative effect on perplexity relative to exact Gibbs sampling but may lead to considerable speedups --up to 700 times if we distribute over 1000 processors, 2) both speedup techniques can be easily combined to lead to further speedups, 3) the proposed techniques seem applicable to a wide range of Bayesian network models.
Joint work with Arthur Acuncion, David Newman and Padhriac Smyth.

MCMC in Probabilistic Databases and Cluster Computing
Andrew McCallum, UMass Amherst

Incorporating uncertainty and probabilistic inference into large-scale databases has posed many challenges, often leading to systems that sacrifice modeling power, scalability, or restrict the class of supported queries. We aim to achieve a good balance among these considerations with an approach in which the underlying relational database represents a single possible world, an imperatively-defined factor graph [1] encodes a distribution over the set of possible worlds, and Markov chain Monte Carlo inference is used for inference [2].
I will describe experiments that operate on hundreds of thousands of entities and relations from both FreeBase and five years of NYTimes articles. I will also discuss methods of distributed MCMC inference on the problem of entity disambiguation for 25k NYTimes person mentions using a cluster of 50 machines.
Joint work with Michael Wick, Sameer Singh, Karl Schultz, Sebastian Riedel, Limin Yao and Gerome Miklau.

[1] Andrew McCallum, Karl Schultz, Sameer Singh. "FACTORIE: Probabilistic Programming via Imperatively Defined Factor Graphs." Neural Information Processing Systems Conference (NIPS), 2009.
[2] Michael Wick, Andrew McCallum, Gerome Miklau. "Scalable Probabilistic Databases with Factor Graphs and MCMC." Very Large Data Bases (VLDB) 2010.

Networks Across Disciplines: Theory and Applications

http://morrislab.med.utoronto.ca/~anna/networksnips2010/

LOCATION
Westin: Nordic
Saturday, 7:30 - 10:30 and 3:30 - 6:30

Edoardo M. Airoldi airoldi@fas.harvard.edu
Harvard University

Anna Goldenberg nyulik@gmail.com
University of Toronto

Jure Leskovec jure@cs.stanford.edu
Stanford University

Quaid Morris quaid.morris@gmail.com
University of Toronto

Abstract

Networks are used across a wide variety of disciplines to describe interactions between entities ---in sociology these are relations between people, such as friendships (Facebook); in biology ---physical interactions between genes; and many others: the Internet, sensor networks, transport networks, ecological networks just to name a few. Computer scientists, physicists and mathematicians search for mechanisms and models that could explain observed networks and analyze their properties. The research into theoretical underpinnings of networks is very heterogeneous and the breadth of existing and possible applications is vast. Yet, many of such works are only known within their specific areas. Many books and articles are written on the subject making it hard to tease out the important unsolved questions especially as richer data becomes available. These issues call for collaborative environment, where scientists from a wide variety of fields could exchange their ideas: theoreticians could learn about new questions in network applications, whereas applied researchers could learn about potential new solutions for their problems. Researchers in different disciplines approach network modeling from complementary angles. For example, in physics, scientists create generative models with the fewest number of parameters and are able to study average behavior of very large networks, whereas in statistics and social science, the focus is often on richer models and smaller networks. Continuous information exchange between these groups can facilitate faster progress in the field of network modelling and analysis.

SCHEDULE

7:30-7:35	Introduction by the organizers
7:35-8:20	Collective Graph Identification Lise Getoor
8:20-9:05	The Trouble with Community Detection Aaron Clauset
9:00-9:30	Coffee Break
9:30-10:15	Bounding Complex Network Models with Bernoulli Graphs Carter Butts
10:15-10:45	Poster Spotlights
10:45-3:30	BREAK (POSTERS)
3:30-4:15	TBD. See workshop's webpage for details
4:15-5:00	Geometry based graph and network models Sayan Mukherjee
5:00-5:15	Coffee break
5:15-6:00	Facebook: Challenges for 2011 Jonathan Chang
6:00-6:30	Panel discussion and closing remarks

INVITED TALKS

Collective Graph Identification

Lise Getoor, Dept. of Computer Science, University of Maryland

The importance of network analysis is growing across many domains, and is fundamental in understanding online social interactions, biological processes, communication, ecological, financial, and transportation networks, and many more. In most of these domains, the networks of interest are not directly observed, but must be inferred from noisy and incomplete data, data that was often generated for purposes other than scientific analysis. In this talk, I will describe graph identification, the process of inferring the hidden network from noisy observational data. In particular, I will describe a collective approach to graph identification, which interleaves the necessary steps in the reconstruction of the network.

The Trouble with Community Detection

Aaron Clauset, Department of Computer Science, University of Colorado, Boulder

Modular structures in complex networks can be extremely important for understanding the functional, dynamical, and evolutionary properties of networks, and are widely believed to be ubiquitous in complex social, biological and technological networks. Most of the empirical evidence in support of this modular hypothesis, however, is indirect and derived from ‘‘community’’ or module detection algorithms. In general, these techniques do not yield unambiguous results and their objective performance in scientific contexts is not well characterized. In this talk, I'll discuss some of the problems with the existing popular community detection frameworks and show that even in simple contexts they can produce highly counter-intuitive results. A consequence is that probably none of the existing claims of modular structure in, for example, biological networks should be trusted and there remains a great deal of work to be done to test the modular-organization hypothesis in such contexts. I'll conclude with some forward-looking thoughts about the general problem of identifying network modules from connectivity data alone, and the likelihood of circumventing these problems using, for instance, notions of functionality.

Bounding Complex Network Models with Bernoulli Graphs
Carter Butts, Department of Sociology, UC Irvine

Complex network models (i.e., stochastic models for networks incorporating heterogeneity and/or dependence among edges) are increasingly widely used in the study of social and other networks, but few techniques other than simulation have been available for studying their behavior. Random graphs with independent edges (i.e., the Bernoulli graphs), on the other hand, are well-studied, and a large literature exists regarding their properties. Here, I demonstrate a method for leveraging this knowledge by constructing families of Bernoulli graphs that bound the behavior of general exponential-family random graphs in a well-defined sense. By studying the behavior of these Bernoulli graph bounds, one can thus constrain the properties of a given random graph. Several applications of this method to the study of complex network models are discussed, including degeneracy identification and robustness testing.

Geometry based graph and network models
Sayan Mukherjee, Statistics, Duke University

We will discuss two graph or network models based on geometric ideas. The first model was motivated by problems in computational molecular biology. The objective is to infer conditional dependece structure between gene products or molecular pathways that are predictive or explain variation in disease phenotypes. This objective will be related to inference of a geometric quantity, the gradient of a regression function. An application to cancer progression will be used to illustrate the ideas. The second model involves the inference of higher-order dependence structure in graphical models. Ideas from computational geometry/topology and spatial point processes are integrated to achieve this, the inference is formulated in a Bayesian paradigm.

Facebook: Challenges for 2011
Jonathan Chang, Facebook

This past year Facebook launched several new products such as Questions and Places, and revamped several products such as Groups and Photos. Each of these products offer rich, structured data to supplement the existing friendship and content graphs. With these new data come new opportunities for machine learning and data analysis ---underpinning the success of these products is our ability to answer questions such as ‘‘Does this user provide high-quality answers to questions?’’ and ‘‘Are these places socially interesting to users?’’ In this talk, I will describe these new data and some approaches to answering these questions. I will also talk about many open problems surrounding these products which may lead to further improvements.

ACCEPTED POSTERS

Active Surveying
Hossam Sharara, Lise Getoor, Myra Norton,

Active Learning on Graphs via Spanning Trees
Nicol`o Cesa-Bianchi, Claudio Gentile, Fabio Vitale, Giovanni Zappella,

Unsupervised discriminative approach to find biomarkers in lung cancer
Anna Goldenberg, Sara Mostafavi, Gerald Quon, Paul C. Boutros, Quaid Morris,

Predicting Node Labels in Large Networks
Sara Mostafavi, Anna Goldenberg, Quaid Morris,

Beyond Keyword Search: Discovering Relevant Scientific Literature
Khalid El-Arini, Carlos Guestrin,

Statistical Mechanics of Semi-Supervised Clustering in Sparse Graphs
Aram Galstyan, Greg Ver Steeg, Armen Allahverdyan,

Co-Evolving Mixed Membership BlockModels
Yoon-Sik Cho, Aram Galstyan, Greg Ver Steeg,

Infinite Multiple Membership Relational Modeling for Complex Networks
Morten Mørup, Mikkel N. Schmidt, Lars Kai Hansen,

Soft Partitioning in Networks via Bayesian Non-negative Matrix Factorization
Ioannis Psorakis, Stephen Roberts, Ben Sheldon,

Estimating Networks with Jumps
Mladen Kolar, Eric P. Xing,

Community Detection in Networks: The Leader-Follower Algorithm
Devavrat Shah, Tauhid R. Zaman,

Node Clustering in Graphs: An Empirical Study
Ramnath Balasubramanyan, Frank Lin, William W. Cohen,

Community Finding: Partitioning Considered Harmful
Fergal Reid, Aaron McDaid, Neil Hurley,

A Latent Space Mapping for Link Prediction
Anthony Brew, Michael Salter-Townshend,

Regularized Output Kernel Regression applied to protein-protein interaction network inference
Céline Brouard, Marie Szafranski, Florence d’Alché-Buc,

RegnANN: network inference using Artificial Neural Networks
Marco Grimaldi, Giuseppe Jurman, Roberto Visintainer,

Introduction to spectral metrics in biological network theory
Roberto Visintainer, Giuseppe Jurman, Marco Grimaldi, Cesare Furlanello,

A Triangle Inequality for p-Resistance
Mark Herbster,

Sampling Graphs with a Prescribed Joint Degree Distribution Using Markov Chains
Isabelle Stanton, Ali Pinar,

Population size estimation and Internet link structure
Abraham Flaxman, Stephen E. Fienberg,

Modeling the Variance of Network Populations with Mixed Kronecker Product Graph Models
Sebastian Moreno, Jennifer Neville, Sergey Kirshner, S.V.N. Vishwanathan,

Exact learning curves for Gaussian process regression on community random graphs
Matthew J. Urry, Peter Sollich,

Higher-order Graphical Models for Classification in Social and Affiliation Networks
Elena Zheleva, Lise Getoor, Sunita Sarawagi,

Four factors influencing effectiveness in email communication networks
Ofer Engel

New Directions in Multiple Kernel Learning

http://doc.ml.tu-berlin.de/mkl_workshop/

LOCATION
Hilton: Mt Currie North
Saturday, 7:30 - 10:30 and 3:30 - 6:30

Marius Kloft mkloft@cs.tu-berlin.de
UC Berkeley / TU Berlin

Ulrich Rueckert rueckert@eecs.berkeley.edu
UC Berkeley

Cheng Soon Ong chengsoon.ong@inf.ethz.ch
ETH Zurich

Alain Rakotomamonjy alain.rakoto@insa-rouen.fr
LITIS, Univ de Rouen

Soeren Sonnenburg Soeren.Sonnenburg@tu-berlin.de
Berlin Institute of Technology

Francis Bach francis.bach@ens.fr
INRIA - Ecole Normale Superieure

Abstract

Research on Multiple Kernel Learning (MKL) has matured to the point where efficient systems can be applied out of the box to various application domains. In contrast to last year's workshop, which evaluated the achievements of MKL in the past decade, this workshop looks beyond the standard setting and investigates new directions for MKL. In particular, we focus on two topics: 1. There are three research areas, which are closely related, but have traditionally been treated separately: learning the kernel, learning distance metrics, and learning the covariance function of a Gaussian process. We therefore would like to bring together researchers from these areas to find a unifying view, explore connections, and exchange ideas. 2. We ask for novel contributions that take new directions, propose innovative approaches, and take unconventional views. This includes research, which goes beyond the limited classical sum-of-kernels setup, finds new ways of combining kernels, or applies MKL in more complex settings. Taking advantage of the broad variety of research topics at NIPS, the workshop aims to foster collaboration across the borders of the traditional multiple kernel learning community.

SCHEDULE

7:30-7:40	Opening remarks
7:40-8:10	Invited Talk: Various Formulations for Learning the Kernel and Structured Sparsity Massimiliano Pontil
8:10-8:40	Invited talk: A Gaussian Process View on MKL Raquel Urtasun
8:40-9:00	Regularization Strategies and Empirical Bayesian Learning for MKL Ryota Tomioka, Taiji Suzuki
9:00-9:30	Coffee Break
9:30-9:50	Online MKL for Structured Prediction Andre F. T. Marins, Noah A. Smith, Eric P. Xing, Pedro M. Q. Aguiar, Mario A.T. Figueiredo
9:50-10:10	Multiple Gaussian Process Models Cedric Archambeau, Francis Bach
10:10-10:30	Multitask Multiple Kernel Learning (MT-MKL) Chris Widmer, Nora C. Toussaint, Yasemin Altun, Gunnar Raetsch
3:30-4:00	Invited talk: Structured Regularization for MKL Guillaume Obozinski
4:00-4:30	Invited talk: Distance Metric Learning for Kernel Machines Kilian Q. Weinberger
4:30-4:50	Poster spotlights
4:50-5:30	Coffee break and poster session continued
5:30-5:50	Co-regularized Spectral Clustering with Multiple Kernels Abhishek Kumar, Piyush Rai, Hal Daume III
5:50-6:30	Panel discussion - Panelists: Guillaume Obozinski (ENS Paris) Massimiliano Pontil (UCL) Raquel Urtasun (TTI Chicago) Kilian Q. Weinberger (Univ. of Washington)

INVITED TALKS

Various Formulations for Learning the Kernel and Structured Sparsity
Massimiliano Pontil, University College London

I will review an approach to learning the kernel, which consists in minimizing a convex objective function over a prescribed set of kernel matrices. I will establish some important properties of this problem and present a reformulation of it from a feature space perspective. A well studied example covered by this setting is multiple kernel learning, in which the set of kernels is the convex hull of a finite set of basic kernels. I will discuss extensions of this setting to more complex kernel families, which involve additional constraints and a continuous parametrization. Some of these examples are motivated by multi-task learning and structured sparsity, which I will describe in some detail during the talk.

A Gaussian Process View on MKL
Raquel Urtasun, TTI Chigaco

Gaussian processes (GPs) provide an appealing probabilistic framework for multiple kernel learning (MKL). For more than a decade, it has been common practice to learn the well known sum-of-kernels by, for example, maximum likelihood estimation. In this talk, I'll first introduce the sum-of-kernels Gaussian process formulation. I'll then show how to go beyond convex formulations by learning the GP covariance. In particular, I'll first introduce parametric forms of the covariance driving connections to co-training. I'll then show how to learn non-parametric covariances via latent spaces. If time permits, I'll talk about multi-task learning as well as multi-output Gaussian processes and show connections to metric learning. I'll demonstrate the performance of some of these approaches in computer vision tasks such as object recognition.

Structured Regularization for MKL
Guillaume Obozinski, ENS Paris

It was realized soon after the introduction of Multiple Kernel Learning that ℓ_1 - regularization and its groupwise extension are related to MKL by dualization. MKL can therefore be viewed as providing an extension of sparsity to function spaces. However, this extension is not limited ℓ_1 - norm. For example, extensions to ℓ_p norms leads to several forms of non-sparse MKL. In this talk, we will discuss how structured sparsity can generically be introduced in the MKL framework, how existing algorithmic approaches extend to this case and how this leads naturally to the notion of structured functional spaces.

Distance Metric Learning for Kernel Machines
Kilian Q. Weinberger, Washington University, St. Louis

Recent work in metric learning has significantly improved the state-of-the-art in k-nearest neighbor classification. However, Support vector machines (with RBF kernels) are arguably the most popular class of classification algorithms that uses distance metrics to compare examples. In this talk I will introduce support vector metric learning (SVML), an algorithm that seamlessly combines both by learning a Mahalanobis metric at the same time as the RBF-SVM decision boundary. SVML is an effective tool for automatically pre-processing data sets for classification, as well as visualizing the structure of SVM decision boundaries. We demonstrate the capabilities (and shortcomings) of our algorithm on 10 benchmark data sets of varying sizes and difficulties.

CONTRIBUTED TALKS

Regularization Strategies and Empirical Bayesian Learning for MKL
Ryota Tomioka, Taiji Suzuki

Multiple kernel learning (MKL) has received considerable attention recently. In this paper, we show how different MKL algorithms can be understood as applications of different types of regularization on the kernel weights. Within the regularization view we consider in this paper, the Tikhonov-regularization-based formulation of MKL allows us to consider a generative probabilistic model behind MKL. Based on this model, we propose learning algorithms for the kernel weights through the maximization of marginalized likelihood.

Online MKL for Structured Prediction
Andre F. T. Martins, Noah A. Smith, Eric P. Xing, Pedro M. Q. Aguiar, Mario A. T. Figueiredo

Training structured predictors often requires a considerable time selecting features or tweak-ing the kernel. Multiple kernel learning (MKL) sidesteps this issue by embedding the kernel learning into the training procedure. Despite the recent progress towards efficiency and scalabil-ity of MKL algorithms, the structured output case remains an open research front. We claim that the existing wrapper-based methods are inadequate for this task. Instead, we propose a new fam-ily of online proximal algorithms able to tackle many variants of MKL and group-LASSO, and for which we show regret, convergence, and gen-eralization bounds. Experiments on handwriting recognition and dependency parsing illustrate the success of the approach.

Multiple Gaussian Process Models
Cedric Archambeau, Francis Bach,

We consider a Gaussian process formulation of the multiple kernel learning problem. The goal is to select the convex combination of kernel matrices that best explains the data and by doing so improve the generalisation on unseen data. Sparsity in the kernel weights is obtained by adopting a hierarchical Bayesian approach: Gaussian process priors are imposed over the latent functions and generalised inverse Gaussians on their associated weights. This construction is equivalent to imposing a product of heavy-tailed process priors over function space. A variational inference algorithm is derived for regression and binary classification.

Multitask Multiple Kernel Learning (MT-MKL)
Chris Widmer, Nora C. Toussaint, Yasemin Altun, Gunnar Raetsch

The lack of sufficient training data is the limiting factor for many Machine Learning applications in Computational Biology. If data is available for several different but related problem domains, Multitask Learning algorithms can be used to learn a model based on all available information. However, combining information from several tasks requires careful consideration of the degree of similarity between tasks. We propse to use the recently published q-Norm Multiple Kernel Learning algorithm to simultaneously learn or refine the similarity matrix between tasks along with the Multitask Learning classifier by formulating the Multitask Learning problem as Multiple Kernel Learning. We demonstrate the performance of our method on two problems from Computational Biology. First, we show that our method is able to improve performance on a splice site dataset with given hierarchical task structure by refining the task relationships. Second, we consider an MHC-I dataset, for which we assume no knowledge about the degree of task relatedness. Here, we are able to learn the task similarity ab initio. Our framework is very general as it allows to incorporate prior knowledge about tasks relationships if available, but is also able to identify task similarities in absence of such prior information. Both variants show promising results in applications from Computational Biology.

Co-regularized Spectral Clustering with Multiple Kernels
Abhishek Kumar, Piyush Rai, Hal Daume III,

In this paper, we propose a spectral clustering algorithm exploiting two (or more) views of the data. Our algorithm takes a co-regularization based approach to clustering which enforces the clusterings across multiple views to agree with eachother. Since each view can be used to define a similarity graph over the data, our algorithm can also be considered as learning with multiple similarity graphs, or equivalently with multiple kernels. We propose an objective function that implicitly combines two (or more) kernels, and leads to an improved clustering performance. Experimental comparisons with a number of baselines on one synthetic and two real-world datasets establish the efficacy of our proposed approach.

POSTERS

Multiple Kernel Testing for SVM-based System Identification
Matthew Higgs, John Shawe-Taylor,

We apply methods of multiple kernel learning to the problem of system identification for multi-dimensional temporal data. Rather than building a full probabilistic model, we take a computationally simple approach that uses out of the box machine learning methods. We attempt to learn the covariance function of a stochastic process via multiple kernel learning. We achieve promising preliminary results and the work suggests an abundance of future theoretical work. We hope to draw on the theory of SVM methods to give a principled learning theory style description of system identification in stochastic processes.

Supervised and Localized Dimensionality Reduction from Multiple Feature Representations or Kernels
Mehmet Goenen, Ethem Alpaydin,

We propose a supervised and localized dimensionality reduction method that combines multiple feature representations or kernels. Each feature representation or kernel is used where it is suitable through a parametric gating model in a supervised manner for efficient dimensionality reduction and classification, and local projection matrices are learned for each feature representation or kernel. The kernel machine parameters, the local projection matrices, and the gating model parameters are optimized using an alternating optimization procedure composed of kernel machine training and gradient-descent updates. Empirical results on benchmark data sets validate the method in terms of classification accuracy, smoothness of the solution, and ease of visualization.

Multiple Kernel Learning for Efficient Conformal Predictions
Vineeth Balasubramanian, Shayok Chakraborty, Sethuraman Panchanathan,

The Conformal Predictions framework is a recent development in machine learning to associate reliable measures of confidence with results in classification and regression. This framework is founded on the principles of algorithmic randomness (Kolmogorov complexity), transductive inference and hypothesis testing. While the formulation of the framework guarantees validity, the efficiency of the framework depends greatly on the choice of the classifier and appropriate kernel functions or parameters. While this framework has extensive potential to be useful in several applications, the lack of efficiency can limit its usability. In this paper, we propose a novel Multiple Kernel Learning (MKL) methodology to maximize efficiency in the CP framework. This method is validated using the k-Nearest Neighbors classifier on a cardiac patient dataset, and our results show promise in using MKL to obtain efficient conformal predictors that can be practically useful.

Operator Multi-Task Gaussian Processes for Solving Differential Equations
Arman Melkumyan,

We consider a Gaussian process formulation of the multiple kernel learning problem. The goal is to select the convex combination of kernel matrices that best explains the data and by doing so improve the generalisation on unseen data. Sparsity in the kernel weights is obtained by adopting a hierarchical Bayesian approach: Gaussian process priors are imposed over the latent functions and generalised inverse Gaussians on their associated weights. This construction is equivalent to imposing a product of heavy-tailed process priors over function space. A variational inference algorithm is derived for regression and binary classification.

Learning Kernels via Margin-and-Radius Ratios
Kun Gai, Guangyun Chen, Changshui Zhang,

Most existing MKL approaches employ the large margin principle to learning kernels. However, we point out that the margin itself can not well describe the goodness of a kernel due to the negligence of the scaling. We use the ratio between the margin and the radius of the minimal enclosing ball of data in the feature space endowed with a kernel, to measure how good the kernel is, and propose a new scaling-invariant formulation for kernel learning. Our presented formulation can handle both linear and nonlinear combination kernels. In linear combination cases, it is also invariant not only to types of norm constraints on combination coefficients but also to initial scalings of basis kernels. By establishing the differentiability of a general type of multilevel optimal value functions, we present a simple and efficient gradient-based kernel learning algorithm. Experiments show that our approach significantly outperforms other state-of-art kernel learning methods.

Currency Forecasting using Multiple Kernel Learning with Financially Motivated Features
Tristan Fletcher, Zakria Hussain, John Shawe-Taylor,

Multiple Kernel Learning (MKL) is used to replicate the signal combination process that trading rules embody when they aggregate multiple sources of financial information when predicting an asset's price movements. A set of financially motivated kernels is constructed for the EURUSD currency pair and is used to predict the direction of price movement for the currency over multiple time horizons. MKL is shown to outperform each of the kernels individually in terms of predictive accuracy. Furthermore, the kernel weightings selected by MKL highlights which of the financial features represented by the kernels are the most informative for predictive tasks.

Numerical Mathematics Challenges in Machine Learning

http://numml.kyb.tuebingen.mpg.de/

LOCATION
Hilton: Diamond Head
Saturday, 7:30 - 10:30 and 3:30 - 6:30

Matthias W. Seeger mseeger@gmail.com
Ecole Polytechnique Federale de Lausanne

Suvrit Sra suvrit@gmail.com
Max Planck Institute for Biological Cybernetics

Abstract

Most machine learning (ML) methods are based on numerical mathematics (NM) concepts, from differential equation solvers over dense matrix factorizations to iterative linear system and eigen-solvers. For problems of moderate size, NM routines can be invoked in a black-box fashion. However, for a growing number of real-world ML applications, this separation is insufficient and turns out to be a limit on further progress.

The increasing complexity of real-world ML problems must be met with layered approaches, where algorithms are long-running and reliable components rather than stand-alone tools tuned individually to each task at hand. Constructing and justifying dependable reductions requires at least some awareness about NM issues. With more and more basic learning problems being solved sufficiently well on the level of prototypes, to advance towards real-world practice the following key properties must be ensured: scalability, reliability, and numerical robustness.

By inviting numerical mathematics researchers with interest in *both* numerical methodology *and* real problems in applications close to machine learning, we will probe realistic routes out of the prototyping sandbox. Our aim is to strengthen dialog between NM, signal processing, and ML. Speakers are briefed to provide specific high-level examples of interest to ML and to point out accessible software. We will initiate discussions about how to best bridge gaps between ML requirements and NM interfaces and terminology.

The workshop will reinforce the community's awakening attention towards critical issues of numerical scalability and robustness in algorithm design and implementation. Further progress on most real-world ML problems is conditional on good numerical practices, understanding basic robustness and reliability issues, and a wider, more informed integration of good numerical software. As most real-world applications come with reliability and scalability requirements that are by and large ignored by most current ML methodology, the impact of pointing out tractable ways for improvement is substantial.

Target audience:
Our workshop is targeted towards practitioners from NIPS, but is of interest to numerical linear algebra researchers as well.

SCHEDULE

7:30-7:45	Opening remarks
7:45-9:00	A Personal Journey: From Signals and Systems to Graphical Models Alan Willsky
9:00-9:30	Coffee Break
9:30-10:00	Efficient space-variant blind deconvolution Stefan Harmeling
10:00-10:30	TBA Matthias Seeger
15:30-16:20	Algebraic Multigrid Methods with Applications to Data Analysis Dan Kushnir
16:20-17:00	Hierarchical Preconditioners for Computer Vision Problems Richard Szeliski
17:00-17:15	Coffee Break
17:15-18:15	Linear Algebra and Machine Learning of Large Informatics Graphs Michael Mahoney
18:15-18:30	Closing remarks and discussion

INVITED SPEAKERS

Invited talk: A Personal Journey: From Signals and Systems to Graphical Models

Alan S. Willsky, Massachusetts Institute of Technology

This talk outlines a meandering line of research, started in 1988, that began with some signal processors and control theorists trying to make statistical sense of the emerging field of wavelet analysis and, to the speaker's surprise, moved into areas that certainly take advantage of his S&S/CT background but evolved into topics in a variety of fields involving efficient numerical methods for large-scale data assimilation and mapping and, eventually, a rapprochement with graphical models and machine learning.

The talk will touch on our early work on multiresolution tree models (motivated by wavelets but only indirectly relevant to them), the way control theorists think of inference and approximate modeling/stochastic realization, with at least one application that rings of the way a machine learning researcher might build a model – but not a mathematical physicist! I'll provide (finally) a real rapprochement with wavelets and then turn to approximate inference on loopy graphs. The first approach builds on an idea used in the solution of PDEs, namely nested dissection, but with some machine learning twists, and some control-theoretic stability issues (showing how control might have a few things to provide to inference algorithm designers).

I will then turn to a topic again related to the ways in which numerical linear algebraists think about solving sparse systems of equations, but in the context of graphical models, this leads to the idea of walk-sum analysis, a surprisingly useful (and at least I think intuitive) idea. Walk-sum analysis then allows us to say some fairly strong things about a variety of iterative algorithms (generally known as either Richardson iterations, Jacobi iterations, or Gauss-Seidel iterations), including adaptive algorithms to guide iterations for fast convergence. Walk-sum analysis is also key in another approach with linear algebraic interpretations, involving so-called feedback vertex sets.

I will also touch on an alarmingly accurate method for computing variances in graphical models that involves using a low-rank approximation to the identity matrix(!) and then return to multiresolution models but now looking at models motivated by two quite different classes of numerical algorithms: multigrid algorithms and multipole algorithms. These algorithms motivate two quite different classes of models, the latter of which requires the introduction of what we refer to as conjugate graphs.

If I have any time and energy left, I will comment on some prospective topics.

Invited talk: Hierarchical Preconditioners for Computer Vision Problems

Richard Szeliski, Microsoft Research Redmond

Invited talk: Linear Algebra and Machine Learning of Large Informatics Graphs

Michael Mahoney, Stanford University

Very large informatics graphs such as large social and information networks typically have properties that render many popular machine learning and data analysis tools largely inappropriate. While this is problematic for these applications, it also suggests that these graphs may be useful as a test case for the development of new algorithmic tools that may then be applicable much more generally. Many of the popular machine learning and data analysis tools rely on linear algebra, and they are typically used by calling traditional numerical linear algebra code as a black box. After briefly reviewing some of the structural properties of large social and information networks that are responsible for the inapplicability of traditional linear algebra and machine learning tools, I will describe several examples of ''new linear algebr'' and ''new machine learning'' that arise from the analysis of such informatics graphs. These new directions involve looking ''inside'' the black box, and they place very different demands on the linear algebra than are traditionally placed by numerical, scientific computing, and small-scale machine learning applications.

Invited talk: Algebraic Multigrid Methods with Applications to Data Analysis

Dan Kushnir, Yale University

Multigrid solvers for solving linear problems (structured or unstructured) are based on iterating between two processes:

- Classical relaxation schemes, which are generally slow to converge but fast to ''smooth'' the error function.
- Approximating the ''smooth'' error function on a coarser level (a coarser grid or, more generally, a smaller graph, typically having one quarter to one half the number of variables). This is done by solving coarse-level equations derived from the fine-level system and the residuals of its current approximate solution, and then interpolating that coarse-level solution to correct the fine-level approximation. The solution of the coarse level equations is obtained by using recursively the same two processes, employing still coarser levels.

The most basic Multigrid solvers are the Geometric solvers (GMG) in which the problem is defined on a geometric grid. Of more interest, in the context of Machine learning, are the Algebraic Multigrid (AMG) solvers designed for unstructured problems. In this tutorial I will introduce the AMG technique, and its adoption for solving numerical problems in the context of Data Analysis. In particular, I will describe several efficient AMG eigensolvers that can be used within spectral graph methods such clustering, image segmentation, and dimensionality reduction. It turns out that the Algebraic Multigrid coarsening procedure itself can yield efficient and accurate algorithms for data clustering and image segmentation without any explicit solution of the related equations.

Invited talk: Efficient space-variant blind deconvolution

Stefan Harmeling, MPI for Biological Cybernetics

Blur in photos due to camera shake, blur in astronomical image sequences due to atmospheric turbulence, and blur in magnetic resonance imaging sequences due to object motion are examples of blur that can not be adequately described as a space-invariant convolution, because such blur varies across the image. In this talk, we present a class of linear transformations, that are expressive enough for space-variant blurs, but at the same time especially designed for efficient matrix-vector-multiplications. Successful results on the above-mentioned examples demonstrate the practical significance of our approach.

Optimization for Machine Learning

http://opt.kyb.tuebingen.mpg.de/

LOCATION
Westin: Emerald A
Friday, 7:30 - 10:30 and 3:30 - 6:30

Suvrit Sra suvrit@gmail.com
Max Planck Institute for Biological Cybernetics

Sebastian Nowozin Sebastian.Nowozin@microsoft.com
Microsoft Research Cambridge

Stephen Wright swright@cs.wisc.edu
University of Wisconsin-Madison

Abstract

Our workshop focuses on optimization theory and practice that is relevant to machine learning. This proposal builds on precedent established by two of our previously well-received NIPS workshops: (@NIPS*08) http://opt2008.kyb.tuebingen.mpg.de/ (@NIPS*09) http://opt.kyb.tuebingen.mpg.de/ Both these workshops had packed (often overpacked) attendance almost throughout the day. This enthusiastic reception reflects the strong interest, relevance, and importance enjoyed by optimization in the greater ML community. One could ask why does optimization attract such continued interest? The answer is simple but telling: optimization lies at the heart of almost every ML algorithm. For some algorithms textbook methods suffice, but the majority require tailoring algorithmic tools from optimization, which in turn depends on a deeper understanding of the ML requirements. In fact, ML applications and researchers are driving some of the most cutting-edge developments in optimization today. The intimate relation of optimization with ML is the key motivation for our workshop, which aims to foster discussion, discovery, and dissemination of the state-of-the-art in optimization, especially in the context of ML. The workshop should realize its aims by: * Providing a platform for increasing the interaction between researchers from optimization, operations research, statistics, scientific computing, and machine learning; * Identifying key problems and challenges that lie at the intersection of optimization and ML; * Narrowing the gap between optimization and ML, to help reduce rediscovery, and thereby accelerating new advances.

ADDITIONAL BACKGROUND AND MOTIVATION
Previous talks at the OPT workshops have covered frameworks for convex programs (D. Bertsekas), the intersection of ML and optimization, especially in the area of SVM training (S. Wright), large-scale learning via stochastic gradient methods and its tradeoffs (L. Bottou, N. Srebro), exploitation of structured sparsity in optimization (Vandenberghe), and randomized methods for extremely large-scale convex optimization (A. Nemirovski). Several important realizations were brought to the fore by these talks, and many of the dominant ideas will appear in our book (to be published by MIT Press) on Optimization for Machine learning. Given the above background it is easy to acknowledge that optimization is indispensable to machine learning. But what more can we say beyond this obvious realization? The ML community's interest in optimization continues to grow. Invited tutorials on optimization will be presented this year at ICML (N. Srebro) and NIPS (S. Wright). The traditional ‘‘point of contact’’ between ML and optimization -SVM -continues to be a driver of research on a number of fronts. Much interest has focused recently on stochastic gradient methods, which can be used in an

SCHEDULE

7:30-7:40	Opening remarks
7:40-8:30	Limited-memory quasi-Newton and Hessian-free Newton methods for non-smooth optimization Mark Schmidt
8:30-8:50	Information-theoretic lower bounds on the oracle complexity of sparse convex optimization Alekh Agarwal
8:50--9:00	Poster spotlights
9:00-9:30	Coffee Break
9:30-10:20	Efficiency of Quasi-Newton Methods on Strictly Positive FunctionsYurii Nesterov
10:20--10:30	Poster spotlights
10:30-3:30	Long Break
3:30-4:20	Safe Feature Elimination in Sparse Learning Laurent El Ghaoui
4:20-4:40	Hierarchical Classification via Orthogonal Transfer Dengyong Zhou
4:40-5:00	An Optimization Based Framework for Dynamic Batch Mode Active Learning Shayok Chakraborty
4:50-5:05	Poster spotlights
5:00-5:30	Coffee break
5:30-5:50	Augmenting Dual Decomposition for MAP Inference Andre Martins
5:50-6:10	An Incremental Subgradient Algorithm for Approximate MAP Estimation in Graphical Models Jeremy Jancsary
6:10-	Poster session

online setting and in settings where data sets are extremely large and high accuracy is not required. Regularized logistic regression is another area that has produced a recent flurry of activity at the intersection of the two communities. Many aspects of stochastic gradient remain to be explored, for example, different algorithmic variants, customizing to the data set structure, convergence analysis, sampling techniques, software, choice of regularization and tradeoff parameters, parallelism. There also needs to be a better understanding of the limitations of these methods, and what can be done to accelerate them or to detect when to switch to alternative strategies. In the logistic regression setting, use of approximate second-order information has been shown to improve convergence, but many algorithmic issues remain. Detection of combined effects predictors (which lead to a huge increase in the number of variables), use of group regularizers, and dealing with the need to handle very large data sets in real time all present challenges.

To avoid becoming lopsided, in our workshop we will also admit the 'not particularly large scale' setting, where one has time to wield substantial computational resources. In this setting, high-accuracy solutions and deep understanding of the lessons contained in the data are needed. Examples valuable to MLers may be exploration of genetic and environmental data to identify risk factors for disease; or problems dealing with setups where the amount of observed data is not huge, but the mathematical models are complex.

INVITED SPEAKERS

Limited-memory quasi-Newton and Hessian-free Newton methods for non-smooth optimization
Mark Schmidt, University of British Columbia

Limited-memory quasi-Newton and Hessian-free Newton methods are two workhorses of unconstrained optimization of high-dimensional smooth objectives. However, in many cases we would like to optimize a high-dimensional unconstrained objective function that is non-smooth due to the presence of a 'simple' non-smooth regularization term. Motivated by problems arising in estimating sparse graphical models, in this talk we focus on strategies for extending limited-memory quasi-Newton and Hessian-free Newton methods for unconstrained optimization to this scenario. We first consider two-metric (sub-) gradient projection methods for problems where the regularizer is separable, and then consider proximal Newton-like methods for group-separable and non-separable regularizers. We will discuss several applications where sparsity-encouraging regularizers are used to estimate graphical model parameters and/or structure, including the estimation of sparse, blockwise-sparse, and structured-sparse models.

Efficiency of Quasi-Newton Methods on Strictly Positive Functions
Yurii Nesterov, Catholic University of Louvain (UCL)

In this talk we consider a new class of convex optimization problems, which admit faster black-box optimization schemes. For analyzing their rate of convergence, we introduce a notion of mixed accuracy of an approximate solution, which is a convenient generalization of the absolute and relative accuracies. We show that for our problem class, a natural Quasi-Newton method is always faster than the standard gradient method. At the same time, after an appropriate normalization, our results can be extended onto the general convex unconstrained minimization problems.

Safe Feature Elimination in Sparse Learning
Laurent El Ghaoui, University of California, Berkeley

We describe methods that allow to quickly eliminate variables (features) in supervised learning problems involving a convex loss function and a l1-norm penalty, leading to a potentially substantial reduction in the number of variables prior to running the supervised learning algorithm. The methods are not heuristic: they only eliminate features that are guaranteed to be absent after solving the learning problem. Our framework applies to a large class of problems, including support vector machine classification, logistic regression and least-squares. The complexity of the feature elimination step is negligible compared to the typical computational effort involved in the sparse supervised learning problem: it grows linearly with the number of features times the number of examples, with much better count if data is sparse. We apply our method to data sets arising in text classification and observe a dramatic reduction of the dimensionality, hence in computational effort required to solve the learning problem, especially when very sparse classifiers are sought. Our method allows to immediately extend the scope of existing algorithms, allowing us to run them on data sets of sizes that were out of their reach before.

ACCEPTED CONTRIBUTIONS

Information-theoretic lower bounds on the oracle complexity of sparse convex optimization
Alekh Agarwal, Peter Bartlett and Martin Wainwright, University of California, Berkeley
Pradeep K. Ravikumar, University of Texas at Austin

Hierarchical Classification via Orthogonal Transfer
Dengyong Zhou, Microsoft Research Lin Xiao, Microsoft Research Mingrui Wu, Microsoft Corporation

An Optimization Based Framework for Dynamic Batch Mode Active Learning
Shayok Chakraborty, Vineeth Balasubramanian and Sethuraman Panchanathan,
Arizona State University

Augmenting Dual Decomposition for MAP Inference
Andre F. T. Martins, Mario A. T. Figueiredo and Noah A. Smith, Instituto Superior Tecnico, Carnegie Mellon University

An Incremental Subgradient Algorithm for Approximate MAP Estimation in Graphical Models
Jeremy Jancsary, Austrian Research Institute for Artificial Intelligence,
Gerald Matz, Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology
Harald Trost, Section for Artificial Intelligence, Medical University of Vienna

Practical Application of Sparse Modeling: Open Issues and New Directions

https://sites.google.com/site/nips10sparsews/

LOCATION
Hilton: Sutcliffe A
Friday, 7:30 - 10:30 and 3:30 - 6:30

Irina Rish rish@us.ibm.com
IBM T.J. Watson Research Center

Alexandru Niculescu-Mizil alexnic@gmail.com
IBM T.J. Watson Research Center

Guillermo Cecchi gcecchi@us.ibm.com
IBM Research

Aurelie C. Lozano aclozano@us.ibm.com
IBM T.J. Watson Research Center

Abstract

Abstract
Sparse modeling is a rapidly developing area at the intersection of statistics, machine-learning and signal processing, focused on the problem of variable selection in high-dimensional datasets. Selection (and, moreover, construction) of a small set of highly predictive variables is central to many applications where the ultimate objective is to enhance our understanding of underlying physical, biological and other natural processes, beyond just building accurate 'black-box' predictors.

Recent years have witnessed a flurry of research on algorithms and theory for sparse modeling, mainly focused on l1-regularized optimization, a convex relaxation of the (NP-hard) smallest subset selection problem. Examples include sparse regression, such as Lasso and its various extensions, such as Elastic Net, fused Lasso, group Lasso, simultaneous (multi-task) Lasso, adaptive Lasso, bootstrap Lasso, etc.), sparse graphical model selection, sparse dimensionality reduction (sparse PCA, CCA, NMF, etc.) and learning dictionaries that allow sparse representations. Applications of these methods are wide-ranging, including computational biology, neuroscience, image processing, stock market prediction and social network analysis, as well as compressed sensing, an extremely fast-growing area of signal processing.

However, is the promise of sparse modeling realized in practice? It turns out that, despite the significant advances in the field, a number of open issues remain when sparse modeling meets real-life applications. Below we only mention a few of them (see the workshop website for a more detailed discussion): stability of sparse models; selection of the ''right'' regularization parameter/ model selection; finding ''right'' representation (dictionary learning); handling structured sparsity; evaluation of the results, interpretability.

We would like to invite researchers working on methodology, theory and especially applications of sparse modeling to share their experiences and insights into both the basic properties of the methods, and the properties of the application domains that make particular methods more (or less) suitable. Moreover, we plan to have a brainstorming session on various open issues, including (but not limited to) the ones mentioned above, and hope to come up with a set of new research directions motivated by problems encountered in practical applications.

SCHEDULE

7:30-7:45	Opening remarks
7:45-8:20	Realistic sparse modeling in genetics and other domains (invited) Saharon Rosset
8:20-8:35	Sparse recovery for Protein Mass Spectrometry data Martin Slawski and Matthias Hein
8:35-8:50	A Sparsity Constraint for Topic Models - Application to Temporal Activity Mining Jagannadan Varadarajan, Remi Emonet, Jean-Marc Odobez
8:50-9:25	Coffee Break/Poster Setup/Poster Session
9:25-10:00	Stability and reproducibility in fMRI analysis(invited) Stephen Strother
10:00-10:30	Discussion: Beyond sparsity - interpretability, stability, reproducibility and other evaluation criteria for sparse modeling in practice
10:30-3:30	Lunch break
3:45-4:20	On High-Dimensional Sparse Structured Input-Output Models, with Applications to Genome-Phenome Association Analysis of Complex Diseases (invited) Eric P. Xing
4:20-4:35	Fast Hard Thresholding with Nesterov's Gradient Method Volkan Cevher and Sina Jafarpour
4:35-4:50	Sparse Exponential Family Latent Variable Models Shakir Mohamed, Katherine Heller, Zoubin Ghahramani
4:50-5:25	Coffee Break/Poster Session
5:25-6:00	Sparsity: from theory to practice? (invited) Francis Bach
6:00-6:30	Discussion: novel types of structure/ regularizers motivated by practical applications? Dictionary learning: beyond linear? What are key open problems raised by practical applications of sparse modeling?

We welcome submissions on various practical aspects of sparse modeling, specifically focusing on the following questions: Does sparse modeling provide a meaningful interpretation of interest to domain experts? What other properties of the sparse models are desirable for better interpretability? How robust is the method with respect to various type of noise in the data? What type of method (e.g., combination of regularizers) is best-suited for a particular application and why? What is the best representation allowing for sparse modeling in your domain? How do you find such a representation efficiently? How is the model evaluated with respect to its structure-recovery quality?

INVITED SPEAKERS

Sparsity: from theory to practice?

Francis Bach, INRIA-ENS

Sparse methods have recently led to a lot of work in several communities, namely machine learning, statistics and signal processing. A significant portion of the effort has been dedicated to the theoretical analysis of such methods and their attractive behavior in high-dimensional settings. However, it can be argued that there are fewer positive results regarding practical applications of such methods. In this talk, I will explore some potential reasons for this gap and some ideas to reduce it, in particular through structured sparsity and dictionary learning techniques.

Realistic sparse modeling in genetics and other domains

Saharon Rosset, Tel-Aviv University

We review some of the considerations in determining which types of sparse models are realistic in real domains, and estimating these models. Key concepts are the dependence between the covariates, the level of expected sparsity, and the type of connections expected between the covariates and the response. A prime example is genome-wide association studies (GWAS), where the typical simplistic approach of estimating the effect of one SNP at a time is consistent with assuming sparse signal on independent covariates -often a reasonable assumption, sometimes problematic. We discuss possible enhancements of GWAS methtodology. We then discuss the range of sparse modeling approaches, and their limitations, focusing on the difficulties of using Lasso and its derivatives on highly correlated covariates, and possible remedies.

Stability and reproducibility in fMRI analysis

Stephen Strother, U. of Toronto

Analysis of fMRI studies of the brain involves a high-dimensional variable selection problem with 10s-100s thousands of spatial voxels/variables and 10s-100s of 3D image volumes/observations. The goal is typically to interpret the extracted salient spatial voxels as activation patterns representing the neural basis of cognition. Therefore their sparsity, stability and reproducibility are of critical scientific concern. I will briefly review these issues as they have appeared in the neuroimaging literature and introduce a split-half, subsampling approach dubbed NPAIRS that we have used for the last decade (Strother et al., NI, 2002; CompStat2010). It provides prediction and pattern reproducibility metrics together with a statistical thresholding mechanism for salient voxel selection, and has links to recent results in subsampling stability theory and sparse variable selection. In particular, we have used NPAIRS to explore empirical tradeoffs between prediction, sparsity and particular stability/reproducibility metrics for regularized linear discriminant models in the fMRI setting. In general we find that comprehensive regularization is much more important than the particular model used, and that prediction and various sparsity constraints (e.g., elastic net) provide somewhat inconsistent objectives for reliable variable selection. Finally our recent results suggest that in typical group fMRI analyses the iid assumption underlying cross-validation, bootstrap and subsampling may be violated making fMRI analysis even more challenging than it currently appears. Joint work with Nathan Churchill and Grigori Yourganov, Rotman Institute/University of Toronto, and Peter Rassmusen, Kristoffer Madsen, and Lars Kai Hansen, Danish Research Centre for MR and Danish Technical University.

On High-Dimensional Sparse Structured Input-Output Models, with Applications to Genome-Phenome Association Analysis of Complex Diseases

Eric P. Xing, CMU

Genome-wide association (GWA) studies have recently become popular as a tool for identifying genetic variables that are responsible for increased disease susceptibility. A modern statistical method for approaching this problem is through model selection (or structure estimation) of Structured Input-Output Regression Models (SIORM) fitted on genetic and phenotypic variation data across a large number of individuals. The inputs of such models bear rich structure, because the cause of many complex disease syndromes involves the complex interplay of a large number of genomic variations that perturb disease-related genes in the context of a regulatory network. Likewise, the outputs of such model are also structured, as patient cohorts are routinely surveyed for a large number of traits such as hundreds of clinical phenotypes and genome-wide profiling for thousands of gene expressions that are interrelated. A Structured Input-Output Regression Model nicely captures all these properties, but raises severe computational and theoretical challenge on consistent model identification. In this talk, I will present models, algorithms, and theories that learn Sparse SIORMs of various kinds in very high dimensional input/output space, with fast and highly scalable optimization procedures, and strong statistical guarantees. I will demonstrate application of our approaches to a number of complex GWA scenarios, including associations to trait networks, to trait clusters, to dynamic traits, under admixed populations, and with epistatic effects. This is joint work with Seyoung Kim, Mladen Kolar, Seunghak Lee, Xi Chen, and Kriti Puniyani and Judie Howrylak.

ACCEPTED CONTRIBUTIONS

Sparse non-linear denoising of fMRI: Performance and pattern reproducibility
Trine J. Abrahamsen, Lars Kai Hansen,

Non-linear embedding improves mental state decoding in fMRI. Here we investigate the possibility of non¬linear denoising by kernel Principal Component Analysis (PCA). The main challenge is the mapping of denoised feature space points back into input space, also referred to as ‘‘the pre-image problem’’. Since the feature space mapping is typically not bijective, pre-image estimation is inherently illposed. In many applications, including functional MRI data, it is of interest to find a sparse signal, hence we investigate sparse reconstruction by Lasso regularization. It is found that sparse estimation provides better predictability and a more reproducible pre-image. Using the combined prediction/ reproducibility metric we may optimize the degree of sparsity and the non-linearity of the kernel embedding.

Fast Hard Thresholding with Nesterov's Gradient Method
Volkan Cevher, Sina Jafarpour,

We provide an algorithmic framework for structured sparse recovery which unifies combinatorial optimization with the non-smooth convex optimization framework by Nesterov [1, 2]. Our algorithm, dubbed Nesterov iterative hard-thresholding (NIHT), is similar to the algebraic pursuits (ALPS) in [3] in spirit: we use the gradient information in the convex data error objective to navigate over the non-convex set of structured sparse signals. Experiments show however that NIHT can empirically outperform ALPS and other state-of-the-art convex optimization-based algorithms in sparse recovery.

An Efficient Proximal Gradient Method for General Structured Structured Sparse Learning
Xi Chen, Qihang Lin, Seyoung Kim, Jaime G. Carbonell, Eric P. Xing

We study the problem of learning regression models regularized by a structured sparsity-inducing penalty which encodes prior structural information. We consider two widely adopted structures as motivating examples: (1) group structure (potentially overlapping) which is encoded via a l1/l2 mixed norm penalty;
(2) graph structure which is encoded in the graph-guided fusion penalty. For both structures, due to the non-separability of the penalties, developing an efficient optimization method has proven quite a challenge. In this paper, we propose a general proximal gradient method which can efficiently solve a class of structured sparse learning problem based on a smooth convex loss and a structured penalties, including our motivating examples. Moreover, our method achieves faster convergence than subgradient method; and is more efficient and scalable than interior point method for second order cone programming, as shown in simulation studies.

Sparse Latent Semantic Analysis
Xi Chen, Yanjun Qi, Bing Bai, Qihang Li, Jaime G. Carbonell,

Latent semantic analysis (LSA), as one of the most popular unsupervised dimension reduction tools, has a wide range of applications in text mining and information retrieval. The key idea of LSA is to learn a projection matrix that maps the high dimensional vector space representations of documents to a lower dimensional latent space, i.e. so called latent topic space. In this paper, we propose a new model called Sparse LSA, which produces a sparse projection matrix via the l1 regularization. Compared to the traditional LSA, Sparse LSA selects only a small number of relevant words for each topic and hence provides a compact representation of topic-word relationships. Moreover, Sparse LSA is computationally very efficient with much less memory usage for storing the projection matrix. Furthermore, we propose the nonnegative Sparse LSA as an extension of Sparse LSA.We conduct experiments on several benchmark datasets and compare Sparse LSA and its extension with several widely used methods, e.g. LSA, Sparse Coding and LDA. Empirical results suggest that Sparse LSA achieves similar performance gains to LSA, but is more efficient in projection computation, storage, and also well explain the topic-word relationships.

An Application of Compressed Sensing to Pulse Doppler Radar
Tom Diethe, Graeme Smith, Zakria Hussain, David Roi Hardoon, John Shawe-Taylor,

The present study shows how the Analogue to Digital Conversion (ADC) sampling rate in a digital radar can be reduced—without reduction in waveform bandwidth—through the use of Compressed Sensing (CS). Real radar data is used to show that through use of chirp or Gabor dictionaries and Basis Pursuit (BP) the Analogue to Digital Conversion (ADC) sampling frequency can be reduced by a factor of 128, to under 1 mega sample per second, while the waveform bandwidth remains 40 MHz. The error on the reconstructed fast-time samples is small enough that accurate range-profiles and range-frequency surfaces can be produced.

Union Support Recovery in Multi-task Learning
Mladen Kolar, John Lafferty, Larry Wasserman,

We sharply characterize the performance of different penalization schemes for the problem of selecting the relevant variables in the multi-task setting. Previous work focuses on the regression problem where conditions on the design matrix complicate the analysis. A clearer and simpler picture emerges by studying the Normal means model. This model, often used in the field of statistics, is a simplified model that provides a laboratory for studying complex procedures.

Sparse Exponential Family Latent Variable Models
Shakir Mohamed, Katherine Heller, Zoubin Ghahramani,

We present a fully developed sparse latent factor model based on Bayesian learning with spike-and-slab priors. This approach has the desirable property that it allows actual zeroes in the latent representation to obtained, while also allowing for a principled approach to handling model uncertainty. Sparsity is combined with an exponential family generalization to allow for modeling of data that may be binary, categorical or real valued. We challenge commonly held beliefs surrounding MCMC by showing that the sparse Bayesian methods can be run in the same time as optimization methods and provide improved results.

Low-Rank Matrix Factorization Using Regularized Logistic Regression
John Paisley, David Blei,

We present a matrix factorization model for dyadic data consisting of counts of successes out of a variable number of trials. We model the problem as a matrix of binomial random variables with underlying probabilities that are logistic transformations of a low-rank, real-valued matrix. We impose sparsity constraints to learn the rank of this underlying matrix using MAP inference. We perform experiments on data from Major League Baseball (years 2000–2009).

Sparse Bayesian Learning for Hidden Markov Models
George Saon, Jen-Tzung Chien,

We introduce Bayesian sensing hidden Markov models (BS-HMMs) to represent sequential data based on a set of state-dependent basis vectors. By incorporating the prior density of sensing weights, the relevance of a feature vector to different bases is determined by the corresponding precision parameters. The BS-HMM parameters, consisting of the basis vectors and the precision matrices of sensing weights and reconstruction errors, are jointly estimated by maximizing the likelihood function, which is marginalized over weight priors. We derive recursive solutions for stable implementation of the three parameters, which are interpreted by the maximum a posteriori estimates of sensing weights. A hybrid dictionary learning and sparse representation is accordingly established for sequential pattern recognition. We show the benefits of BS-HMMs through some experiments on large vocabulary continuous speech recognition.

Sparse recovery for Protein Mass Spectrometry data
Martin Slawski, Matthias Hein,

Extraction of peptide masses from a raw protein mass spectrum (MS) is a challenging problem in computational biology. We discuss several structural characteristics of MS data, notably non-negativity and heteroscedastic noise that make standard sparse recovery methods exhibit inferior performance as compared to more targeted approaches. In particular, we suggest non-negative least squares followed by thresholding as a simple, user-friendly alternative, which yields very promising results in practice.

Sparse Preference Learning
Evgeni Tsivtsivadze, Tom Heskes,

We propose a novel sparse preference learning/ranking algorithm. Our algorithm approximates the true utility function by a weighted sum of basis functions using the squared loss on pairs of data points, and is a generalization of the matching pursuit method. It can operate both in a supervised and a semi-supervised setting and allows efficient search for multiple, near-optimal solutions. In our experiments we demonstrate that the proposed algorithm outperforms several state-ofthe-art learning methods when taking into account unlabeled data and performs comparably in a supervised learning scenario, while providing sparser solution.

The $l_{1,p}$ Group-Lasso for Multi-Task Learning
Julia E. Vogt, Volker Roth,

The Group-Lasso is a well known tool for joint regularization in machine learning methods. While the l1,2 version has been studied in detail, there are still open questions regarding the uniqueness of solutions and the efficiency of algorithms for other l1,p variants. We characterize conditions for uniqueness of solutions for the l1,p Group-Lasso for all p with 1 ≤ p ≤∞. We present a simple test for uniqueness of solutions, and we derive a highly efficient active set algorithm that can deal with input dimensions in the millions. We compare the prediction performance of all variants of the Group-Lasso in a multi-task learning setting where the aim is to solve many learning problems in parallel which are coupled via the Group-Lasso constraint. We show that the l1,2 version consistently outperforms all other l1,p counterparts in terms of prediction accuracy.

A Sparsity Constraint for Topic Models -Application to Temporal Activity Mining
Jagannadan Varadarajan, R´emi Emonet, Jean-Marc Odobez,

We address the mining of sequential activity patterns from document logs given as word-time occurrences. We achieve this using topics that model both the co-occurrence and the temporal order in which words occur within a temporal window. Discovering such topics, which is particularly hard when multiple activities can occur simultaneously, is conducted through the joint inference of the temporal topics and of their starting times, allowing the implicit alignment of the same activity occurrences in the document. A current issue is that while we would like topic starting times to be represented by sparse distributions, this is not achieved in practice. Thus, in this paper, we propose a method that encourages sparsity, by adding regularization constraints on the searched distributions. The constraints can be used with most topic models (e.g. PLSA, LDA) and lead to a simple modified version of the EM standard optimization procedure. The effect of the sparsity constraint on our activity model and the robustness improvement in the presence of difference noises have been validated on synthetic data. Its effectiveness is also illustrated in video activity analysis, where the discovered topics capture frequent patterns that implicitly represent typical trajectories of scene objects.

Sparse Structured Dictionary Learning for Brain Resting-State Activity Modeling
Gaël Varoquaux, Alexandre Gramfort, Bertrand Thirion, Rodolphe Jenatton, Guillaume Obozinski, Francis Bach,

In this paper, we introduce a generative model to study brain resting-state time series. These signals are represented as linear combinations of latent spatial maps, which we obtain via matrix factorization techniques developed for dictionary learning. In particular, we propose to learn the spatial components with specific structural constraints, e.g, small localized clusters of voxels, which can be achieved with sparsity-inducing regularization schemes recently used for dictionary learning. While brain resting-state time-series are generally the object of exploratory data analysis, our model provides a natural framework for model selection and quantitative evaluation. We show that our approach yields improved estimates as assessed by the likelihood on unseen data, while exhibiting interpretable spatial components, that match known areas of interest in the brain.

Predictive Models in Personalized Medicine

https://sites.google.com/site/personalmedmodels

LOCATION
Hilton: Sutcliffe B
Saturday, 7:30 - 10:30 and 3:30 - 6:30

Faisal Farooq f.farooq@siemens.com
Siemens Medical Solutions, USA Inc.

Glenn Fung glenn.fung@siemens.com
Siemens Healthcare

Romer Rosales romer.rosales@siemens.com
Siemens Healthcare

Shipeng Yu shipeng.yu@siemens.com
Siemens Healthcare

Jude Shavlik shavlik@cs.wisc.edu
University of Wisconsin at Madison

Balaji R. Krishnapuram balaji.krishnapuram@siemens.com
Siemens Medical Solutions USA, Inc.

Raju Kucherlapati rkucherlapati@partners.org
Department of Medicine, Harvard University

SCHEDULE

7:30-8:00	Personalized Medicine Overview R. Bharat Rao
8:00-8:45	TBA Erwin Bottinger
8:45-9:00	Personalizing Cancer Therapy via Machine Learning Sriraam Natarajan
9:00-9:30	Poster Session with Coffee Break
9:30-10:15	Learning Predictive Models with Potential to Aid in Clinical Decisions David Page
10:15-10:30	Virus Genotype Prediction for Personalized Amortized Treatment Optimization Betty Y. Cheng
3:30-4:15	TBA
4:15-4:30	Machine Learning for Guiding Neurosurgery in Epilepsy Justin Dauwels
4:30-4:45	Tumour heterogeneity is a central obstacle to predictive models of patient prognosis in lung cancer Gerald Quon
4:45-5:00	Learning individual and population level traits from clinical temporal data Suchi Saria
5:00-5:15	Poster Session with Coffee Break
5:15-6:00	TBA
6:00-6:30	Panel discussion

Abstract

The purpose of this cross-discipline workshop is to bring together machine learning and healthcare researchers interested in problems and applications of predictive models in the field of personalized medicine. The goal of the workshop will be to bridge the gap between the theory of predictive models and the applications and needs of the healthcare community. There will be exchange of ideas, identification of important and challenging applications and discovery of possible synergies. Ideally this will spur discussion and collaboration between the two disciplines and result in collaborative grant submissions. The emphasis will be on the mathematical and engineering aspects of predictive models and how it relates to practical medical problems. Although related in a broad sense, the workshop does not directly overlap with the fields of Bioinformatics and Biostatistics. Although, predictive modeling for healthcare has been explored by biostatisticians for several decades, this workshop focuses on substantially different needs and problems that are better addressed by modern machine learning technologies. For example, how should we organize clinical trials to validate the clinical utility of predictive models for personalized therapy selection? The traditional bio-statistical approach for running trials on a large cohort of homogeneous patients would not suffice for the new paradigm and new methods are needed. On the other hand bioinformatics typically deals with the analysis of genomic and proteomic data to answer questions of relevance to basic science. For example, identification of sequences in the genome corresponding to genes, identification of gene regulatory networks etc. This workshop does not focus on issues of basic science; rather, we focus on predictive models that combine all available patient data (including imaging, pathology, lab, genomics etc.) to impact point of care decision making. More recently, as part of American Re-investment and Recovery Act (ARRA), the US government set aside significant amounts of grant funds for cross-disciplinary research in use of information technology in improving health outcomes, quality of care and selection of therapy. The workshop program will consist of presentations by invited speakers from both machine learning and personalized medicine fields and by authors of extended abstracts submitted to the workshop. In addition, there will be a slot for a panel discussion to identify important problems, applications and synergies between the two scientific disciplines.

INVITED SPEAKERS

Learning Predictive Models with Potential to Aid in Clinical Decisions
David Page, University of Wisconsin at Madison

ACCEPTED CONTRIBUTIONS

Personalizing Cancer Therapy via Machine Learning
Marty Tenenbaum, CollabRx Inc. Alan Fern, Oregon State University Lise Getoor, University of Maryland at College Park Michael Littman, Rutgers University Vikash Mansinghka, Navia Systems Sriraam Natarajan, University of Wisconsin at Madison David Page, University of Wisconsin at Madison Jeff Shrager, CollabRx Inc. Yoram Singer, Google Inc. Prasad Tadepalli, Oregon State University

Virus Genotype Prediction for Personalized Amortized Treatment Optimization
Betty Y. Cheng, Carnegie Mellon University
Jaime G. Carbonell, Carnegie Mellon University

Machine Learning for Guiding Neurosurgery in Epilepsy
Justin Dauwels, Nanyang Technological University, Singapore
Emad Eskandar, Massachusetts General Hospital
Andy Cole, Massachusetts General Hospital
Dan Hoch, Massachusetts General Hospital
Rodrigo Zepeda, Massachusetts General Hospital
Sydney S. Cash, Massachusetts General Hospital

Tumour heterogeneity is a central obstacle to predictive models of patient prognosis in lung cancer
Gerald Quon, University of Toronto
Syed Haider, Ontario Institute for Cancer Research
Ang Cui, Banting and Best Department of Medical Research
Paul C. Boutros, Ontario Institute for Cancer Research
Quaid Morris, Banting and Best Department of Medical Research

Learning individual and population level traits from clinical temporal data
Suchi Saria, Stanford University
Daphne Koller, Stanford University
Anna Penn, Stanford University

Robust Statistical Learning

http://www.cs.utexas.edu/~sai/robustml

LOCATION
Hilton: Mt. Currie South
Friday, 7:30 - 10:30 and 3:30 - 6:30

Pradeep K. Ravikumar pradeepr@cs.utexas.edu
University of Texas, Austin

Constantine Caramanis caramanis@mail.utexas.edu
The University of Texas

Sujay Sanghavi sanghavi@mail.utexas.edu
University of Texas, Austin

Abstract

At the core of statistical machine learning is to infer conclusions from data, typically using statistical models that describe probabilistic relationships among the underlying variables. Such modeling allows us to make strong predictions even from limited data by leveraging specific problem structure. However on the flip side, when the specific model assumptions do not exactly hold, the resulting methods may deteriorate severely. A simple example: even a few corrupted points, or points with a few corrupted entries, can severely throw off standard SVD-based PCA. The goal of this workshop is to investigate this ‘‘robust learning’’ setting where the data deviate from the model assumptions in a variety of different ways. Depending on what is known about the deviations, we can have a spectrum of approaches: (a) Dirty Models: Statistical models that impose ‘‘clean’’ structural assumptions such as sparsity, low-rank etc. have proven very effective at imposing bias without being overly restrictive. A superposition of two (or more) such clean models can provide a method that is also robust. For example, approximating data by the sum of a sparse matrix and a low-rank one leads to PCA that is robust to corrupted entries. (b) Robust Optimization: Most statistical learning methods implicitly or explicitly have an underlying optimization problem. Robust optimization uses techniques from convexity and duality, to construct solutions that are immunized from some bounded level of uncertainty, typically expressed as bounded (but otherwise arbitrary, i.e., adversarial) perturbations of the decision parameters. (c) Classical Robust Statistics; Adversarial Learning: There has been a large body of work on classical robust statistics, which develops estimation methods that are robust to misspecified modeling assumptions in general, and do not model the outliers specifically. While this area is still quite active, it has a long history, with many results developed in the 60s, 70s and 80s. There has also been significant recent work in adversarial machine learning. Thus, we see that while there has been a resurgence of robust learning methods (broadly understood) in recent years, it seems to be largely coming from different communities that rarely interact: (classical) robust statistics, adversarial machine learning, robust optimization, and multi-structured or dirty model learning. It is the aim of this workshop to bring together researchers from these different communities, and identify common intuitions underlying such robust learning methods. Indeed, with increasingly high-dimensional and ‘‘dirty’’ real world data that do not conform to clean modeling assumptions, this is a vital necessity.

SCHEDULE

7.30-7.40	Introduction, Welcome
7.40-8.15	TBA Constantine Caramanis
8.15-9.05	Robustness in statistics and related issues in machine learning Peter Bickel
9.05-9.20	Coffee Break
9.20-9.55	TBA Emmanuel Candes
9.55-10.30	Dirty Statistical Models Pradeep Ravikumar/Sujay Sanghavi
10.30-10.50	Spotlights
15.10-15.35	TBA
15.35-16.10	TBA Nina Balcan
16.10-16.45	TBA Chiranjib Bhattacharyya
16.45-17.00	Coffee Break
17.00-17.35	TBA Laurent El Ghaoui
17.35-18.10	Panel Discussion

INVITED SPEAKERS

Robustness in statistics and related issues in machine learning
Peter Bickel, UC Berkeley

I will survey some of the major formulations of robustness in statistical theory and optimization problems that they lead to. I will then relate them to questions in machine learning and possible extensions.

Robustness Models in Learning
Laurent El Ghaoui, UC Berkeley

Robust optimization is a methodology traditionally applied to handle measurement errors or uncertainty about future data in a decision problem. It has a natural place in machine learning applications. In fact, classical learning models such as SVMs or penalized regression can be interpreted in terms of robustness models. This interpretation leads to new insights into important aspects such as consistency and sparsity of solutions. Building on these connections, we show that robust optimization ideas can lead to new, more performant learning models that take into account data structure, such as those arising in text classification, where the data is often boolean or integer-valued.

Robustness models can also be used in the context of a large-scale learning problem with dense data, based on the idea of thresholding and taking into account the resulting error when solving the learning problem.

Robust Principal Component Analysis?
Emmanuel Candes, Stanford University

This talk is about a curious phenomenon. Suppose we have a data matrix, which is the superposition of a low-rank component and a sparse component. Can we recover each component individually? We prove that under some suitable assumptions, it is possible to recover both the low-rank and the sparse components exactly by solving a very convenient convex program called Principal Component Pursuit; among all feasible decompositions, simply minimize a weighted combination of the nuclear norm and of the L1 norm. This suggests the possibility of a principled approach to robust principal component analysis since our methodology and results assert that one can recover the principal components of a data matrix even though a positive fraction of its entries are arbitrarily corrupted. This extends to the situation where a fraction of the entries are missing as well. We discuss an algorithm for solving this optimization problem emphasizing the suitability of this approach for large scale problems, and present applications in the area of video surveillance, where our methodology allows for the detection of objects in a cluttered background, and in the area of face recognition, where it offers a principled way of removing shadows and specularities in images of faces.
Joint work with X. Li, Y. Ma and J. Wright.

Robust Hierarchical Clustering
Maria Florina Balcan, Georgia Institute of Technology

One of the most widely used techniques for data clustering is agglomerative clustering. Such algorithms have been long used across many different fields ranging from computational biology to social sciences to computer vision in part because their output is easy to interpret. Unfortunately, it is well known, however, that many of the classic agglomerative clustering algorithms are not robust to noise. In this talk, we describe new robust algorithms for bottom-up agglomerative clustering. We show that these algorithms can be used to cluster accurately in cases where the data satisfies a number of natural properties and where the traditional agglomerative algorithms fail. We also describe how to adapt our algorithms to the inductive setting where our given data is only a small random sample of the entire data set.

Learning, Robustness, and Structure
Constantine Caramanis, University of Texas at Austin

Robust optimization is a paradigm that uses ideas from convexity and duality, to immunize solutions of convex problems to bounded uncertainty in the parameters of the problem. Machine learning is fundamentally about making decisions under uncertainty, and optimization has long been a central tool; thus at a high level there is no surprise that robust optimization should have a role to play. We discuss some recent developments in the use of robust optimization and machine learning, highlighting its use as a methodology to construct algorithms with special properties. We illustrate these ideas using Support Vector Machines and Lasso. We show that both can be re-interpreted as robust optimization problems. Then, using robustness, one can obtain consistency, and sparsity, among other properties, therefore showing that SVMs are consistent because they are robust, and Lasso is sparse because it is robust.

wstalktitleDirty Statistical Models PradeepRavikumar SujaySanghavi, University of Texas at Austin

ACCEPTED CONTRIBUTIONS

High-Dimensional Robust Structure Learning of Ising Models on Sparse Random Graphs
Animashree Anandkumar, Vincent Y. F. Tan and Alan S. Willsky, MIT

Learning from Noisy Data under Distributional Assumptions
Nicolo Cesa-Bianchi, Universita degli Studi di Milano
Shai Shalev-Shwartz, The Hebrew University
Ohad Shamir, Microsoft Research

Robust Matrix Decomposition with Outliers
Daniel Hsu and Tong Zhang, Rutgers University
Sham M. Kakade, University of Pennsylvania

Regularization via Statistical Stability
Chinghway Lim, UC Berkeley Bin Yu, UC Berkeley

Weighted Neighborhood Linkage
Pramod Gupta, Georgia Institute of Technology
Maria Florina Balcan, Georgia Institute of Technology

Tensors, Kernels, and Machine Learning

http://csmr.ca.sandia.gov/~dfgleic/tkml2010/

LOCATION
Westin: Nordic
Friday, 7:30 - 10:30 and 3:30 - 6:30

Tamara G. Kolda tgkolda@sandia.gov
Sandia National Laboratories

Vicente Malave vmalave@cogsci.ucsd.edu
University of California, San Diego

David F. Gleich dfgleic@sandia.gov
Sandia National Laboratories

Johan Suykens johan.suykens@esat.kuleuven.be
K.U. Leuven

Marco Signoretto marco.signoretto@gmail.com
KULeuven

Andreas Argyriou argyriou@ttic.edu
Toyota Technological Institute

Abstract
Tensors are a generalization of vectors and matrices to high dimensions. The goal of this workshop is to explore the links between tensors and information processing. We expect that many problems in, for example, machine learning and kernel methods can benefit from being expressing as tensor problems; conversely, the tensor community may learn from the estimation techniques commonly used in information processing and from some of the kernel extensions to nonlinear models. On the other hand, standard tensor-based techniques can only deliver multi-linear models. As a consequence, they may suffer from limited discriminative power. A properly defined kernel-based extension might overcome this limitation. The statistical machine learning community has much to offer on different aspects such as learning (supervised, unsupervised and semi-supervised) and generalization, regularization techniques, loss functions and model selection.

The goal of this workshop is to promote the cross-fertilization between machine learning and tensor-based techniques. This workshop is appropriate for anyone who wishes to learn more about tensor methods and/or share their machine learning or kernel techniques with the tensor community; conversely, we invite contributions from tensor experts seeking to use tensors for problems in machine learning and information processing.
We hope to discuss the following topics:

- Applications using tensors for information processing (e.g. image recognition, EEG, text analysis, diffusion weighted tensor imaging, etc.) as well as the appropriateness of tensors models for various information processing tasks.
- Specialized tensor decompositions that may be of interest for informatics processing (e.g. nonnegative factorizations, specialized objective functions or constraints, symmetric factorizations, handling missing data, handling special types of noise).
- Information processing techniques that have connections to tensor representations and factorizations, such as nonlinear kernel methods, multi-task learning, specialized learning algorithms that can be adapted to tensor factorizations.
- Theoretical questions of interest in applying tensor information processing methods (e.g. questions surrounding tensor rank, extension of nuclear norm to tensors).

SCHEDULE

7:30-8:15	Applications of Tensor Decomposition in Data Mining and Machine Learning (Invited) Morten Mrup
8:15-8:35	Tensor Products of Kernel Modules Omar de la Cruz with Susan Holmes, Alex Barnett, Hua Teng
8:35-8:55	Non-Redundant Tensor Decomposition Deniz Erdogmus with Olexiy Kyrgyzov
8:55-9:10	Coffee Break and Posters
9:10-9:40	Sparse Methods for Matrix Factorization (Invited) Francis Bach
9:40-10:00	Sum of Separable Functions in Machine Learning Jochen Garcke
10:00-10:30	Discussion -- When are tensor factorizations the right model for the data? What is the difference between these and closely related methods?
10:30-3:30	Recess and Posters
3:30-4:00	Cholesky Decomposable Tensors and Multilinear Mercer Kernels (Invited) Lek-Heng Lim
4:00-4:20	Making Tensor Factorizations Robust to Non-Gaussian Noise Eric Chi with Tamara G. Kolda
4:20-4:40	Probabilistic Tensor Factorization and Model Selection Cyril Goutte with Massih-Reza Amini
4:40-5:00	Three-Way DEDICOM for Relational Learning Maximilian Nickel with Volker Tresp
5:00-5:20	Coffee Break and Posters
5:20-5:40	Convex Multilinear Estimation and Operatorial Representations Marco Signoretto with Lieven De Lathauwer, Johan A. K. Suykens
5:40-6:00	On the Extension of Trace Norm to Tensors Ryota Tomioka with Kohei Hayashi, Hisashi Kashima
6:00-6:30	Discussion -- What theoretical questions need to be answered to advance the utility of these methods? Are these methods of practical to use in concert with current tensor and machine learning technology?

INVITED SPEAKERS

Applications of Tensor Decomposition in Data Mining and Machine Learning

Morten Mørup, Technical University of Denmark

Tensor (multi-way) decomposition is becoming an important tool for data mining and machine learning, fueled by the computational power of modern computers enabling researchers to analyze large scale tensorial structured data that only a few years ago would have been impossible. Tensor factorizations have several advantages over (two-way) matrix factorizations for unsupervised learning/exploratory data analysis including uniqueness of the optimal solution and component identification even when most of the data is missing. Furthermore, tensor decomposition techniques explicitly exploit the multi-way structure that is lost when collapsing some of the modes of the tensor in order to analyze the data by regular matrix factorization approaches. Tensor decomposition is being applied to new fields every year and there is no doubt the future will bring many exciting new applications. The aim of this tutorial is to introduce the basic concepts of tensor decompositions and to demonstrate some of the many benefits and challenges of modeling data multi-way for a wide variety of problem domains. The talk will focus primarily on tensor decomposition of neuroimaging data such as EEG and fMRI where an important challenge is to extract the consistent, reproducible patterns of activation across multiple trials, subjects and/or conditions while facing low signal to noise ratios. Finally, it will be discussed how Bayesian methods from machine learning can be used to address import challenges in tensor decomposition such as regularization tuning and model selection.

Sparse Methods for Matrix Factorization

Francis Bach, INRIA/Ecole Normale Superieure

Matrix factorization problems occur in many situations, like image processing, audio processing, bioinformatics, multi-task or unsupervised learning. In this talk, I will present recent contributions and some associated open problems, in particular on structured dictionary learning and low-rank decompositions. (joint work with R. Jenatton, J. Mairal, G. Obozinski, J. Ponce).
Cholesky Decomposable Tensors and Multilinear Mercer Kernels Lek-Heng Lim, University of Chicago We will discuss two topics in this talk to illustrate multilinearity in neuroimaging and in learning respectively.

A real quadratic form $f(x)= x^T \mathrm{Ax}$ may be expressed as a sum of squares of linear forms $f(x)= \Sigma_k (\beta_{k1}x_1 + \beta_{kn}x_n)^2$ if the associated symmetric matrix A is positive semidefinite. We will examine a generalization of this notion: degree-p forms that are expressible as a sum of p-powers of linear forms $f(x)= \Sigma_k (\beta_{k1}x_1 + \beta_{kn}x_n)^p$ (p even). In this case it is not sufficient that the symmetric tensor associated with f be positive semidefinite although that is a necessary condition. We call such tensors Cholesky decomposable tensors and show that they are surprisingly well-behaved. The sum-of-powers decomposition is then a higher-order analogue of the spectral decomposition of a positive semidefinite matrix. It arises naturally in diffusion MRI and gives the most accurate method for extracting nerve fibers crossing. This is joint work with T. Schultz. A multilinear Mercer kernel (defined here for trilinear) $K(x, y, z)$ is one that is symmetric, continuous, and satisfies a reproducing property that there exists an orthonormal basis $\{\varphi k \in L^2(X, \mu)\}$ such that (i) for any $f(x)= \Sigma_k C_k \varphi_k (z)$ one must have

$$\int K(x, y, z)\, f(x) = \Sigma/k\ Ck\ \varphi k\ (x)\ \varphi k\ (y)$$

and (ii) for any $g(x, y)= \Sigma_k C_k \varphi_k (x) \varphi_k (y)$, one must have

$$\iint K(x, y, z)\, g(x, y)\, d\mu(x)\, d\mu(y) = \Sigma/k\ Ck\ \varphi(z)$$

It is easy to show that such kernels must exist and that

$$K(x, y, z) = \Sigma/k\ \varphi k\ (x)\ \varphi k\ (y)\ \varphi k\ (z)$$

We will see that these share many common properties with the usual (bilinear) Mercer kernel and that a multilinear SVM may be defined based on this notion.

POSTER CONTRIBUTIONS

Compression of Electroencephalograms using Tensor Decompositions

Justin Dauwels, Nanyang Technological University, Singapore
Srinivasan Kannan, Indian Institude of Technology Madras
Ramasubba Reddy M., RIKEN
Andrzej Cichocki, RIKEN

Function-Valued Reproducting Kernel Hilbert Spaces and Applications

Hachem Kadri, INRIA Lille -Nord Europe Philippe Preux, INRIA Lille/Universite de Lille Emmanuel Duflos, INRIA Lille/Ecole Centrale de Lille St´ephane Canu, INSA de Rouen Manual Davy, Ecole Centrale de Lille

Multiple-biomarker tensor analysis for tuberculosis lineage identification

Cagri Ozcaglar, Rensselaer Polytechnic Institute Amina Shabbeer, Rensselaer Polytechnic Institute Scott Vandenberg, Siena College Buelent Yener, Rensselaer Polytechnic Institute Kristin P. Bennett, Rensselaer Polytechnic Institute

Probabilistic Latent Tensor Factorization

Y. Kenan Yilmaz, Boi University gazic
A. Taylan Cemgil, Boi University

Transfer Learning Via Rich Generative Models.

http://www.mit.edu/~rsalakhu/workshop_nips2010/index.html

LOCATION
Westin: Emerald A
Saturday, 7:30 - 10:30 and 3:30 - 6:30

Ruslan R. Salakhutdinov rsalakhu@mit.edu
Massachusetts Institute of Technology

Ryan P. Adams rpa@cs.toronto.edu
University of Toronto

Josh Tenenbaum jbt@mit.edu
Massachusetts Institute of Technology

Zoubin Ghahramani zoubin@eng.cam.ac.uk
University of Cambridge & CMU

Tom Griffiths tomgriffiths@berkeley.edu
University of California, Berkeley

Abstract
Intelligent systems must be capable of transferring previously-learned abstract knowledge to new concepts, given only a small or noisy set of examples. This transfer of higher order information to new learning tasks lies at the core of many problems in the fields of computer vision, cognitive science, machine learning, speech perception and natural language processing.

Over the last decade, there has been considerable progress in developing cross-task transfer (e.g., multi-task learning and semi-supervised learning) using both discriminative and generative approaches. However, many existing learning systems today can not cope with new tasks for which they have not been specifically trained. Even when applied to related tasks, trained systems often display unstable behavior. More recently, researchers have begun developing new approaches to building rich generative models that are capable of extracting useful, high-level structured representations from high-dimensional sensory input. The learned representations have been shown to give promising results for solving a multitude of novel learning tasks, even though these tasks may be unknown when the generative model is being trained. A few notable examples include learning of Deep Belief Networks, Deep Boltzmann Machines, deep nonparametric Bayesian models, as well as Bayesian models inspired by human learning.

''Learning to learn'' new concepts via rich generative models has emerged as one of the most promising areas of research in both machine learning and cognitive science. Although there has been recent progress, existing computational models are still far from being able to represent, identify and learn the wide variety of possible patterns and structure in real-world data. The goal of this workshop is to assess the current state of the field and explore new directions in both theoretical foundations and empirical applications.

SCHEDULE

7:30-7:50	Opening remarks
7:50-8:30	Bayesian Learning of Deep Sparse Generative Models Zoubin Ghahramani
8:30-9:10	Learning Rich Generative Models for Human Cognition Josh Tenenbaum
9:10-9:30	Discussion
9:10-9:30	Coffee Break
9:30-10:10	Transfer Learning in Language: We've Got a Long Way to Go Hal Daume III
10:10-10:30	Contributed Talk: Modeling Transfer Learning and Taxonomy Induction with the Hierarchical Dirichlet Process Kevin Canini
10:30--	Poster Session
3:30-4:10	Visual Learning via Topics, Transformations, and Trees Erik Sudderth
4:10-4:50	Learning to Detect Rare Objects using Frequent Objects Antonio Torralba
4:50-5:05	Transfer Learning Challenge
5:05-5:30	Coffee break
5:30-6:10	Deep Learning Generalizing from Related but Out-of-distribution Examples Yoshua Bengio
6:10-6:30	Panel discussion / Closing Remarks

INVITED SPEAKERS

Bayesian Learning of Deep Sparse Generative Models
Zoubin Ghahramani, University of Cambridge

I will start with some thoughts on transfer learning and probabilistic modelling, focusing on several ways in which one can formalise transfer learning and relate it to general probabilistic modelling. I will then describe our work (joint with Ryan Adams and Hanna Wallach) on learning the structure of deep sparse directed graphical models. Unusually for the deep learning school, our work takes a fully Bayesian approach to the problem and develops a nonparametric methodology for learning the model structure (depth, width, connectivity, and unit type) from the data. Rather than imposing strong prior knowledge the Bayesian methodology is used to remove many arbitrary choices in the model architecture and let the data do (most of) the speaking for itself. I will describe several applications to unsupervised learning from image data sets and try to relate it all back to transfer learning with generative models.

Learning Rich Generative Models for Human Cognition
Josh Tenenbaum, MIT

Transfer Learning in Language: We've Got a Long Way to Go Hal Daume III, University of Maryland, College Park Human language is messy, and machine learning has done a lot to tame this messiness. There are tons of facets to language processing, and while the common approach is to run a bunch of component systems in a pipeline, there is mounting evidence that this is a bad idea. Enter transfer learning and multitask learning. Unfortunately for language processing people, but fortunately for us, there are tons of open-ended problems in transfer learning for language due to the sorts of data and annotations that we have easy access to in the language domain. This talk will highlight some successful attempts to use transfer learning (generative and otherwise) in language, but will also talk a good deal about what we, language folks, want from you, machine learning folks, in the future.

Modeling Transfer Learning and Taxonomy Induction with the Hierarchical Dirichlet Process
Kevin Canini, UC Berkeley

Many categorization models and algorithms treat categories in isolation, learning and representing each category independently from the others. However, in settings where a learner must simultaneously learn about multiple categories, it seems natural and beneficial to allow information to be shared between categories. In such cases, algorithms which introduce dependencies between the representations of different categories could increase learning speed and accuracy by avoiding having to relearn redundant information appearing in multiple categories. Modeling these dependent structures is a type of transfer learning and can occur whenever there are significant similarities between the structures and/or contents of categories. We describe two cases where this phenomenon occurs: in the sequential learning of partially overlapping categories, and in the simultaneous learning of a taxonomy of completely overlapping categories. We propose a Bayesian generative model --the hierarchical Dirichlet process (HDP) --as a general categorization algorithm which exhibits transfer learning capabilities for both of these cases. Drawing on the large body of work in studying human categorization and the proficiency of people in quickly learning complex category structures, we evaluate the HDP using experiments with human learners.

Visual Learning via Topics, Transformations, and Trees
Erik Sudderth, Brown University

Learning to Detect Rare Objects using Frequent Objects
Antonio Torralba, MIT

Deep Learning Generalizing from Related but Out-of-distribution Examples Yoshua Bengio, University of Montreal Recent theoretical and empirical work in statistical machine learning has demonstrated the potential of learning algorithms for deep architectures, i.e., function classes obtained by composing multiple levels of representation. The hypothesis evaluated here is that intermediate levels of representation, because they can be shared across tasks and examples from different but related distributions, can yield more benefits. Comparative experiments were performed on a large-scale handwritten character recognition setting with 62 classes (upper case, lower case, digits), using both a multi-task setting and perturbed examples in order to obtain out-of-distribution examples. The results agree with the hypothesis, and also show that a deep learner did beat previously published results and reached human-level performance. In another set of experiments involving text categorization, we show that deep learners can generalize better across domains, i.e., generalize to domain B while being trained on domain A with no example from domain B, presumably because they have extracted higher-level abstractions that happen to be shared across both domains.

ACCEPTED CONTRIBUTIONS

Deep Generative Models for Bio-sequences
Joseph Bockhorst,

Modeling Transfer Learning and Taxonomy Induction with the Hierarchical Dirichlet Process
Kevin Canini, Tom Griffiths,

Coping with New User Problems: Transfer Learning in Accelerometer-Based Human Activity Recognition
Hirotaka Hachiya, Masashi Sugiyama, Naonori Ueda,

Cross-View Transfer Learning for Automatic Speech Recognition
Jing Huang, Xiaodong Cui, Jen-Tzung Chien,

Feature-Preserving Embeddings for Topic Transfer
Peter Krafft, Sridhar Mahadevan,

Learning novel visual concepts from few examples
Brenden Lake, Jason Gross, Ruslan R. Salakhutdinov, Josh Tenenbaum,

Blind Domain Transfer for Named Entity Recognition using Generative Latent Topic Models
Ramesh Nallapati, Mihai Surdeanu, Christopher D. Manning,

Multitask Learning using Mixture of Factor Analyzers
Piyush Rai, Hal Daume III,

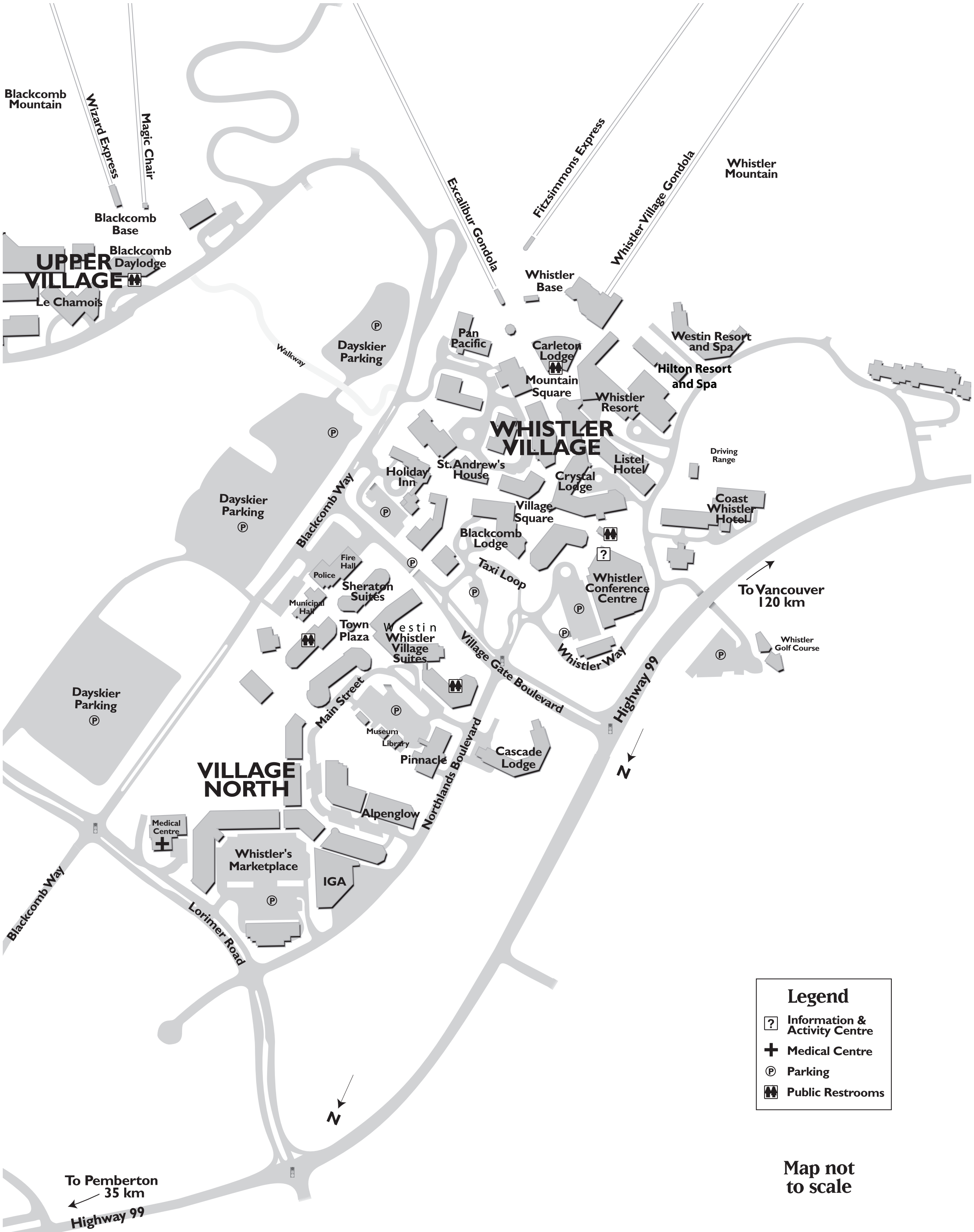
Random guillotine partitions of the Cantor cube and statistical applications to relational data
Dan Roy, Yee Whye Teh,

Transfer learning via Semi-Supervised Multi-task Gaussian Processes
Grigorios Skolidis, Guido Sanguinetti,

Gated Boltzmann Machine for Recognition under Occlusion
Yichuan Tang,

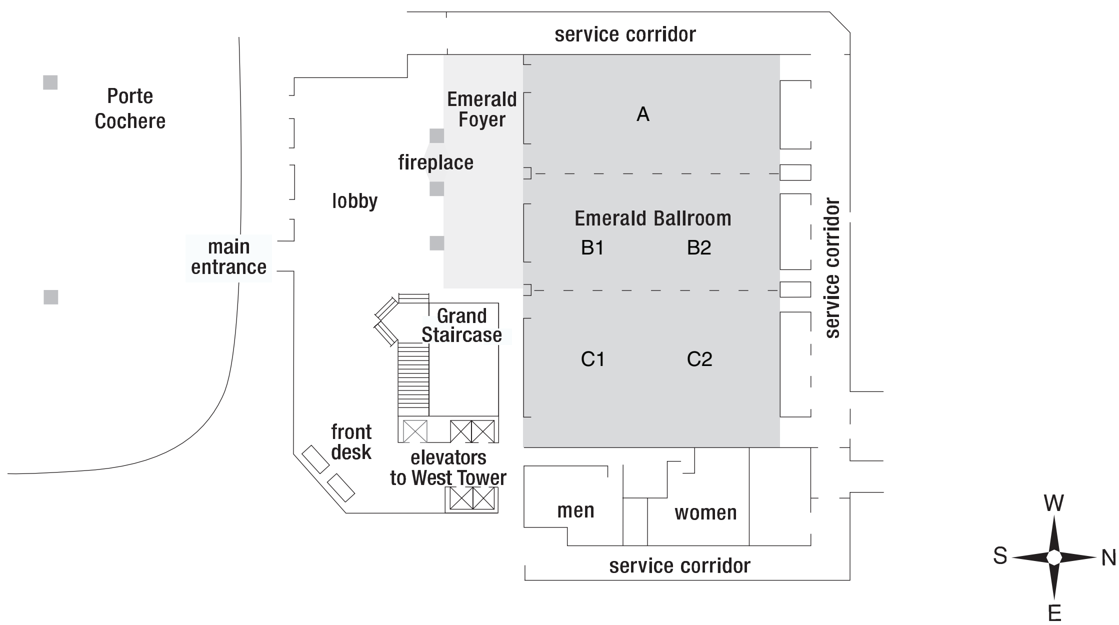
Transferring Learning via Learned Transformations
Christopher Wilson, Dan Ventura,

Whistler Map

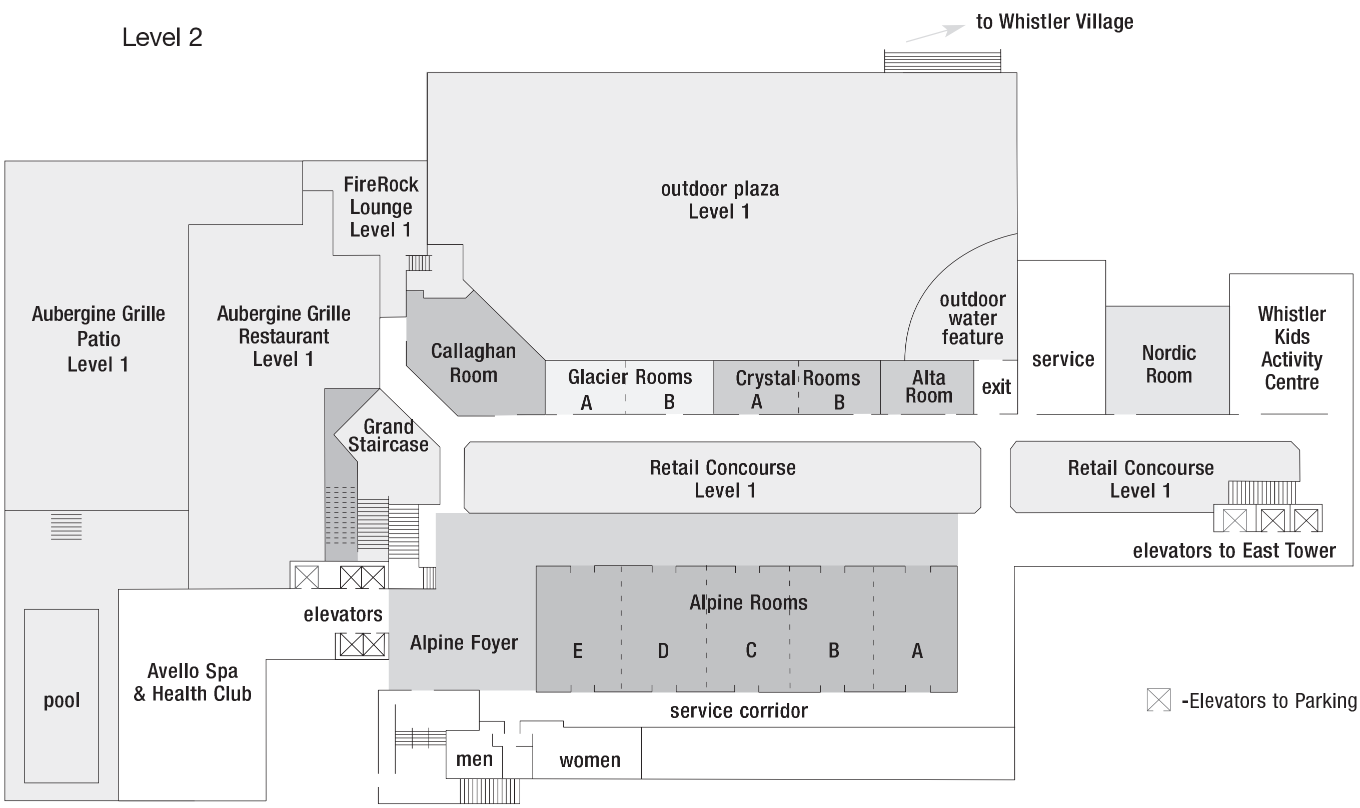


Westin Resort Workshop Rooms

Lobby Level

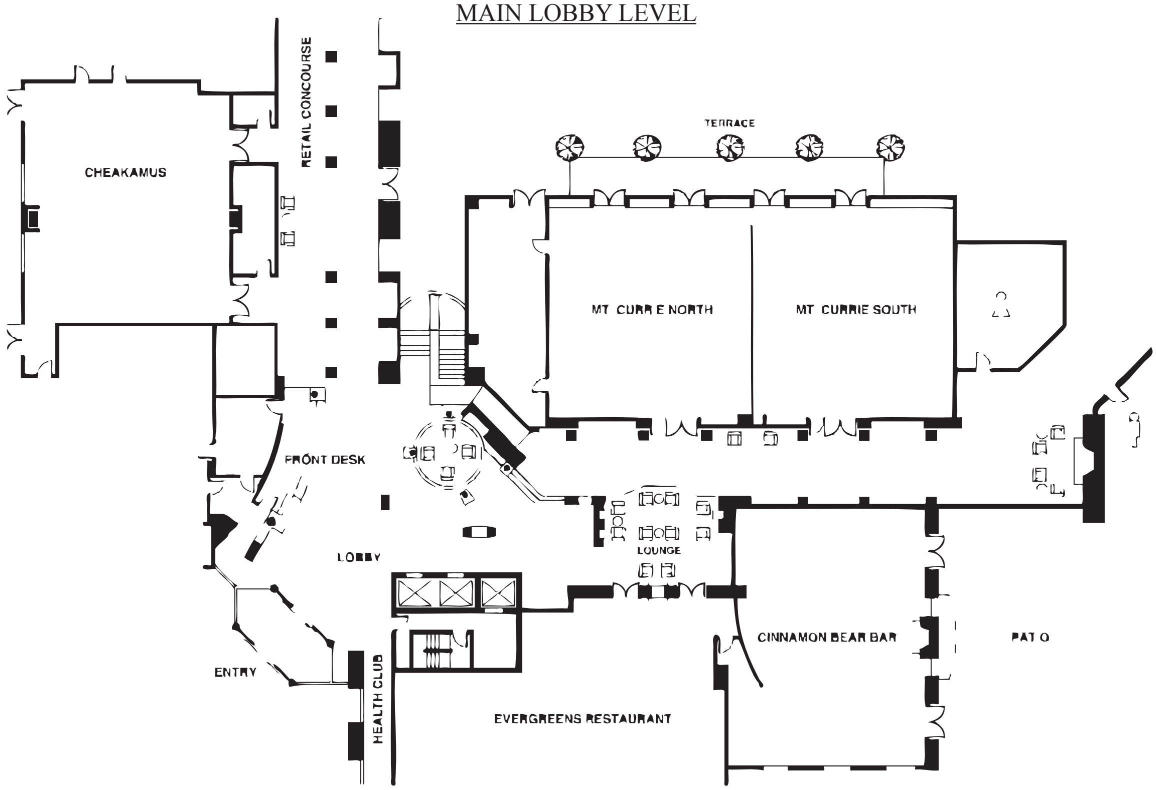


Level 2

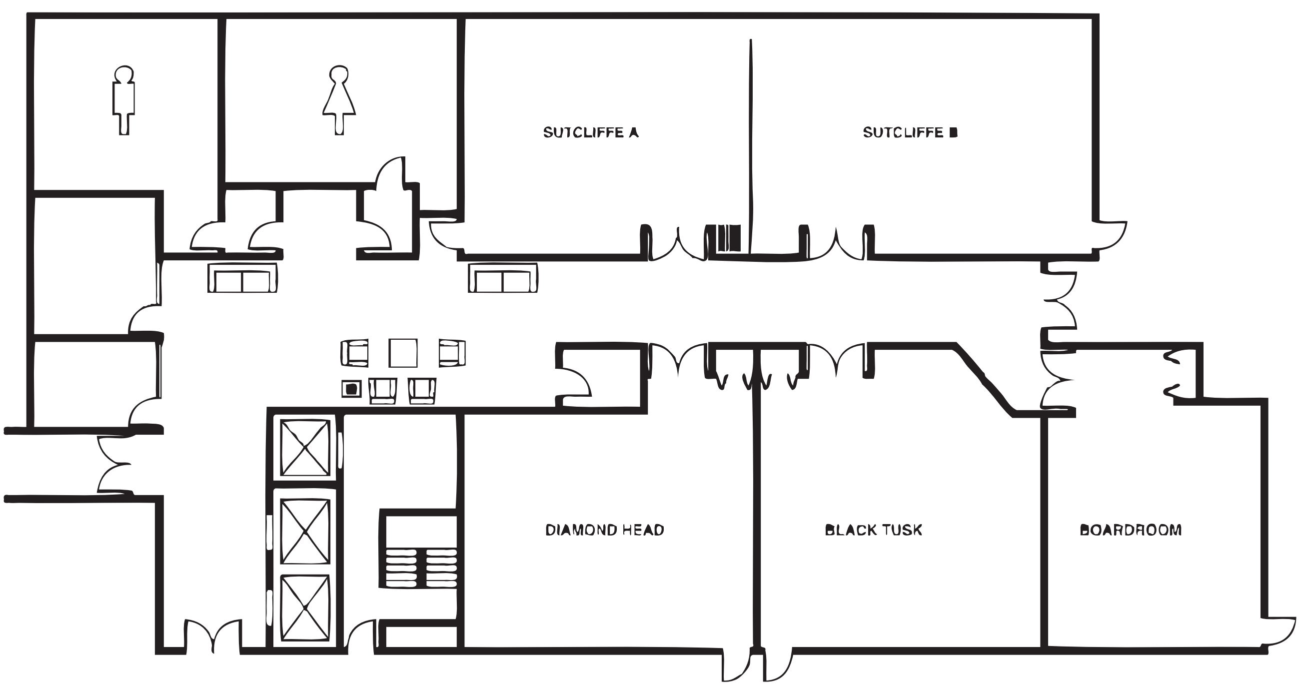


Hilton Resort Workshop Rooms

Hilton
Whistler Resort
<u>MAIN LOBBY LEVEL</u>



LOWER LEVEL



AUTHOR INDEX

AUTHOR INDEX

AUTHOR INDEX

NOTES

N I
2010
P S

N I
2010
P S

N I
2010
P S