NIPS 2009

NEURAL INFORMATION PROCESSING SYSTEMS CONFERENCE

Mini Symposia & Workshops

TUTORIALS
December 7, 2009
Hyatt Regency
Vancouver, BC, Canada

CONFERENCE SESSIONS
December 7-10, 2009
Hyatt Regency
Vancouver, BC, Canada

MINI SYMPOSIA
December 10, 2009
Hyatt Regency
Vancouver, BC, Canada

WORKSHOP
December 11-12, 2009
The Westin Resort & Spa
The Hilton Whistler Resort & Spa
Whistler, BC, Canada

Neural Information Processing Systems Foundation

N I P S 2 0 0 9

Mini Symposia & Workshops

TUTORIALS
December 7, 2009
Hyatt Regency
Vancouver, BC, Canada

CONFERENCE SESSIONS
December 7-10, 2009
Hyatt Regency
Vancouver, BC, Canada

MINI SYMPOSIA
December 10, 2009
Hyatt Regency
Vancouver, BC, Canada

WORKSHOP
December 11-12, 2009
The Westin Resort & Spa
The Hilton Whistler Resort & Spa
Whistler, BC, Canada

Sponsored by the Neural Information Processing System Foundation, Inc

There are 5 Mini Symposia and 28 workshops covering a rather eclectic range of topics from assistive technologies to combinatorics to neuroscience to robotics to sustainability to vision science to bioinformatics to learning theory to internet related problems. They provide an exciting preview for future trends in Neural Information Processing Systems and in this way they complement the main conference.

Neural Information Processing Systems Foundation

Contents

Organizing Committee . . . 7
Program Committee . . . 7
NIPS Foundation Offices and Board Members . . . 8
Sponsors . . . 9
Core Logistics Team . . . 9

Schedule **11**

Overview **13**

MS1 Assistive Machine Learning for People with Disabilities . . . 17
MS2 Causality and Time Series Analysis . . . 20
MS3 Machine Learning for Sustainability . . . 23
MS4 Machine Learning in Computational Biology . . . 25
MS5 Partially Observable Reinforcement Learning . . . 27
Cognitive Dynamic Radio . . . 30
WS1 Adaptive Sensing, Active Learning, and Experimental Design . . . 31
WS2 Advances in Ranking . . . 35
WS3 Analysis and Design of Algorithms for Interactive Machine Learning (ADA-IML'09) . . . 38
WS4 Analyzing Networks and Learning with Graphs . . . 41
WS5 Applications for Topic Models: Text and Beyond . . . 42
WS6 Approximate Learning of Large Scale Graphical Models: Theory and Applications . . . 48
WS7 Nonparametric Bayes . . . 51
WS8 Bounded-rational analyses of human cognition: Bayesian models, approximate inference, and the brain . . . 60
WS9 Clustering: Science or art? Towards principled approaches . . . 62
WS10 Connectivity Inference in Neuroimaging . . . 64
WS11 The Curse of Dimensionality Problem: How Can the Brain Solve It? . . . 67
WS12 Deep Learning for Speech Recognition and Related Applications . . . 69
WS13 Discrete Optimization in Machine Learning: Submodularity, Polyhedra and Sparsity . . . 71
WS14 The Generative and Discriminative Learning Interface . . . 73
WS15 Grammar Induction, Representation of Language and Language Learning . . . 78
WS16 Kernels for Multiple Outputs and Multi-task Learning: Frequentist and Bayesian Points of View . . . 82
WS17 Large-Scale Machine Learning: Parallelism and Massive Datasets . . . 85
WS18 Learning from Multiple Sources with Applications to Robotics . . . 89
WS19 Learning with Orderings . . . 91

WS20 Machine Learning in Computational Biology 95
WS21 Manifolds, sparsity, and structured models: When can low-dimensional geometry really help? 97
WS22 Probabilistic Approaches for Robotics and Control 99
WS23 Normative electrophysiology: explaining cellular properties of neurons from first principles 104
WS24 Optimization for Machine Learning 109
WS25 Statistical Machine Learning for Visual Analytics 113
WS26 Temporal Segmentation: Perspectives from Statistics, Machine Learning, and Signal Processing 116
WS27 Transfer Learning for Structured Data 119
WS28 Understanding Multiple Kernel Learning Methods 121

Index **124**

Maps **131**

Notes **134**

Organizing Committee

General Chairs	Yoshua Bengio, University of Montreal Dale Schuurmans, University of Alberta
Program Chairs	John Lafferty, Carnegie Mellon University Chris Williams, University of Edinburgh
Tutorials Chair	Yann LeCun, New York University
Workshop Chair	Richard Zemel, University of Toronto
Demonstration Chair	Hans Peter Graf, NEC Laboratories America, Inc.
Publications Chair and Electronic Proceedings Chair	Aron Culotta, Southeastern Louisiana University
Volunteers Chair	NIPS Foundation Office
Publicity Chair	NIPS Foundation Office

Program Committee

Jean-Yves Audibert, Ecole des Ponts ParisTech
David Blei, Princeton University
Kwabena Boahen, Stanford University
Michael Bowling, University of Alberta
Nicolo Cesa-Bianchi, University of Milan
Aaron Courville, University of Montreal
Koby Crammer, University of Pennsylvania
Nathaniel Daw, New York University
David Dunson, Duke University
Paolo Frasconi, University of Florence
Nir Friedman, Hebrew University of Jerusalem
Arthur Gretton, Carnegie Mellon University and Max Planck Institute
Matthias Hein, Saarland University
Aapo Hyvarinen, University of Helsinki
Thorsten Joachims, Cornell University
Mark Johnson, Brown University
Charles Kemp, Carnegie Mellon University
John Lafferty, Carnegie Mellon University
Wee Sun Lee, National University of Singapore
Tai Sing Lee, Carnegie Mellon University
Jon McAuliffe, University of Pennsylvania
Yael Niv, Princeton University
Robert Nowak, University of Wisconsin, Madison
Pascal Poupart, University of Waterloo
Carl Rasmussen, University of Cambridge
Erik Sudderth, Brown University
Ben Taskar, University of Pennsylvania
Antonio Torralba, Massachusetts Institute of Technology
Bill Triggs, Laboratoire Jean Kuntzmann, CNRS
Sethu Vijayakumar, University of Edinburgh
Chris Williams, University of Edinburgh
Andrew Zisserman, University of Oxford

NIPS Foundation Officers and Board Members

Sponsors

NIPS gratefully acknowledges the generosity of those individuals and organizations who have provided financial support for the NIPS 2009 conference. The financial support enabled us to sponsor student travel and participation, the outstanding student paper awards, the demonstration track and the opening buffet.

Artificial Intelligence
Google
Microsoft Research
IBM Research
D.E. Shaw & Co.
Yahoo Labs
Alberta Ingenuity Center for Machine Learning
MITACS
Intel
Pascal
Toyota
Springer

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:

Yucheng Low, Workflow Master
Lee Campbell
Chris Hiestand
Sheri Leone
Mary Ellen Perry

Schedule

Thursday, December 10th

13.30-16.30	Mini Symposia	Hyatt Vancouver
14.00-18.00	Buses depart Vancouver Hyatt for Westin Resort and Spa	
17.00-20.30	Registration	Westin Emerald foyer
21.30-22.30	Workshop Keynote Talk	Westin Emerald

Friday, December 11th

6.30–8.00	Breakfast	Westin Emerald
7.00-11.00	Registration	Westin Emerald foyer
7.30–10.30	Workshop sessions	Westin and Hilton
9.00–9.30	Coffee stations	Westin and Hilton
15.30–18.30	Workshop sessions continue	Westin and Hilton
17.00–17.30	Coffee stations	Westin and Hilton

Saturday, December 12th

6.30–8.00	Breakfast	Westin Emerald
7.00-11.00	Registration	Westin Emerald foyer
7.30-11.30	Workshop sessions	Westin and Hilton
9.00–9.30	Coffee stations	Westin and Hilton
15.30–18.30	Workshop sessions continue	Westin and Hilton
17.00–17.30	Coffee stations	Westin and Hilton
19.30 – 22.30	Banquet and wrap up meeting	Westin Emerald

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.

Overview

Thursday, December 10th

Assistive Machine Learning for People with Disabilities

13.30–16.30 .... Hyatt Regency: Regency B **MS1**

Causality and Time Series Analysis

13.30–16.30 .... Hyatt Regency: Regency A **MS2**

Machine Learning for Sustainability

13.30–16.30 .... Hyatt Regency: Regency D **MS3**

Machine Learning in Computational Biology

13.30–16.30 .... Hyatt Regency: Regency C **MS4**

Partially Observable Reinforcement Learning

13.30–16.30 .... Hyatt Regency: Regency EF **MS5**

Opening Remarks

21:30-21:40 .... Westin: Emerald

Workshop Keynote Talk: Simon Haykin

21:40-22:30 .... Westin: Emerald

Friday, December 11th

Adaptive Sensing, Active Learning, and Experimental Design

07:30-10:30 and 15:30-18:30 .... Westin: Emerald A **WS1**

Advances in Ranking

07:30-10:30 and 15:30-18:30 .... Hilton: Diamond Head **WS2**

Analyzing Networks and Learning with Graphs

07:30-10:30 and 15:30-18:30 .... Westin: Nordic **WS4**

Applications for Topic Models: Text and Beyond

07:30-10:30 and 15:30-18:30 .... Westin: Callaghan **WS5**

Clustering: Science or art? Towards principled approaches

07:30-10:30 and 15:30-18:30 .... Hilton: Mt. Currie South **WS9**

The Curse of Dimensionality Problem: How Can the Brain Solve It?

07:30-10:30 and 15:30-18:30 .... Hilton: Black Tusk **WS11**

Discrete Optimization in Machine Learning: Submodularity, Polyhedra and Sparsity

07:30-10:30 and 15:30-18:30 .... Hilton: Cheakamus **WS13**

Grammar Induction, Representation of Language and Language Learning

07:30-10:30 and 15:30-18:30 .... Hilton: Sutcliffe A **WS15**

Large-Scale Machine Learning: Parallelism and Massive Datasets

07:30-10:30 and 15:30-18:30 .... Hilton: Mt. Currie North **WS17**

Machine Learning in Computational Biology

07:30-10:30 and 15:30-18:30 .... Westin: Alpine DE **WS20**

Probabilistic Approaches for Robotics and Control

07:30-10:30 and 15:30-18:30 .... Westin: Alpine BC **WS22**

Normative electrophysiology: explaining cellular properties of neurons from first principles

07:30-10:30 and 16:00-18:30 .... Westin: Alpine A **WS23**

Statistical Machine Learning for Visual Analytics

07:30-10:30 and 15:30-18:45 .... Westin: Glacier **WS25**

Understanding Multiple Kernel Learning Methods

07:30-10:30 and 15:30-18:30 .... Hilton: Sutcliffe B **WS28**

Saturday, December 12th

Analysis and Design of Algorithms for Interactive Machine Learning (ADA-IML'09)

07:30-10:30 and 15:30-18:30 .... Hilton: Black Tusk **WS3**

Approximate Learning of Large Scale Graphical Models: Theory and Applications

07:30-10:30 and 15:30-18:30 .... Hilton: Mt. Currie North **WS6**

Nonparametric Bayes

07:30-10:30 and 15:30-18:30 .... Westin: Emerald A **WS7**

Bounded-rational analyses of human cognition: Bayesian models, approximate inference, and the brain

07:30-10:30 and 15:30-18:30 .... Westin: Alpine BC **WS8**

Connectivity Inference in Neuroimaging

07:30-10:30 and 15:30-18:30 .... Westin: Nordic **WS10**

Deep Learning for Speech Recognition and Related Applications

07:30-10:30 and 15:30-18:30 .... Hilton: Cheakamus **WS12**

The Generative and Discriminative Learning Interface

07:30-10:45 and 15:50-18:45 .... Westin: Alpine DE **WS14**

Kernels for Multiple Outputs and Multi-task Learning: Frequentist and Bayesian Points of View

07:30-10:30 and 15:30-18:30 .... Westin: Alpine A **WS16**

Learning from Multiple Sources with Applications to Robotics

07:30-10:30 and 15:30-18:30 .... Hilton: Sutcliffe A **WS18**

Learning with Orderings

07:30-10:30 and 15:30-18:30 .... Hilton: Diamond Head **WS19**

Manifolds, sparsity, and structured models: When can low-dimensional geometry really help?

07:30-10:30 and 15:30-18:30 .... Hilton: Mt. Currie South **WS21**

Optimization for Machine Learning

07:30-10:30 and 15:30-18:30 .... Hilton: Sutcliffe B **WS24**

Temporal Segmentation: Perspectives from Statistics, Machine Learning, and Signal Processing

07:30-10:30 and 15:30-18:30 .... Westin: Glacier **WS26**

Transfer Learning for Structured Data

07:30-10:30 and 15:30-18:30 .... Westin: Callaghan **WS27**

December 10, 2008, 13.30–16.30 Hyatt Regency: Regency B **MS1**

Assistive Machine Learning for People with Disabilities

http://www.davidroihardoon.com/AMD09/Description.html

Fernando Pérez-Cruz fernando@tsc.uc3m.es
Universidad Carlos III de Madrid
Emilio Parrado-Hernández emipar@tsc.uc3m.es
Universidad Carlos III de Madrid
David R. Hardoon drhardoon@i2r.a-star.edu.sg
Institute for Infocomm Research (I^2R)
Jaisiel Sánchez-Madrid jmadrid@technosite.es
Technosite

Abstract

Nowadays, there are massive amounts of heterogeneous electronic information available on the Web. People with disabilities, among other groups potentially influenced by the digital gap, face great barriers when trying to access information. Sometimes their disability makes their interaction the ICT environment (eg., computers, mobile phones, multimedia players and other hardware devices) more difficult. Furthermore, the contents are delivered in such formats that cannot be accessed by people with disability and the elderly. The challenge for their complete integration in information society has to be analyzed from different technology approaches.

Recent developments in Machine Learning are improving the way people with disabilities access to digital information resources. From the hardware perspective, Machine Learning can be a core part for the correct design of accessible interaction systems of such users with computers (such as BCI). From the contents perspective, Machine Learning can provide tools to adapt contents (for instance changing the modality in which it is accessed) to users with special needs. From the users' perspective, Machine Learning can help constructing a good user modeling, as well as the particular context in which the information is accessed.

1:30- 2:00	**Toward Brain Computer Interfacing: Algorithms for on-line Differentiation of Neuroelectric Activities** K.-R. Müller
2:00-2:30	**Machine learning for brain-computer interfaces** J. Hill
2:30-2:45	**An Efficient P300-based Brain-Computer Interface with Minimal Calibration Time** F. Lotte and C. Guan
2:45-3:00	**Extracting Gait Spatiotemporal Properties from Parkinson's disease Patients** A. Sama, A. Catalá, A. Rodríguez-Molinero and C. Angulo
3:15-3:45	**Machine Learning applied to multi-modal interaction, adaptive interfaces and ubiquitous assistive technologies** J. Madrid
3:45-4:00	**Human-Centered Machine Learning in a Social Interaction Assistant for Individuals with Visual Impairments** V. Balasubramanian, S. Chakraborty, S. Krishna and S. Panchanathan
4:00-4:15	**Toward Text-to-Picture Synthesis** A. B. Goldberg, J. Rosin, X. Zhu and C. R. Dyer

4:15-4:30	**Fast and Flexible Selection with a Single Switch** T. Broderick and D.J.C. MacKay
4:30-4:45	**Data Mining based User Modeling Systems for Web Personalization applied to people with disabilities** J. Abascal, O. Arbelaitz, J. Muguerza and I. Perona
4:45-5:00	**Perspective on the Goals and Complexities of Inclusive Design** A. Z. Perkins

Toward Brain Computer Interfacing: Algorithms for on-line Differentiation of Neuroelectric Activities
Klaus-Robert Muller, T.U. Berlin
Brain Computer Interfacing (BCI) aims at making use of brain signals for e.g. the control of objects, spelling, gaming and so on. This talk will first provide a brief overview of Brain Computer Interface from a machine learning and signal processing perspective. In particular it shows the wealth, the complexity and the difficulties of the data available, a truely enormous challenge: In real-time a multi-variate very strongly noise contaminated data stream is to be processed and neuroelectric activities are to be accurately differentiated in real time. Finally, I report in more detail about the Berlin Brain Computer (BBCI) Interface that is based on EEG signals and take the audience all the way from the measured signal, the preprocessing and filtering, the classification to the respective application. BCI as a new channel for man-machine communication is discussed in a clincial setting and for gaming. This is joint work with Benjamin Blankertz, Michael Tangermann, Claudia Sanelli, Carmen Vidaurre, Thorsten Dickhaus (TU Berlin), Steven Lemm, Guido Nolte, Andreas Ziehe, Florin Popescu (Fraunhofer FIRST, Berlin) Gabriel Curio, Vadim Nikulin (Charite, Berlin) and further member of the Berlin Brain Computer Interface team, see www.bbci.de.

Machine learning for brain-computer interfaces
Jeremy Hill, Max Planck Institute
Brain-computer interfaces (BCI) aim to be the ultimate in assistive technology: decoding a user's intentions directly from brain signals without involving any muscles or peripheral nerves. Thus, some classes of BCI potentially offer hope for users with even the most extreme cases of paralysis, such as in late-stage Amyotrophic Lateral Sclerosis, where nothing else currently allows communication of any kind. Other lines in BCI research aim to restore lost motor function in as natural a way as possible, reconnecting and in some cases re-training motor-cortical areas to control prosthetic, or previously paretic, limbs. Research and development are progressing on both invasive and non-invasive fronts, although BCI has yet to make a breakthrough to widespread clinical application.

The high-noise high-dimensional nature of brain-signals, particularly in non-invasive approaches and in patient populations, make robust decoding techniques a necessity. Generally, the approach has been to use relatively simple feature extraction techniques, such as template matching and band-power estimation, coupled to simple linear classifiers. This has led to a prevailing view among applied BCI researchers that (sophisticated) machine-learning is irrelevant since "it doesn't matter what classifier you use once you've done your preprocessing right and extracted the right features." I shall show a few examples of how this runs counter to both the empirical reality and the spirit of what needs to be done to bring BCI into clinical application. Along the way I'll highlight some of the interesting problems that remain open for machine-learners.

Machine Learning applied to multi-modal interaction, adaptive interfaces and ubiquitous assistive technologies
Jaisiel Madrid, Technosite
The presentation will describe the challenge of eInclusion in the technological design process, which impedes the complete integration of people with disabilities and elderly in Information Society. To face this challenge, the INREDIS project aims to face individual needs of users instead of addressing the needs of the average user, by proposing basic technologies that enables the creation of personalized channels for communication and interaction with the technological environment. For this purpose, Machine Learning

can help constructing effective methods to reflect user needs, preferences and expectations (and their evolution over time) on user interfaces, consequently improving satisfaction and performance. In particular, academia and industry within the INREDIS consortium explore together the potential of Machine Learning on multimodal services and ubiquitous assistive technologies, as well as adaptive user interfaces according to user and technological capabilities.

December 10, 2008, 13.30–16.30 Hyatt Regency: Regency A **MS2**

Causality and Time Series Analysis

http://clopinet.com/isabelle/Projects/NIPS2009

Florin Popescu florin.popescu@first.fraunhofer.de
Fraunhofer FIRST
Guido Nolte guido.nolte@first.fraunhofer.de
Fraunhofer FIRST
Isabelle Guyon guyon@clopinet.com
Clopinet

Abstract

This symposium addresses a topic that has spurred vigorous scientific debate of late in the fields of neuroscience and machine learning: causality in time-series data. In neuroscience, causal inference in brain signal activity (EEG, MEG, fMRI, etc.) is challenged by relatively rough prior knowledge of brain connectivity and by sensor limitations (mixing of sources). On the machine learning side, as the Causality workshop last year's NIPS conference has evidenced for static (non-time series) data, there are issues of whether or not graphical models (directed acyclic graphs) pioneered by Judea Pearl, Peter Spirtes, and others can reliably provide a cornerstone of causal inference, whereas in neuroscience there are issues of whether Granger type causality inference is appropriate given the source mixing problem, traditionally addressed by ICA methods. Further topics, yet to be fully explored, are non-linearity, non-Gaussianity and full causal graph inference in high-dimensional time series data. Many ideas in causality research have been developed by and are of direct interest and relevance to researchers from fields beyond ML and neuroscience: economics (i.e. the Nobel Prize winning work of the late Clive Granger, which we will pay tribute to), process and controls engineering, sociology, etc. Despite the long-standing challenges of time-series causality, both theoretical and computational, the recent emergence of cornerstone developments and efficient computational learning methods all point to the likely growth of activity in this seminal topic.

Along with the stimulating discussion of recent research on time-series causality, we will present and highlight time-series datasets added to the Causality Workbench, which have grown out of last year's Causality challenge and NIPS workshop, some of which are neuroscience related.

13:30	Welcome and Introduction
13:35	**Granger causality and dynamic structural systems** Halbert White
14:15	**Time series causality inference using the Phase Slope Index** Guido Nolte
14:40	Coffee break
15:00	**Granger causality in brain connectivity studies using functional Magnetic Resonance Imaging (fMRI) data** Rainer Goebel
15:25	**Graphical Causal Models for Time Series Econometrics: Some Recent Developments and Applications** Alessio Moneta
15:50	**Open-access datasets for time series causality discovery validation** Isabelle Guyon

Granger causality and dynamic structural systems
Halbert White, Department of Economics, University of California San Diego
Xun Lu, Department of Economics, University of California San Diego
Using a generally applicable dynamic structural system of equations, we give natural definitions of direct and total structural causality applicable to both structural VARs and recursive structures representing time-series natural experiments. These concepts enable us to forge a previously missing link between Granger (G-) causality and structural causality by showing that, given a corresponding conditional form of exogeneity, G- causality holds if and only if a corresponding form of structural causality holds. Of importance for applications is the structural characterization of finite-order G-causality, which forms the basisfor most empirical work. We show that conditional exogeneity is necessary for valid structural inference and prove that in the absence of structural causality, conditional exogeneity is equivalent to G non-causality. We provide practical new G-causality and conditional exogeneity tests and describe their use in testing for structural causality.

Time series causality inference using the Phase Slope Index
Florin Popescu, Fraunhofer Institute FIRST, Berlin
Guido Nolte, Fraunhofer Institute FIRST, Berlin
A method recently introduced by Nolte et. al (Phys Rev Lett 100:23401, 2008) estimates the causal direction of interactions robustly with respect to instantaneous mixtures of independent sources with arbitrary spectral content, i.e. in observations which are dominated by non-white spatially correlated noise and in which dynamic structural interaction plays little part. The method, named Phase Slope Index (PSI), is unlikely to assign causality in the case of lack of dynamic interaction among time series, unlike Granger causality for linear systems. Results show that PSI does not yield false positives even in the case of nonlinear interactions. The meaning of instaneous noise mixtures in different data domains will be discussed in the context of correct correlation vs. causation inference, and the theoretical relationship of PSI to other time-series causality inference methods will be expanded upon.

Granger causality in brain connectivity studies using functional Magnetic Resonance Imaging (fMRI) data
Alard Roebroeck, Faculty of Psychology & Neuroscience, Maastricht University
Rainer Goebel, Faculty of Psychology & Neuroscience, Maastricht University
This talk will discuss the application of Granger causality to fMRI data in the form of Granger causality mapping (GCM), which is used to explore directed influences between neuronal populations (effective connectivity) in fMRI data. The method does not rely on a priori specification of a model that contains pre-selected regions and connections between them. This distinguishes it from other fMRI effective connectivity approaches that aim at testing or contrasting specific hypotheses about neuronal interactions. Instead, GCM relies on the Granger causality concept to define the existence and direction of influence from temporal information in the data. The problems of limited temporal resolution in fMRI, and the hemodynamic source of the signal that makes direct interpretation of fMRI Granger causality as neuronal influence difficult, will be discussed.

Graphical causal models for time series econometrics: some recent developments and applications
Alessio Moneta, Max Planck Institute for Economics, Jena
In this talk, I shall assess the empirical plausibility of the real business cycle view that shocks to real variables are the dominant sources of economic fluctuations and that monetary policy shocks play an insignificant role in determining the behavior of real variables. I reconsider the vector autoregressive model of King et al. (Am Econ Rev 81:819?840, 1991), but propose an alternative identification method, based on graphical causal models. This method selects the contemporaneous causal structure using the information incorporated in the partial correlations among the residuals. The residuals orthogonalization which follows and the study of the impulse response functions confirm the results of King et al.: permanent productivity shocks are not the dominant sources of aggregate fluctuations in US economy.

Open-access datasets for time series causality discovery validation
Isabelle Guyon, Clopinet, Berkeley
The CausalityWorkbench project provides an environment to test causal discovery algorithms. Via a web

portal (http://clopinet.com/causality), we provide a number of resources, including a repository of datasets, models, and software packages, and a virtual laboratory allowing users to benchmark causal discovery algorithms by performing virtual experiments to study artificial causal systems. We regularly organize competitions. Our repository already includes several time dependent datasets from a variety of domains: system biology, neurosciences, physics, manufacturing, and marketing. We will invite new contributions and present our plan for upcoming evaluations of causal models for time series applications.

December 10, 2008, 13.30–16.30 Hyatt Regency: Regency D **MS3**

Machine Learning for Sustainability

http://www.cs.stanford.edu/group/nips09-mlsust

J. Zico Kolter kolter@cs.stanford.edu
Stanford University
Thomas G. Dietterich tgd@cs.orst.ed
Oregon State University
Andrew Y. Ng ang@cs.stanford.edu
Stanford University

Abstract

The world has a sustainability problem. Humans currently consume an average of 16TW of power (and rising), more than 86% of which comes from (unsustainable) fossil fuels. There is a range of estimates as to when this supply will run out, but this is a scenario that may well happen within our lifetimes. Even more pressing is the effect that such fuels have on our climate: given no attempts to reduce the world's fossil fuel usage, even the most conservative climate models predict that the world temperature will increase by over five degrees (Fahrenheit) in the next 90 years, an increase that could cause ecological disasters on a global scale. Building a sustainable infrastructure for energy and ecosystems is shaping up to be one of the grand scientific and political challenges of the 21st century. Furthermore, there is a growing consensus that many aspects of sustainability are fundamentally information systems problems, tasks where machine learning can play a significant role.

This mini-symposium will bring together leading researchers with both machine learning backgrounds and energy/sustainability backgrounds to address the question: How can machine learning help address the world's sustainability problem? The mini-symposium will also seek to answer: What is the current state of work directed at sustainability, energy, and ecology in the machine learning, operations research, and optimization communities? What are the primary scientific and technical challenges in information processing for sustainability? And finally, what are (and what aren't) areas where machine learning can make a genuine impact on the science of sustainability?

Because this is an emerging field of research, the talks at this symposium will aimed at the general NIPS audience. There is a growing number of researchers working in sustainability, but even more broadly, we think that such problems have the potential to advance basic machine learning in a manner similar to other important applications, such as computer vision, natural language processing, and computational biology. Sustainability problems offer an equally rich set of domains, and solutions to these problems will have a genuine impact on the world.

1:30–1:40	Opening Remarks
1:40–2:05	**Machine learning for the NYC power grid: lessons learned and the future** David L. Waltz
2:05–2:30	**What it takes to win the carbon war. Why even AI is needed.** Saul Griffith
2:30–3:00	Coffee Break
3:00-3:25	**Ecological Science and Policy: Challenges for Machine Learning** Thomas G. Dietterich
3:25-3:50	**Optimizing Information Gathering in Environmental Monitoring** Carlos Guestrin

3:50-4:15	**Approximate Dynamic Programming in Energy Resource Management** Warren Powell
4:15–4:25	Closing Remarks

December 10, 2008, 13.30–16.30 Hyatt Regency: Regency C **MS4**

Machine Learning in Computational Biology

http://www.fml.tuebingen.mpg.de/nipscompbio/mlcb-2009/mini-symposium-program

Jean-Phillipe Vert Jean-Philippe.Vert@ensmp.fr
ParisTech
Yanjun Qi yanjun@nec-labs.com
NEC Labs America
Gal Chechik gal.chechik@gmail.com
Google
Alexander Zien Alexander.Zien@tuebingen.mpg.de
LIFE Biosystems GmbH
Tomer Hertz thertz@fhcrc.org
Fred Hutchinson Cancer Research Center
William Noble noble@gs.washington.edu
University of Washington

Abstract

The field of computational biology has seen dramatic growth over the past fewyears, both in terms of new available data, new scientific questions, and newchallenges for learning and inference. In particular, biological data are oftenrelationally structured and highly diverse, well-suited to approaches thatcombine multiple weak evidence from heterogeneous sources. These data mayinclude sequenced genomes of a variety of organisms, gene expression data frommultiple technologies, protein expression data, protein sequence and 3Dstructural data, protein interactions, gene ontology and pathway databases,genetic variation data (such as SNPs), and an enormous amount of textual datain the biological and medical literature. New types of scientific and clinicalproblems require the development of novel supervised and unsupervised learningmethods that can use these growing resources. Furthermore, next generationsequencing technologies are yielding terabyte scale data sets that requirenovel algorithmic solutions. The goal of this min-symposium is to presentemerging problems and machine learning techniques in computational biology.

13:30-14:25	**Understanding Gene Regulatory Networks and Their Variations** Daphne Koller
14:25-15:20	**Using Networks to Elucidate Disease and Drugs** Roded Sharan
15:20-16:15	**Novel Applications of Computational Biology in Infectious Disease Interventions** Elizabeth Halloran

Understanding Gene Regulatory Networks and Their Variations
Daphne Koller, Stanford University
A key biological question is to uncover the regulatory networks in a cellular system and to understand how this network varies across individuals, cell types, and environmental conditions. In this talk I will describe work that uses machine learning techniques to reconstruct gene regulatory networks from gene expression data. Specifically, we exploit novel forms of Bayesian regularized regression to enable transfer between multiple related learning problems, such as between different individuals or between different cell types. We demonstrate applications in two domains: understanding the effect of individual genetic variation on

gene regulation and its effect on phenotypes including human disease; and understanding the regulatory mechanisms underling immune system cell differentiation.

Using Networks to Elucidate Disease and Drugs
Roded Sharan, Tel-Aviv University
In recent years, there is a tremendous growth in large scale networks describing diseases and drugs. These allow for the first time a systems-level analysis of the molecular basis of disease and the therapeutic properties of drugs. In my talk I will describe several recent works in this direction, aiming to associate genes, protein complexes and pathways with disease and to uncover therapeutic areas, targets and side effects of drugs.

Novel Applications of Computational Biology in Infectious Disease Interventions
Elizabeth Halloran, Fred Hutchinson Cancer Research Center
Interventions in infectious diseases are increasingly relying on computational biology and genomic methods. Estimating changes in viral genetic diversity in a population could be a new potential method to evaluate vaccination strategies in populations. Transgenic mosquitoes immune to a pathogen are being developed to replace the native mosquito vector of a number of vector-borne diseases. High throughput methods are being used to elucidate mechanisms of immune memory with the promise of developing better vaccines. Large-scale computer simulation models are useful for exploring interventions and could benefit from input from network and graph theory. In this talk, we discuss a few novel applications of computational biology in understanding infectious diseases and interventions.

December 10, 2008, 13.30–16.30 Hyatt Regency: Regency EF **MS5**

Partially Observable Reinforcement Learning

http://www.hutter1.net/ai/porlsymp.htm

Marcus Hutter marcus.hutter@anu.edu.au
Australian National University
William Uther William.Uther@nicta.com.au
NICTA and University of New South Wales
Pascal Poupart ppoupart@cs.uwaterloo.ca
University of Waterloo
Kee Siong Ng KeeSiong.Ng@nicta.com.au
NICTA and The Australian National University

Abstract

For many years, the reinforcement learning community primarily focused on sequential decision making in fully observable but unknown domains while the planning under uncertainty community focused on known but partially observable domains. Since most problems are both partially observable and (at least partially) unknown, recent years have seen a surge of interest in combining the related, but often different, algorithmic machineries developed in the two communities. The time thus seems ripe for a symposium that brings these two communities together and reviews recent advances in this convergence.

A reinforcement learning agent for a partially observable environment is often broken into two parts: 1) the inference of an environment model from data; and 2) the solution of the associated control/planning problem. There has been significant progress on both these fronts in recent years. Both linear and non-linear models of various forms can now be learned from history data. Modern POMDP solvers can also now handle some models with millions of states. This symposium brings together five active researchers in PORL research to present some state-of-the-art developments.

13:30 - 13:45	**An Introduction and Overview of Approaches to PORL** William Uther
13:50 - 14:20	**What to Model?** Satinder Singh
14:25 - 14:55	**Structured Hierarchical Bayesian Priors For Modeling Dynamical Systems** David Wingate
15:00-15:10	Coffee break
15:10 - 15:40	**Algorithmic and Theoretical Properties of Bayesian Reinforcement Learning** Pascal Poupart
15:45 - 16.15	**Principled Large-Scale POMDP Learning** Marcus Hutter
16:15 - 16:30	**A Monte Carlo AIXI Approximation** Joel Veness

An Introduction and Overview of Approaches to PORL
William Uther, NICTA and University of New South Wales

Inferring a model of a partially observable environment from a sequence of actions, observations and rewards has been approached in a number of different ways. This introduction for the mini-symposium will give a brief overview of the range of techniques found in the literature and describe some relationships between them.

What to Model?
Satinder Singh, University of Michigan
In applications of reinforcement learning to engineering and control problems, it is often quite clear what models to build, i.e., the state variables and actions are well defined and the modeling challenge is to find a good parametric representation and efficient learning algorithm. In more AI settings, where the specific tasks have not themselves been the object of many person-years of engineering effort, e.g., a robot equipped with a camera and other sensors in a building expected to do all sorts of chores, it is far less clear what model to build. What is important to predict? What secondary things do we need to be able to predict in order to make predictions of the primary things we want to predict? Ideally, one would want to learn models whose complexity depends only on the complexity of the prediction questions we want the model to be able to answer and not on the complexity of the real world. These are significant and understudied challenges towards progress in AI. In this talk I will present some thoughts and preliminary results pertaining to these challenges.

Structured Hierarchical Bayesian Priors For Modeling Dynamical Systems
David Wingate, Massachusetts Institute of Technology
Josh Tenenbaum, Massachusetts Institute of Technology
Noah Goodman, Massachusetts Institute of Technology
Hierarchical Bayesian models with interesting (possibly nonparametric) priors can be used to learn models of structured but partially observable domains. Elemental distributions (such as multinomials or Dirichlet processes) can be composed, for example, to co-cluster states and perceptual features with an unknown number of clusters; such clustering improves generalization and decreases sample complexity of learning models. The same types of priors can also be used to reason about complex, compositional options which are triggered in response to features of the environment. In this talk, we'll discuss examples of both structured world models (including object-oriented models) and structured policies, and how an agent can plan using both.

Algorithmic and Theoretical Properties of Bayesian Reinforcement Learning
Pascal Poupart, University of Waterloo
Reinforcement Learning in partially observable domains is a notoriously hard problem. When taking a Bayesian perspective, reinforcement learning becomes a problem of planning in a special partially observable Markov decision process (POMDP) where the unknown parameters (e.g., transition dynamics, observation probabilities and reward distribution) are treated as state features that the learner maintains beliefs about. In this talk, I will describe some of the algorithmic and theoretical properties of this POMDP. More specifically, I will discuss how this POMDP provides a natural formulation of non-myopic active learning and how the exploration/exploitation tradeoff is naturally optimized. I will also show that the optimal value function is the upper surface of a set of mixtures of Dirichlets which is piecewise-linear and convex. These properties will then be used to derive approximate dynamic programming procedures such as point-based value iteration for offline planning.

Principled Large-Scale POMDP Learning
Marcus Hutter, The Australian National University
Partially Observable Markov Decision Processes (POMDPs) are a very important generalization of MDPs. Nature is still assumed to be an MDP, but the states of nature are only partially observed via some non-injective or probabilistic function. While POMDP planning is well-defined but (harder than NP) hard, a general theory of learning POMDPs is missing. A different approach is to work with histories and directly map them to a finite approximation of a (belief) MDP. A Coding/MDL/Bayes-inspired criterion can be used to learn this mapping. This reduction significantly expands the scope of many existing reinforcement learning algorithms. The approach can be extended to factored MDPs, resulting in the first principled general-purpose learning algorithm for large POMDPs.

A Monte Carlo AIXI Approximation

Joel Veness, University of New South Wales

Marcus Hutter's AIXI agent provides a mathematically rigorous, optimality notion for general reinforcement learning agents. An interesting open question is to what extent this ideal can be approximated in a computationally efficient manner. This talk will focus on a particular, direct approximation of AIXI. This approximation can be broken into two main parts: Solomonoff's universal distribution is approximated by an efficiently computable Bayesian mixture of Prediction Suffix Trees, and the finite horizon expectimax operation is approximated via Monte-Carlo Tree Search. Although this approximation is undoubtedly crude, it has already achieved impressive results on small, noisy, partially observable and stochastic domains.

Cognitive Dynamic Radio

Simon Haykin

McMaster University

haykin@mcmaster.ca

Abstract

In this Keynote Talk, I will focus on a new generation of dynamic systems that are enabled with cognition. Specifically, I will highlight three applications:

- Cognitive Radio;
- Cognitive Mobile Assistants; and
- Cognitive Radar

the study of which constitutes the main thrust of my current research program.

This introductory material will pave the way for me to link up with the visual brain, which, is characterized by the perception-action cycle.

For the last part of my talk I will expand on this cycle, focusing on cognitive radar, in the context of which I will do the following:

- describe a new generation of nonlinear filters that we have named "cubature Kalman filters" that provide the best known approximation to the Bayesian filter in an information-theoretic sense when operating in a Gaussian environment; and
- present experimental results (using computer simulations) on cognitive tracking radar that will demonstrate the power of cognition.

December 11, 2009, 07:30-10:30 and 15:30-18:30 Westin: Emerald A **WS1**

Adaptive Sensing, Active Learning, and Experimental Design

http://www.isr.ist.utl.pt/~rmcantin/nips2009.php

Rui Castro rmcastro@ee.columbia.edu
Columbia University
Nando de Freitas nando@cs.ubc.ca
University of British Columbia
Ruben Martinez-Cantin rmcantin@gmail.com
Instituto Superior Tecnico

Abstract

The fields of active learning, adaptive sensing and sequential experimental design have seen a growing interest over the last decades in a number of communities, ranging from machine learning and statistics to biology and computer vision. Broadly speaking, all active and adaptive approaches focus on closing the loop between data analysis and acquisition. Said in a different way the goal is to use information collected in past samples to adjust and improve the future sampling and learning processes, in the spirit of the twenty questions game. These fields typically address the problem in very diverse ways, and using different problem formulations. The main objective of this workshop is to bring these communities together, share ideas and knowledge, and cross-fertilize the various fields. Most of the theoretical work in the area of adaptive sensing and active learning has remained quite distant from the realm of practical applications (with a few notable exceptions). In less-than-ideal settings, many modeling assumptions are only approximately true, and hence closed-loop (active) methods as described need to be very robust in other to: (i) guarantee consistency, in the sense that the proposed method must not fail dramatically; (ii) improve on the performance of open-loop (passive) procedures whenever favorable conditions are met. Due to the feedback nature of closed-loop procedures these are often prone to failure when modeling assumptions are only approximately met, and this has been observed by many when deploying practical algorithms. By bringing together both theoreticians and practitioners from the fields of computer vision and robotics, statistics, signal and information processing and machine learning it will be possible to identify promising directions for active learning at large, and address these points in a satisfactory way.

7:30-7:45	Opening remarks
7:45-8:15	**TBA** Andreas Krause
8:15-8:45	**Gaussian Process Response Surface Optimization** Dan Lizotte
8:45-9:00	Discussion
9:00-9:30	Coffee Break
9:30-10:00	Posters
10:00-10:30	**Active Learning in Robotics** Luis Montesano
15:45-16:15	**Large Scale Nonlinear Bayesian Experimental Design: Adaptive Compressive Sensing in the Real World** Matthias Seeger
16:15-16:45	**Playing 20 Questions with the Homunculus: Optimal Experimental**

	Design for Neurophysiology Jeremy Lewi
16:45-17:00	Discussion
17:00-17:30	Coffee Break
17:30-18:00	**The True Sample Complexity of Active Learning** Maria-Florina Balcan
18:00-18:30	Panel discussion

INVITED SPEAKERS

The True Sample Complexity of Active Learning
Maria-Florina Balcan, Georgia Tech
We describe a new perspective on the sample complexity of active learning. In many situations where it was generally believed that active learning does not help, we show that active learning does help in the limit, often with exponential improvements in sample complexity. These new insights arise from a subtle variation on the traditional definition of sample complexity, not previously recognized in the active learning literature.

TBA
Andreas Krause, Caltech

Gaussian Process Response Surface Optimization
Daniel Lizotte, University of Michigan
Response surface methods construct a Gaussian process model of an objective function based on all observed data points. The model is then used to compute which point the method should acquire next in its search for the global optimum of the objective. These optimization methods can be very efficient in terms of the number of objective function evaluations used, but existing formulations have drawbacks: Although they are intended to be "black-box," these methods are sensitive to the initial choice of function evaluations, they are not invariant to shifting and scaling of the objective function, and their experimental evaluation to date has been limited. We examine each of these issues and present rules of thumb for deploying response surface methods in practice. Along the way, we will discuss the idea of quantifying the difficulty of global optimization problems that are drawn from Gaussian process models.

Active learning in robotics
Luis Montesano, University of Zaragoza
Robot systems have to be able to adapt their behavior and acquire new abilities. To achieve this, learning has become a crucial component of many successful robotic systems. However, data is gathered by interaction with the environment or with humans, requiring time and energy. In this talk, we will discuss some examples where active strategies reduce the amount of information/samples required to learn new models, skills or tasks. We will cover different robotic problems ranging from discovering the robot structure, learning new skills to interact with objects and imitation learning through active inverse reinforcement learning.

Playing 20 questions with the homunculus: optimal experimental design for neurophysiology
Jeremy Lewi, Georgia Tech
Liam Paninski, Columbia University
Neurophysiology is in many ways analogous to the game of twenty questions. One of the fundamental paradigms in experimental neuroscience is to stimulate the brain, measure the response, and then infer what the brain is doing. As in the game twenty questions, success depends critically on intelligently picking the next stimulus or question based on the data already gathered. We frame this problem in the context of neurophysiology by modeling a neuron as a generalized linear model. We present methods for constructing the stimulus which will provide the most information for deciding which GLM provides the best model of the neuron. We show that for purely spatial stimuli, we can reduce the problem to a tractable 2-d optimization which can be solved in near real time. We also consider the case of constructing near optimal stimuli when the

stimulus has complex spatio-temporal structure such as a sound. We validate our methods using simulations in which the data was obtained from auditory experiments with Zebra Finch.

Large Scale Nonlinear Bayesian Experimental Design: Adaptive Compressive Sensing in the Real World

Matthias Seeger, Saarland University, Saarbruecken

How to best acquire a real world image for nonlinear sparse reconstruction? While out of scope of current compressive sensing theory, this problem can be addressed by nonlinear sequential Bayesian experimental design, if approximate Bayesian inference is scaled up to high-resolution images by way of novel variational relaxations. We provide results of a study aiming to speed up magnetic resonance imaging by optimized undersampling, one of the most important potential applications of compressive sensing yet. In nonlinear experimental design, decisions depend on previously obtained responses, linking it to the more familiar problem of active learning. We will outline the basic properties of the former, to facilitate theory transfer with the latter. In acquisition optimization, the goal is a high-dimensional spatial signal rather than a binary label, the driving statistic is the posterior covariance matrix. Meaningful analysis must not be based on common assumptions of unstructured exact sparsity, but on weaker heavy-tails assumptions, and has to focus on approximate rather than intractable exact Bayesian inference. Recent convex variational approximations based on standard computational primitives may be promising targets towards such analyses of real practical relevance.

ACCEPTED CONTRIBUTIONS

Gaussian Processes for Global Optimization

Roman Garnett,
Michael A. Osborne,
Stephen J. Roberts,

Active Data Selection with Faults and Changepoints

Michael A. Osborne,
Roman Garnett,
Stephen J. Roberts,

Decision-theoretic Planning under Uncertainty for Cooperative Active Perception

Matthijs T.J. Spaan,
Pedro U. Lima,

Active Filtering for robotic tactile learning

Hannes P. Saal,
Jo-Anne Ting,
Sethu Vijayakumar,

Active Learning with an ERM Oracle

Alina Beygelzimer,
Daniel Hsu,
John Langford,
Tong Zhang,

Dynamic sensing policies for monitoring arterial road traffic

Aude Hofleitner,
Saurabh Amin,
Ryan Herring,
Pieter Abbeel,

Alexandre Bayen,

Robust Selective Sampling from Single and Multiple Teachers
Ofer Dekel,
Claudio Gentile,
Karthik Sridharan,

Gaussian Porcess Bandits: An Experimental Design Apprach
Niranjan Srinivas,
Andreas Krause,
Sham M Kakade,
Matthias Seeger,

Efficient Resource-constrained Retrospective Analysis of Long Video Sequences
Daozheng Chen,
Mustafa Bilgic,
Lise Getoor,
David Jacobs,

Activized Learning: Transforming Passive to Active with Improved Label Complexity
Steve Hanneke,

December 11, 2009, 07:30-10:30 and 15:30-18:30 Hilton: Diamond Head **WS2**

Advances in Ranking

http://web.mit.edu/shivani/www/Ranking-NIPS-09

Shivani Agarwal shivani@mit.edu
MIT
Chris Burges chris.burges@microsoft.com
Microsoft Research
Koby Crammer koby@ee.technion.ac.il
Technion

Abstract

Ranking problems are increasingly recognized as a new class of statistical learning problems that are distinct from the classical learning problems of classification and regression. Such problems arise in a wide variety of domains: in information retrieval, one wants to rank documents according to relevance to a query; in natural language processing, one wants to rank alternative parses or translations of a sentence; in collaborative filtering, one wants to rank items according to a user's likes and dislikes; in computational biology, one wants to rank genes according to relevance to a disease. Consequently, there has been much interest in ranking in recent years, with a variety of methods being developed and a whole host of new applications being discovered.

This workshop aims to bring together researchers interested in the area to share their perspectives, identify persisting challenges as well as opportunities for meaningful dialogue and collaboration, and to discuss possible directions for further advances and applications in the future.

07:30-07:35	Opening Remarks
07:35-08:20	**KEYNOTE LECTURE: Classical Methods for Ranked Data: A Review** Persi Diaconis (Stanford University)
08:20-08:40	**A Decision-Theoretic Model of Rank Aggregation** Tyler Lu and Craig Boutilier (University of Toronto)
08:40-09:00	**Inferring Rankings Under Constrained Sensing** Srikanth Jagabathula and Devavrat Shah (MIT)
09:00-09:15	Coffee Break
09:15-09:50	**INVITED TALK: Statistical Ranking and Combinatorial Hodge Theory** Lek-Heng Lim (UC Berkeley)
09:50-10:10	**Learning Preferences from Atomic Gradients** Michael Wick, Khashayar Rohanimanesh, Aron Culotta and Andrew McCallum (University of Massachusetts Amherst)
10:10-10:30	**Half Transductive Ranking** Bing Bai (NEC Labs), Jason Weston (Google Research), David Grangier (NEC Labs), Ronan Collobert (NEC Labs), Corinna Cortes (Google Research) and Mehryar Mohri (Google Research and New York University)
	[Posters available for preview and discussion]

15:30-16:05	**INVITED TALK: Matchbox: General Purpose, Large Scale, Online Bayesian Recommendations** Ralf Herbrich (Microsoft Research Cambridge)
16:05-16:25	**Scalable Online Learning to Rank from Clicks** Filip Radlinski, Aleksandrs Slivkins and Sreenivas Gollapudi (Microsoft Research Cambridge)
16:25-17:20	**Poster Session and Coffee Break**
17:20-17:45	**Learning to Rank Through the Wisdom of Crowds** Jennifer Wortman Vaughan (Harvard University)
17:45-18:25	**Panel Discussion**
18:25-18:30	Closing Remarks

Classical Methods for Ranked Data: A Review

Persi Diaconis, Stanford University

This review talk covers spectral and metric techniques for working with ranked and partially ranked data. These methods were developed for large data sets involving few ranked items. Nowadays, cases with many ranked items are also of interest.

Statistical Ranking and Combinatorial Hodge Theory

Lek-Heng Lim, University of California, Berkeley

We discuss a number of techniques for obtaining a global ranking from data that may be incomplete and imbalanced – characteristics almost universal to modern data sets coming from e-commerce and internet applications. We are primarily interested in score or rating-based cardinal data. From raw ranking data, we construct pairwise rankings, represented as edge flows on an appropriate graph. Our statistical ranking method uses the graph Helmholtzian, the graph theoretic analogue of the Helmholtz operator or vector Laplacian, in much the same way the graph Laplacian is an analogue of the Laplace operator or scalar Laplacian.

We study the graph Helmholtzian using combinatorial Hodge theory: we show that every edge flow representing pairwise ranking can be resolved into two orthogonal components, a gradient flow (acyclic) that represents the L2-optimal global ranking and a divergence-free flow (cyclic) that measures the validity of the global ranking obtained – if this is large, then the data does not have a meaningful global ranking. This divergence-free flow can be further decomposed orthogonally into a curl flow (locally cyclic) and a harmonic flow (locally acyclic but globally cyclic); these provides information on whether inconsistency arises locally or globally. An obvious advantage over the NP-hard Kemeny optimization is that discrete Hodge decomposition may be computed via standard linear least squares regression. We discuss relations with Kemeny optimization, Borda count, and Kendall-Smith consistency index from social choice theory and statistics.

Time permitting, we will also discuss (1) L1-projection of edge flows and correlation maximization over bounded divergence-free flows, (2) L1-approximate sparse cyclic ranking and correlation maximization over bounded curl-free flows, (3) a new twist where we used nuclear norm minimization over skew symmetric matrices to determine a globally consistent ranking.

The main part of this work is joint with Yuan Yao. This talk also features collaboration with David Gleich, Xiaoye Jiang, and Yinye Ye.

Matchbox: General Purpose, Large Scale, Online Bayesian Recommendations
Ralf Herbrich, Microsoft Research Cambridge
I will present a scalable probabilistic model for generating personalised recommendations of items to users of a web service. The system makes use of content information in the form of user and item meta data in combination with collaborative filtering information from previous user behaviour in order to predict the value of an item for a user. Users and items are represented by feature vectors which are mapped into a low-dimensional 'trait space' in which affinity is measured by inner products.

Efficient inference is achieved by approximate message passing involving a combination of Expectation Propagation (EP) and Variational Message Passing. The model was designed from the ground up to be general purpose and practical. It can be combined with different types of model for feedback in order to learn about user-item preferences and a dynamics model can be added to allow for changing item popularity and changing user tastes. The model can also be trained online meaning that new data can be taken account of immediately so recommendations are always up to date.

In the second part of the talk I will discuss some diverse applications of Matchbox. The system is currently being tested in a number of domains:

- content on the MSN portal,
- automatic algorithm selection for Microsoft Solver Foundation, and
- game and multimedia recommendations for Xbox Marketplace.

This is joint work with David Stern and Thore Graepel.

December 12, 2009, 07:30-10:30 and 15:30-18:30 Hilton: Black Tusk **WS3**

Analysis and Design of Algorithms for Interactive Machine Learning (ADA-IML'09)

http://research.microsoft.com/~sumitb/adaiml09

Sumit Basu sumitb@microsoft.com
Microsoft Research
Ashish Kapoor akapoor@microsoft.com
Microsoft Research

Abstract

The traditional role of the human operator in machine learning problems is that of a batch labeler, whose work is done before the learning even begins. However, there is an important class of problems in which the human is interacting directly with the learning algorithm as it learns. Canonical problem scenarios which fall into this space include active learning, interactive clustering, query by selection, learning to rank, and others. Such problems are characterized by three main factors:

1. the algorithm requires input from the human during training, in the form of labels, feedback, parameter guidance, etc.

2. the user cannot express an explicit loss function to optimize, either because it is impractical to label a large training set or because they can only express implicit preferences.

3. the stopping criterion is performance that's "good enough" in the eyes of the user.

The goal of this workshop is to focus on the machine learning techniques that apply to these problems, both in terms of surveying the major paradigms and sharing information about new work in this area. Through a combination of invited talks, discussions, and posters, we hope to gain a better understanding of the available algorithms and best practices for this space, as well as their inherent limitations.

7:30-8:00	Introduction; Developing a Syllabus for Interactive Machine Learning
8:00-8:30	**Invited Talk: The Need for User Interaction and Feedback in Clustering** Rich Caruana
8:30-9:00	**Invited Talk: Using Personalization to Tame Information Overload** Carlos Guestrin
9:00-9:30	Coffee Break
9:30-10:30	Poster preview talks
3:30-4:00	**Invited Talk: Learning and Evaluating Interactive Segmentation Systems** Pushmeet Kohli
4:00-5:00	Poster session
5:00-5:30	Coffee Break
5:30-6:30	**Invited Talk: Human Machine Co-Learning** Jerry Zhu

6:00-6:30 Open Problems, Challenges, Opportunities

POSTERS

Designing for End-User Interactive Concept Learning in CueFlik
Saleema Amershi,
James Fogarty,
Ashish Kapoor,
Desney Tan,

Clustering with Interactive Feedback
Pranjal Awasthi,
Maria Florina-Balcan,
Avrim Blum,

Online Active Learning Using Conformal Predictions
Vineeth Balasubramanian,
Shayok Chakraborty,
Sethuraman Panchanathan,

Interactively Reviewing Large Image Sets
Scott Blunsden,
Cristina Versino,

Beyond Feature Relevance: Incorporating Rich User Feedback Into Interactive Machine Learning Applications
Krzysztof Gajos,

Multi-Class Active Learning with Binary User Feedback
Ajay J. Joshi,
Fatih Porikli,
Nikolaos Papanikolopoulos,

Interactive Learning on Multiple Binary Classification Problems
Goo Jun,
Alexander Liu,
Joydeep Ghosh,

Leveraging People and Computers for NLP
Karrie Karahalios,
Tony Bergstrom,

Interactively Shaping Agents via Human Feedback: TheTAMER Framework
W. Bradley Knox,
Peter Stone,

Software, Psychophysics, and Selection: Towards Anthropocentric Data Analysis
Joshua M. Lewis,

A Contextual-Bandit Approach to Personalized News Article Recommendation
Lihong Li,
Wei Chu,
John Langford,
Robert E. Schapire,

Interactive Structural Learning for Image and Video Analysis
Xu Miao,
Rajesh P. N. Rao,
Shin'ichi Satoh,

Achieving Small Regret Using an Interactive Learning Approach to Imitation Learning
Stephane Ross,
J. Andrew Bagnell,

Interactive Feature Space Construction
Kevin Small,
Dan Roth,

Online Gradient Descent Using Interactive User Feedback
Yisong Yue,

Actively Cutting Graphs: Think Globally, Cut Locally
Alice X. Zheng,
John Dunagan,
Ashish Kapoor,

December 11, 2009, 07:30-10:30 and 15:30-18:30 Westin: Nordic **WS4**

Analyzing Networks and Learning with Graphs

http://snap.stanford.edu/nipsgraphs2009

Edoardo Airoldi airoldi@fas.harvard.edu
Harvard University
Jon Kleinberg kleinber@cs.cornell.edu
Cornell University
Jure Leskovec jure@cs.stanford.edu
Stanford University
Josh Tenenbaum jbt@mit.edu
MIT

Abstract

Recent research in machine learning and statistics has seen the proliferation of computational methods for analyzing networks and learning with graphs. These methods support progress in many application areas, including the social sciences, biology, medicine, neuroscience, physics, finance, and economics.

The primary goal of the workshop is to actively promote a concerted effort to address statistical, methodological and computational issues that arise when modeling and analyzing large collection of data that are largely represented as static and/or dynamic graphs. To this end, we aim at bringing together researchers from applied disciplines such as sociology, economics, medicine and biology, together with researchers from more theoretical disciplines such as mathematics and physics, within our community of statisticians and computer scientists. Different communities use diverse ideas and mathematical tools; our goal is to to foster cross-disciplinary collaborations and intellectual exchange.

Presentations will include novel graph models, the application of established models to new domains, theoretical and computational issues, limitations of current graph methods and directions for future research.

Check website for schedule
Speakers listed below

Jennifer Chayes, Microsoft Research
Matthew Jackson, Stanford University
Ravi Kumar, Yahoo! Research
Martina Morris, University of Washington
Cosma Shalizi, Carnegie Mellon University

December 11, 2009, 07:30-10:30 and 15:30-18:30 Westin: Callaghan **WS5**

Applications for Topic Models: Text and Beyond

http://nips2009.topicmodels.net

David Blei blei@cs.princeton.edu
Princeton University
Jordan Boyd-Graber jbg@umiacs.umd.edu
University of Maryland
Jonathan Chang jcone@princeton.edu
Facebook
Katherine Heller kheller@gmail.com
University of Cambridge
Hanna Wallach wallach@cs.umass.edu
University of Massachusetts, Amherst

Abstract

Statistical topic models are a class of Bayesian latent variable models, originally developed for analyzing the semantic content of large document corpora. With the increasing availability of other large, heterogeneous data collections, topic models have been adapted to model data from fields as diverse as computer vision, finance, bioinformatics, cognitive science, music, and the social sciences. While the underlying models are often extremely similar, these communities use topic models in different ways in order to achieve different goals. This one-day workshop will bring together topic modeling researchers from multiple disciplines, providing an opportunity for attendees to meet, present their work and share ideas, as well as inform the wider NIPS community about current research in topic modeling. This workshop will address the following specific goals:

- Identify and formalize open research areas
- Propose, explore, and discuss new application areas
- Discuss how best to facilitate transfer of research ideas between application domains
- Direct future work and generate new application areas
- Explore novel modeling approaches and collaborative research directions

The workshop will consist of invited talks by established researchers from multiple research communities, contributed talks, two poster sessions, and a panel discussion.

7:30 - 8:10 **Opening Remarks & Overview/Survey of Topic Modeling Research**
Hanna Wallach

8:10 - 8:50 **Modeling Dynamic Network Tomography**
Eric Xing

8:50 - 9:05 **Contributed Talk: A Probabilistic Topic Model for Music Analysis**
Diane Hu and Lawrence Saul

9:05 - 9:35 Poster Session 1

9:35 - 9:50 **Contributed Talk: Modeling Influence in Text Corpora**
Sean Gerrish and David Blei

9:50 - 10:30	**From Bag-of-Words to Total Scene Understanding: Evolution of Topic Models in Visual Recognition** Fei-Fei Li
3:30 - 4:10	**Topic Models and Adaptor Grammars** Mark Johnson
4:10 - 4:25	**Contributed Talk: Financial Topic Models** Gabriel Doyle and Charles Elkan
4:25 - 4:40	**Contributed Talk: Complexity of Inference in Topic Models** David Sontag and Daniel Roy
4:40 - 4:55	**Contributed Talk: Reconstructing Pompeian Households** David Mimno
4:55 - 5:25	Poster Session 2
5:25 - 6:05	**Modeling Language Learning: Some History, Commentary and News** Thomas Landauer
6:05 - 6:30	Panel Discussion/Closing Remarks
7:00 -	Dinner at Kypriaki (http://www.kypriaki.net/location.html)

Invited Talk: Modeling Dynamic Network Tomography
Eric Xing, Carnegie Mellon University
A plausible representation of the relational information among entities in dynamic systems such as a social community or a living cell is a stochastic network that is topologically rewiring and semantically evolving over time. While there is a rich literature in modeling static or temporally invariant networks, until recently, little has been done toward modeling the dynamic processes underlying rewiring networks. In this talk, I will present a model-based approach to analyze what we will refer to as the dynamic tomography of such time-evolving networks. This approach builds on a time-evolving mixed membership stochastic blockmodel, which is reminiscent of a dynamic topic model. It offers an intuitive but powerful tool to infer and visualize the semantic underpinnings of each actor, such as its social roles or biological functions, underlying the observed network topologies; and it overcomes a number of limitations of many current network inference techniques. I will show some empirical analyses using our model of a social network between monks, a dynamic email communication network between the Enron employees, and a rewiring gene interaction network of fruit fly collected during its full life cycle. In all cases, our model reveals interesting patterns of the dynamic roles of the actors.

Invited Talk: From Bag-of-Words to Total Scene Understanding: Evolution of Topic Models in Visual Recognition
Fei-Fei Li, Stanford University
Starting from the original Bag of Words (BoW) formulation of images, the vision community has come a long way in using topic models to solve visual recognition problems. In this talk, I'll sample a number of representative work by us and others that illustrate this evolution. I will particularly focus on issues that are related to representing and learning high-level visual concepts such as scenes, objects, and pictures-and-words. I will show that by using sophisticated representation of detailed image information, topics models can offer a powerful representation for scene context, object segmentation, annotation, and high-level visual concept understanding. Last but not the least, I will discuss both pros and cons of using topic models for vision.

Invited Talk: Topic Models and Adaptor Grammars
Mark Johnson, Brown University
Adaptor grammars are a non-parametric Bayesian extension of Probabilistic Context-Free Grammars that can express a variety of different Hierarchical Dirichlet or Pitman-Yor Processes. Not surprisingly, Adaptor

Grammars are closely related to Topic Models. After introducing Adaptor Grammars, this talk will focus on the relationship between Adaptor Grammars and Topic Models and describe what they have in common and the ways in which they differ.

Invited Talk: Modeling language learning: some history, commentary and news
Thomas Landauer, University of Colorado at Boulder
History: In the 90s,while trying to overcome the vocabulary problem in information retrieval, we discovered that SVD combines words into passages in much the same way as for humans, thus "Latent Semantic Analysis." A spectrum of past applications will be mentioned. Commentary: The big difference from TOPICS is their objective functions, how words combine versus how they cluster. A popular misunderstanding is that LSA measures how often words occur together in passages. It doesn't. The news: A new LSA application measures the separate growth of knowledge for any individual student for every word in a corpus.

Contributed Talk: A Probabilistic Topic Model for Music Analysis
Diane Hu, University of California, San Diego
Lawrence Saul, University of California, San Diego
We describe a probabilistic model for learning musical key-profiles from symbolic and audio files of polyphonic, classical music. Our model is based on Latent Dirichlet Allocation (LDA), a statistical approach for discovering hidden topics in large corpora of text. In our adaptation of LDA, music files play the role of text documents, groups of musical notes play the role of words, and musical key- profiles play the role of topics. We show how these learnt key-profiles can be used to determine the key of a musical piece and track its harmonic modulations.

Contributed Talk: Modeling Influence in Text Corpora
Sean Gerrish, Princeton University
David Blei, Princeton University
Identifying the most influential documents in a corpus is an important problem in a wide range of fields, ranging from information science and historiography to text summarization and news aggregation. We propose using changes in the linguistic content of these documents over time to predict the importance of individual documents within the collection and describe a dynamic topic model for both quantifying and qualifying the impact of each document in the corpus.

Contributed Talk: Financial Topic Models
Gabriel Doyle, University of California, San Diego
Charles Elkan, University of California, San Diego
We apply topic models to financial data to obtain a more accurate view of economic networks than that supplied by traditional economic statistics. The learned topic models can serve as a substitute for or a complement to more complicated network analysis. Initial results on S&P500 stock market data show that topic models are able to obtain meaningful stock categories from unsupervised data and show promise in revealing network-like statistics about the stock market. We also discuss the characteristics of an ideal topic model for financial data.

Contributed Talk: Complexity of Inference in Topic Models
David Sontag, Massachusetts Institute of Technology
Daniel Roy, Massachusetts Institute of Technology
We consider the computational complexity of finding the MAP assignment of topics to words in Latent Dirichlet Allocation. We show that, when the effective number of topics per document is small, exact inference takes polynomial time. In contrast, we show that, when a document has a large number of topics, finding the MAP assignment in LDA is NP-hard. Our results motivate further study of the structure in real-world topic models, and raise a number of questions about the requirements for accurate inference during both learning and test-time use of topic models.

Contributed Talk: Reconstructing Pompeian Households
David Mimno, University of Massachusetts, Amherst
A database of objects discovered in houses in the Roman city of Pompeii provides a unique view of ordinary life in an ancient city. Experts have used this collection to study the structure of Roman households, exploring the distribution and variability of tasks in architectural spaces, but such approaches are necessarily affected

by modern cultural assumptions. In this study we present a data-driven approach to household archeology, treating it as an unsupervised labeling problem, that attempts to provide a more objective complement to human interpretation.

Spherical Topic Models (Session 1)
Joseph Reisinger,
Austin Waters,
Bryan Silverthorn,
Raymond Mooney,

Undirected Topic Models (Session 1)
Ruslan Salakhutdinov,
Geoffrey Hinton,

Generating Status Hierarchies from Meeting Transcripts Using the Author-Topic Model (Session 1)
David Broniatowski,

Software Analysis with Unsupervised Topic Models (Session 1)
Erik Linstead,
Lindsey Hughes,
Cristina Lopes,
Pierre Baldi,

Adaptation of Topic Model to New Domains Using Recursive Bayes (Session 1))
Ying-Lang Chang,
Jen-Tzung Chien,

Modeling Shared Tastes in Online Communities (Session 1)
Laura Dietz,

Application of Lexical Topic Models to Protein Interaction Sentence Prediction (Session 1)
Tamara Polajnar,
Mark Girolami,

A Time and Space Dependent Topic Model for Unsupervised Activity Perception in Video (Session 1)
Eric Wang,
Lawrence Carin,

Audio Scene Understanding using Topic Models (Session 1)
Samuel Kim,
Shiva Sundaram,
Panayiotis Georgiou,
Shrikanth Narayanan,

Stopwords and Stylometry: A Latent Dirichlet Allocation Approach (Session 1)
Arun R.,
Saradha R.,
V. Suresh,

C.E. Veni Madhavan,
M. Narasimha Murty,

Learning to Summarize using Coherence (Session 1)
Pradipto Das,
Rohini Srihari,

Focused Topic Models (Session 1)
Sinead Williamson,
Chong Wang,
Katherine Heller,
David Blei,

Applications of Topics Models To Analysis of Disaster-Related Twitter Data (Session 1)
Kirill Kireyev,
Leysia Palen,
Kenneth Anderson,

A Semantic Question / Answering System using Topic Models (Session 1)
Asli Celikyilmaz,

Finding Topics in Emails: Is LDA Enough? (Session 1)
Shafiq Joty,
Giuseppe Carenini,
Gabriel Murray,
Raymond Ng,

Topic Models for Audio Mixture Analysis (Session 2)
Paris Smaragdis,
Madhusudana Shashanka,
Bhiksha Raj,

Timelines: Revealing The Birth and Evolution of Ideas In Text Stream Using Infinite Dynamic Topic Models (Session 2)
Amr Ahmed,
Eric Xing,

Modeling Concept-Attribute Structure (Session 2)
Joseph Reisinger,
Marius Pasca,

Segmented Topic Model for Text Classification and Speech Recognition (Session 2)
Chuang-Hua Chueh,
Jen-Tzung Chien,

Writer Identification in Offline Handwriting Using Topic Models (Session 2)
Anurag Bhardwaj,
Manavender Malgireddy,

Venu Govindaraju,

Implicit Communication Detection Using Topics Model on Asynchronous Communication Data (Session 2)
Charles Panaccione,
Peter Folz,

Topic Modeling for the Social Sciences (Session 2)
Daniel Ramage,
Evan Rosen,
Jason Chuang,
Chris Manning,
Daniel McFarland,

Author Disambiguation: A Nonparametric Topic and Co-authorship Model (Session 2)
Andrew Dai,
Amos Storkey,

Speeding Up Gibbs Sampling by Variable Grouping (Session 2)
Evgeniy Bart,

Modeling Tag Dependencies in Tagged Documents (Session 2)
Timothy Rubin,
America Holloway,
Padhraic Smyth,
Mark Steyvers,

Data Portraiture and Topic Models (Session 2)
Aaron Zinman,
Doug Fritz,

Who Talks to Whom: Modeling Latent Structures in Dialogue Documents (Session 2)
Bailu Ding,
Jiang-Ming Yang,
Chong Wang,
Rui Cai,
Zhiwei Li,
Lei Zhang,

Topic Models for Semantically Annotated Document Collections (Session 2)
Markus Bundschus,
Volker Tresp,
Hans-Peter Kriegel,

December 12, 2009, 07:30-10:30 and 15:30-18:30 Hilton: Mt. Currie North **WS6**

Approximate Learning of Large Scale Graphical Models: Theory and Applications

http://www.cs.toronto.edu/~rsalakhu/workshop_nips2009/

Ruslan Salakhutdinov rsalakhu@mit.edu
CSAIL, MIT
Amir Globerson gamir@cs.huji.ac.il
The Hebrew University of Jerusalem
David Sontag dsontag@mit.edu
CSAIL, MIT

Abstract

Undirected graphical models provide a powerful framework for representing dependency structure between random variables. Learning the parameters of undirected models plays a crucial role in solving key problems in many machine learning applications, including natural language processing, visual object recognition, speech perception, information retrieval, computational biology, and many others. Learning in undirected graphical models of large treewidth is difficult because of the hard inference problem induced by the partition function for maximum likelihood learning, or by finding the MAP assignment for margin-based loss functions. The goal of this workshop is to assess the current state of the field and explore new directions for both theoretical foundations and empirical applications. The workshop will be built around the following key topics: the use of approximate inference in learning, learning in graphical models with latent variables, learning in models with deep architectures, theoretical analysis/guarantees of learning algorithms, as well as their scalability and success at real-world applications. Through a series of invited talks and a panel discussion, this workshop will bring together machine learning researchers working on approximate inference in learning to discuss key challenges and to identify promising directions for future investigation.

7:30-7:40	**Introduction** Organizers
7:40-8:20	**Approximate Inference in Natural Language Processing** Noah Smith
8:20-9:00	**TBA** Ben Taskar
9:00-9:10	Coffee Break
9:10-9:50	**Jointly Maximum Margin and Maximum Entropy Learning of Graphical Models** Eric Xing
9:50-10:30	**Large-Scale Learning and Inference: What We Have Learned with Markov Logic Networks** Pedro Domingos
3:30-4:00	**Parameter Learning using Approximate MAP Inference** Pawan Kumar
4:00-4:40	**Training Structured Predictors for Novel Loss Functions** David McAllester
4:40-5:10	**TBA** Bill Freeman

5:10-5:20 Coffee Break

5:20-6:00 **Image Retrieval using Short Binary Codes**
Geoffrey Hinton

6:00:-6:30 **Discussion and Conclusions**
Organizers

Approximate Inference in Natural Language Processing
Noah Smith, Carnegie Mellon University

TBA
Ben Taskar, University of Pennsylvania

Jointly Maximum Margin and Maximum Entropy Learning of Graphical Models
Eric Xing, Carnegie Mellon University
Inferring structured predictions based on correlated covariates remains a central problem in many fields, including NLP, computer vision, and computational biology. Popular paradigms for training structured input/output models include the maximum (conditional) likelihood estimation, which leads to the well-known CRF; and the max-margin learning, which leads to the structured SVM (a.k.a. M3N), each enjoys some advantages, as well as weaknesses. In this talk, I present a new general framework called Maximum Entropy Discrimination Markov Networks (MEDN), which integrates the margin-based and likelihood-based approaches and combines and extends their merits. This new learning paradigm naturally facilitates integration of the generative and discriminative principles under a unified framework, and the basic strategies can be generalized to learn arbitrary graphical models, such as the generative Bayesian networks or models with structured hidden variables. I will discuss a number of theoretical properties of this model, and show applications of MEDN to learning fully supervised structured i/o model, max-margin structured i/o models with hidden variables, and a max-margin LDA model for jointly discovering discriminative latent topic representations and predicting document label/score of text documents, with compelling performance in each case.

Large-Scale Learning and Inference: What We Have Learned with Markov Logic Networks
Pedro Domingos, University of Washington
Markov logic allows very large and rich graphical models to be compactly specified. Current learning and inference algorithms for Markov logic can routinely handle models with millions of variables, billions of features, thousands of latent variables, and strong dependencies. In this talk I will give an overview of the main ideas in these algorithms, including weighted satisfiability, MCMC with deterministic dependencies, lazy inference, lifted inference, relational cutting planes, scaled conjugate gradient, relational clustering and relational pathfinding. I will also discuss the lessons learned in developing successive generations of these algorithms and promising ideas for the next round of scaling up. (Joint work with Stanley Kok, Daniel Lowd, Hoifung Poon, Matt Richardson, Parag Singla, Marc Sumner, and Jue Wang.)

Parameter Learning using Approximate MAP Inference
Pawan Kumar, Stanford University
In recent years, machine learning has seen the development of a series of algorithms for parameter learning that avoid estimating the partition function and instead, rely on accurate approximate MAP inference. Within this framework, we consider two new topics.

In the first part, we discuss parameter learning in a semi-supervised scenario. Specifically, we focus on a region-based scene segmentation model that explains an image in terms of its underlying regions (a set of connected pixels that provide discriminative features) and their semantic labels (such as sky, grass or foreground). While it is easy to obtain (partial) ground-truth labeling for the pixels of a training image, it is not possible for a human annotator to provide us with the best set of regions (those that result in the most discriminative features). To address this issue, we develop a novel iterative MAP inference algorithm which selects the best subset of regions from a large dictionary using convex relaxations. We use our algorithm to "complete" the ground-truth labeling (i.e. infer the regions) which allows us to employ the highly successful max-margin training regime. We compare our approach with the state of the art methods and demonstrate significant improvements.

In the second part, we discuss a new learning framework for general log-linear models based on contrastive objectives. A contrastive objective considers a set of "interesting" assignments and attempts to push up the probability of the correct instantiation at the expense of the other interesting assignments. In contrast to our approach, related methods such as pseudo-likelihood and contrastive divergence compare the correct instantiation only to nearby instantiations, which can be problematic when there is a high-scoring instantiation far away from the correct one. We present some of the theoretical properties and practical advantages of our method, including the ability to learn a log-linear model using only (approximate) MAP inference. We also show results of applying our method to some simple synthetic examples, where it significantly outperforms pseudo-likelihood.

Training Structured Predictors for Novel Loss Functions

David McAllester, Toyota Technological Institute at Chicago

As a motivation we consider the PASCAL image segmentation challenge. Given an image and a target class, such as person, the challenge is to segment the image into regions occupied by objects in that class (person foreground) and regions not occupied by that class (non-person background). At the present state of the art the lowest pixel error rate is achieved by predicting all background. However, the challenge is evaluated with an intersection over union score with the property that the all-background prediction scores zero. This raises the question of how one incorporates a particular loss function into the training of a structured predictor. A standard approach is to incorporate the desired loss into the structured hinge loss and observe that, for any loss, the structured hinge loss is an upper bound on the desired loss. However, this upper bound is quite loose and it is far from clear that the structured hinge loss is an appropriate or useful way to handle the PASCAL evaluation measure. This talk reviews various approaches to this problem and presents a new training algorithm we call the good-label-bad-label algorithm. We prove that in the data-rich regime the good-label-bad-label algorithm follows the gradient of the training loss assuming only that we can perform inference in the given graphical model. The algorithm is structurally similar to, but significantly different from, stochastic subgradient descent on the structured hinge loss (which does not follow the loss gradient).

Image retrieval using short binary codes

Geoffrey Hinton, University of Toronto

The obvious way to find images that are semantically similar to a query image is to solve the object recognition problem. In the meantime, it is possible to extract a feature vector from each image and to retrieve images with similar features. If the features are binary they are cheap to store and match. If they are also highly abstract (e.g. indoor vs outdoor) and roughly orthogonal they will work well for image retrieval. I will describe a method of extracting such binary features using deep belief networks. I will then show how binary codes can be used retrieve a shortlist of semantically similar images extremely rapidly in a time that is independent of the size of the database. This is work in progress with Alex Krizhevsky.

TBA

Bill Freeman, MIT

DECEMBER 12, 2009, 07:30-10:30 AND 15:30-18:30 WESTIN: EMERALD A **WS7**

Nonparametric Bayes

http://npbayes-2009.wikidot.com

Dilan Gorur dilan@gatsby.ucl.ac.uk
GATSBY COMPUTATIONAL NEUROSCIENCE UNIT, UCL
Francois Caron Francois.Caron@inria.fr
INRIA BORDEAUX SUD-OUEST
Yee Whye Teh ywteh@gatsby.ucl.ac.uk
GATSBY COMPUTATIONAL NEUROSCIENCE UNIT, UCL
David Dunson dunson@stat.duke.edu
DUKE UNIVERSITY
Zoubin Ghahramani zoubin@eng.cam.ac.uk
UNIVERSITY OF CAMBRIDGE
Michael I. Jordan jordan@cs.berkeley.edu
UNIVERSITY OF CALIFORNIA AT BERKELEY

Abstract

One of the major problems driving current research in statistical machine learning is the search for ways to exploit highly-structured models that are both expressive and tractable. Bayesian nonparametrics provides a framework for developing robust and flexible models that can accurately represent the complex structure in the data. Model flexibility is achieved by assigning priors with unbounded capacity and overfitting is prevented by integrating out all parameters and latent variables.

We aim to bring together researchers to create a forum for discussing recent advances in Bayesian nonparametrics, to understand better the asymptotic properties of the models and to inspire research on new techniques for better models and inference algorithms. The workshop will focus mainly on two important issues. 1) Theoretical properties of complex Bayesian nonparametric models, in particular asymptotics (e.g. consistency, rates of convergence, and Bernstein von-Mises results). 2) Practical matters to enable the use of Bayesian nonparametrics in real world applications such as developing general purpose software, discussion of an objective or empirical Bayes treatment. Each focus will be given a specific session during the workshop.

7:30-8:10	**A Brief Overview of Nonparametric Bayesian Models** ZOUBIN GHAHRAMANI
8:10-8:30	**Conjugate Projective Limits** PETER ORBANZ
8:30-9:00	**Convergence of posterior distributions in infinite dimension — a decade of success stories** SUBHASHIS GHOSHAL
9:00-9:40	Poster Session & Coffee break
9:40-10:00	**Approximation of conditional densities by smooth mixtures of regressions** ANDRIY NORETS
10:00-10:30	Discussion
15:30-16:00	**Nonparametric Bayesian models of human cognition** THOMAS L. GRIFFITHS

16:00-16:30	**Practical Aspects of Bayesian Nonparametrics** Alejandro Jara
16:30-17:00	**On the role of sequential Monte Carlo algorithms for complex nonparametric mixture models** Abel Rodriguez
17:00-17:20	Poster Session & Coffee break
17:20-17:50	**Modeling dependent distributions with Gaussian processes** Surya Tokdar
17:50-18:30	Discussion and wrap up

A Brief Overview of Nonparametric Bayesian Models
Zoubin Ghahramani, University of Cambridge
The flexibility of nonparametric Bayesian (NPB) methods for data modelling has generated an explosion of interest in the last decade in both Statistics and Machine Learning communities. I will give an overview of some of the main NPB models, and focus on the relationships between them. I plan to give a whirlwind tour of the Gaussian process, Dirichlet process (DP) and Beta process, the associated Chinese restaurant and Indian buffet, times series models such as the infinite HMM (sometimes called the HDP-HMM), hierarchical models such as Kingman's coalescent and the Dirichlet diffusion tree, dependent models such as the depedent Dirichlet process, and other topics such as completely random measures and stick-breaking constructions, time permitting.

Conjugate Projective Limits
Peter Orbanz, University of Cambridge
Bayesian nonparametric models can be regarded as Bayesian models on infinite-dimensional spaces. These infinite-dimensional distributions can be constructed from finite-dimensional ones using the tools of stochastic process theory. An example is the construction of the Gaussian process constructed from Gaussian distributions. My talk will address the question which finite-dimensional distributions are suitable for the construction of nonparametric Bayesian models with useful statistical properties. By a proper choice of finite-dimensional models used in the construction, the nonparametric Bayesian model can be guaranteed to be conjugate, and to have a sufficient statistic. I will briefly discuss for which models these constructions follow a generic recipe, and for which cases we have to expect mathematical complications.

Convergence of posterior distributions in infinite dimension — a decade of success stories
Subhashis Ghoshal, North Carolina State University
It was long realized that for parametric inference problems, posterior distributions based on a large class of reasonable prior distributions possess very desirable large sample convergence properties, even if viewed from purely frequentist angles. For nonparametric or semiparametric problems, the story gets complicated, but still good frequentist convergence properties are enjoyed by Bayesian methods if a prior distribution is carefully constructed. The last ten years have witnessed the most significant progress in the study of consistency, convergence rates and finer frequentist properties. It is now well understood that the properties are controlled by the concentration of prior mass near the true value, as well as the effective size of the model, measured in terms of the metric entropy. Results have poured in for independent and identically distributed data, independent and non-identically distributed data and dependent data, as well as for a wide spectrum of inference problems such as density estimation, nonparametric regression, classification, and so on. Nonparametric mixtures, random series and Gaussian processes play particularly significant roles in the construction of the "right" priors. In this talk, we try to outline the most significant developments that took place in the last decade. In particular, we emphasize the ability of the posterior distribution to effortlessly choose the right model and adapt to the unknown level of smoothness.

Approximation of conditional densities by smooth mixtures of regressions
Andriy Norets, Princeton University

This paper shows that large nonparametric classes of conditional multivariate densities can be approximated in the Kullback-Leibler distance by different specifications of finite mixtures of normal regressions in which normal means and variances and mixing probabilities can depend on variables in the conditioning set (covariates). These models are a special case of models known as mixtures of experts in statistics and computer science literature. Flexible specifications include models in which only mixing probabilities, modeled by multinomial logit, depend on the covariates and, in the univariate case, models in which only means of the mixed normals depend flexibly on the covariates. Modeling the variance of the mixed normals by flexible functions of the covariates can weaken restrictions on the class of the approximable densities. Obtained results can be generalized to mixtures of general location scale densities. Rates of convergence and easy to interpret bounds are also obtained for different model specifications. These approximation results can be useful for proving consistency of Bayesian and maximum likelihood density estimators based on these models. The results also have interesting implications for applied researchers.

Nonparametric Bayesian models of human cognition

Thomas L. Griffiths, University of California, Berkeley

Human learners are capable of adapting the way representations of the properties of objects in response to statistical information. For example, we can from clusters based on visual information, and decide what features of objects are important based on the other objects to which we compare them. Nonparametric Bayesian models provide a way to provide a rational account of such representational flexibility, indicating how an ideal learner would interpret relevant statistical information. In particular, by allowing hypothesis spaces of unbounded complexity, nonparametric Bayesian models potentially provide a more satisfying account of the rich representations entertained by human learners. I will summarize the results of recent studies examining how ideas from nonparametric Bayesian statistics lead to models of human cognition, and discuss some of the challenges that thinking about human learning poses for this approach.

Practical Aspects of Bayesian Nonparametrics

Alejandro Jara, Universidad de Concepcion, Chile

In this talk I will discuss practical aspects associated to the implementation of Bayesian semi- and non-parametrics models. The emphasis of the talk will be on three different aspects: (A) the discussion of the most popular Bayesian nonparametric models, (B) the role of the parameter identification in Bayesian semiparametric model building, and (C) computational issues associated to Bayesian nonparametric inference. In (A), the most popular Bayesian methods for function estimation are reviewed. In (B), I'll discuss the limitations of the statistical inferences in Bayesian semiparametric models. Specifically, I'll discuss the role of the parameter identifiability in the model specification and show that, although the lack of identification present no difficulties to a Bayesian analysis in the sense that a prior is transformed into a posterior using the sampling model and the probability calculus, if the interest focuses on a unidentified parameter then such formal assurances have little practical value. From a computational point of view, identification problems imply ridges in the posterior distribution and MCMC methods can be difficult to implement in these situations. Finally, since the main obstacle for the practical use of Bayesian nonparametric methods has been the lack of easy-to-use estimation tools, I will introduce a simple, yet comprehensive, set of programs for the implementation of Bayesian non- and semi-parametric models in R, DPpackage. I will discuss the general syntax and design philosophy of DPpackage and describe the currently available functions. The main features and usage of DPpackage will be illustrated using simulated and real data analyses.

On the role of sequential Monte Carlo algorithms for complex nonparametric mixture models

Abel Rodriguez, University of California, Santa Cruz

This talk will explore the role that sequential Monte Carlo (SMC) algorithm can play in learning complex Bayesian nonparametric mixture models. In particular the talk revolves around four themes: 1) models that are sequential in nature (e.g., the infinite hidden Markov model), 2) models that are not sequential in nature but where more standard Monte Carlo algorithm can be difficult to implement (e.g., the nested Dirichlet process and some of its extensions), 3) problems where model comparison is a key inference issue, and 4) problems with large sample sizes where parallelization (and particularly graphical processing units, GPUs) can provide dramatic speed-ups.

Modeling dependent distributions with Gaussian processes
Surya Tokdar, Duke University
I would talk about the use of Gaussian processes (GP) to model a family of dependent distributions in a non-parametric, non-Gaussian setting. Examples include density regression, spatial GLM, multi-site discrete-valued time series, etc. All these models can be induced by a Gaussian process on the product space of the variable of interest (or a latent version of it) and the variable that indexes the family membership (covariates, site locations etc). Dependence among the distributions is easily encoded through the covariance function of this Gaussian process. I'd briefly highlight nice theoretical properties of such processes and then discuss in detail issues with model fitting, particularly with MCMC exploration of the resulting posterior. I'd start with the well-known big-N problem of Gaussian processes and talk about the Predictive Process (PP) approach. Then I will focus on the special needs of the product-space construction and how to adapt PP to handle this. Next I'd stress on the often neglected issue of mixing of the GP covariance parameters. This mixing behaves notoriously when the underlying GP function cannot be integrated out (as is done in regression or spatial models with Gaussian errors). I'd elaborate on a useful strategy to overcome this. I'd end with some further thoughts on GP and PP for complex, high-dimensional models and functional data analysis.

POSTERS

Building Graph Structures from the Beta Process
Noel Welsh,
Jeremy Wyatt,
One of the major problems driving current research in statistical machine learning is the search for ways to exploit highly-structured models that are both expressive and tractable. Bayesian nonparametrics provides a framework for developing robust and flexible models that can accurately represent the complex structure in the data. Model flexibility is achieved by assigning priors with unbounded capacity and overfitting is prevented by integrating out all parameters and latent variables.

We aim to bring together researchers to create a forum for discussing recent advances in Bayesian nonparametrics, to understand better the asymptotic properties of the models and to inspire research on new techniques for better models and inference algorithms. The workshop will focus mainly on two important issues. 1) Theoretical properties of complex Bayesian nonparametric models, in particular asymptotics (e.g. consistency, rates of convergence, and Bernstein von-Mises results). 2) Practical matters to enable the use of Bayesian nonparametrics in real world applications such as developing general purpose software, discussion of an objective or empirical Bayes treatment. Each focus will be given a specific session during the workshop.

Collapsed Variational Inference for Time-varying Dirichlet Process Mixture Models
Amr Ahmed,
Eric Xing,
One of the major problems driving current research in statistical machine learning is the search for ways to exploit highly-structured models that are both expressive and tractable. Bayesian nonparametrics provides a framework for developing robust and flexible models that can accurately represent the complex structure in the data. Model flexibility is achieved by assigning priors with unbounded capacity and overfitting is prevented by integrating out all parameters and latent variables.

We aim to bring together researchers to create a forum for discussing recent advances in Bayesian nonparametrics, to understand better the asymptotic properties of the models and to inspire research on new techniques for better models and inference algorithms. The workshop will focus mainly on two important issues. 1) Theoretical properties of complex Bayesian nonparametric models, in particular asymptotics (e.g. consistency, rates of convergence, and Bernstein von-Mises results). 2) Practical matters to enable the use of Bayesian nonparametrics in real world applications such as developing general purpose software, discussion of an objective or empirical Bayes treatment. Each focus will be given a specific session during the workshop.

Conditional Simultaneous Draws from Hierarchical Chinese Restaurant Processes
Takaki Makino,
Shunsuke Takei,

Daichi Mochihashi,
Issei Sato,
Toshihisa Takagi,
One of the major problems driving current research in statistical machine learning is the search for ways to exploit highly-structured models that are both expressive and tractable. Bayesian nonparametrics provides a framework for developing robust and flexible models that can accurately represent the complex structure in the data. Model flexibility is achieved by assigning priors with unbounded capacity and overfitting is prevented by integrating out all parameters and latent variables.

We aim to bring together researchers to create a forum for discussing recent advances in Bayesian nonparametrics, to understand better the asymptotic properties of the models and to inspire research on new techniques for better models and inference algorithms. The workshop will focus mainly on two important issues. 1) Theoretical properties of complex Bayesian nonparametric models, in particular asymptotics (e.g. consistency, rates of convergence, and Bernstein von-Mises results). 2) Practical matters to enable the use of Bayesian nonparametrics in real world applications such as developing general purpose software, discussion of an objective or empirical Bayes treatment. Each focus will be given a specific session during the workshop.

Cross-categorization: A Method for Discovering Multiple Overlapping Clusterings
Vikash Mansinghka,
Eric Jonas,
Cap Petschulat,
Beau Cronin,
Patrick Shafto,
Joshua Tenenbaum,
One of the major problems driving current research in statistical machine learning is the search for ways to exploit highly-structured models that are both expressive and tractable. Bayesian nonparametrics provides a framework for developing robust and flexible models that can accurately represent the complex structure in the data. Model flexibility is achieved by assigning priors with unbounded capacity and overfitting is prevented by integrating out all parameters and latent variables.

We aim to bring together researchers to create a forum for discussing recent advances in Bayesian nonparametrics, to understand better the asymptotic properties of the models and to inspire research on new techniques for better models and inference algorithms. The workshop will focus mainly on two important issues. 1) Theoretical properties of complex Bayesian nonparametric models, in particular asymptotics (e.g. consistency, rates of convergence, and Bernstein von-Mises results). 2) Practical matters to enable the use of Bayesian nonparametrics in real world applications such as developing general purpose software, discussion of an objective or empirical Bayes treatment. Each focus will be given a specific session during the workshop.

Fast Search for Infinite Latent Feature Models
Piyush Rai,
Hal Daume III,
One of the major problems driving current research in statistical machine learning is the search for ways to exploit highly-structured models that are both expressive and tractable. Bayesian nonparametrics provides a framework for developing robust and flexible models that can accurately represent the complex structure in the data. Model flexibility is achieved by assigning priors with unbounded capacity and overfitting is prevented by integrating out all parameters and latent variables.

We aim to bring together researchers to create a forum for discussing recent advances in Bayesian nonparametrics, to understand better the asymptotic properties of the models and to inspire research on new techniques for better models and inference algorithms. The workshop will focus mainly on two important issues. 1) Theoretical properties of complex Bayesian nonparametric models, in particular asymptotics (e.g. consistency, rates of convergence, and Bernstein von-Mises results). 2) Practical matters to enable the use of Bayesian nonparametrics in real world applications such as developing general purpose software, discussion of an objective or empirical Bayes treatment. Each focus will be given a specific session during the workshop.

Metric Entropy and Gaussian Bandits
Steffen Grlder,
Jean-Yves Audibert,
Manfred Opper,
John Shawe-Taylor,
One of the major problems driving current research in statistical machine learning is the search for ways to exploit highly-structured models that are both expressive and tractable. Bayesian nonparametrics provides a framework for developing robust and flexible models that can accurately represent the complex structure in the data. Model flexibility is achieved by assigning priors with unbounded capacity and overfitting is prevented by integrating out all parameters and latent variables.

We aim to bring together researchers to create a forum for discussing recent advances in Bayesian nonparametrics, to understand better the asymptotic properties of the models and to inspire research on new techniques for better models and inference algorithms. The workshop will focus mainly on two important issues. 1) Theoretical properties of complex Bayesian nonparametric models, in particular asymptotics (e.g. consistency, rates of convergence, and Bernstein von-Mises results). 2) Practical matters to enable the use of Bayesian nonparametrics in real world applications such as developing general purpose software, discussion of an objective or empirical Bayes treatment. Each focus will be given a specific session during the workshop.

Modeling Associations among Multivariate Longitudinal Categorical Variables in Survey Data: a Semiparametric Bayesian Approach
Sylvie Tchumtchoua,
Dipak K. Dey,
One of the major problems driving current research in statistical machine learning is the search for ways to exploit highly-structured models that are both expressive and tractable. Bayesian nonparametrics provides a framework for developing robust and flexible models that can accurately represent the complex structure in the data. Model flexibility is achieved by assigning priors with unbounded capacity and overfitting is prevented by integrating out all parameters and latent variables.

We aim to bring together researchers to create a forum for discussing recent advances in Bayesian nonparametrics, to understand better the asymptotic properties of the models and to inspire research on new techniques for better models and inference algorithms. The workshop will focus mainly on two important issues. 1) Theoretical properties of complex Bayesian nonparametric models, in particular asymptotics (e.g. consistency, rates of convergence, and Bernstein von-Mises results). 2) Practical matters to enable the use of Bayesian nonparametrics in real world applications such as developing general purpose software, discussion of an objective or empirical Bayes treatment. Each focus will be given a specific session during the workshop.

Nonparametric Bayesian Author Disambiguation
Andrew M. Dai,
Amos J. Storkey,
One of the major problems driving current research in statistical machine learning is the search for ways to exploit highly-structured models that are both expressive and tractable. Bayesian nonparametrics provides a framework for developing robust and flexible models that can accurately represent the complex structure in the data. Model flexibility is achieved by assigning priors with unbounded capacity and overfitting is prevented by integrating out all parameters and latent variables.

We aim to bring together researchers to create a forum for discussing recent advances in Bayesian nonparametrics, to understand better the asymptotic properties of the models and to inspire research on new techniques for better models and inference algorithms. The workshop will focus mainly on two important issues. 1) Theoretical properties of complex Bayesian nonparametric models, in particular asymptotics (e.g. consistency, rates of convergence, and Bernstein von-Mises results). 2) Practical matters to enable the use of Bayesian nonparametrics in real world applications such as developing general purpose software, discussion of an objective or empirical Bayes treatment. Each focus will be given a specific session during the workshop.

Nonparametric Bayesian Co-Clustering Ensembles
Pu Wang,
Carlotta Domeniconi,
Kathryn B. Laskey,
One of the major problems driving current research in statistical machine learning is the search for ways to exploit highly-structured models that are both expressive and tractable. Bayesian nonparametrics provides a framework for developing robust and flexible models that can accurately represent the complex structure in the data. Model flexibility is achieved by assigning priors with unbounded capacity and overfitting is prevented by integrating out all parameters and latent variables.

We aim to bring together researchers to create a forum for discussing recent advances in Bayesian nonparametrics, to understand better the asymptotic properties of the models and to inspire research on new techniques for better models and inference algorithms. The workshop will focus mainly on two important issues. 1) Theoretical properties of complex Bayesian nonparametric models, in particular asymptotics (e.g. consistency, rates of convergence, and Bernstein von-Mises results). 2) Practical matters to enable the use of Bayesian nonparametrics in real world applications such as developing general purpose software, discussion of an objective or empirical Bayes treatment. Each focus will be given a specific session during the workshop.

Nonparametric Bayesian Local Partition Model for Multi-task Reinforcement Learning in POMDPs
Chenghui Cai,
Xuejun Liao,
Lawrence Carin,
One of the major problems driving current research in statistical machine learning is the search for ways to exploit highly-structured models that are both expressive and tractable. Bayesian nonparametrics provides a framework for developing robust and flexible models that can accurately represent the complex structure in the data. Model flexibility is achieved by assigning priors with unbounded capacity and overfitting is prevented by integrating out all parameters and latent variables.

We aim to bring together researchers to create a forum for discussing recent advances in Bayesian nonparametrics, to understand better the asymptotic properties of the models and to inspire research on new techniques for better models and inference algorithms. The workshop will focus mainly on two important issues. 1) Theoretical properties of complex Bayesian nonparametric models, in particular asymptotics (e.g. consistency, rates of convergence, and Bernstein von-Mises results). 2) Practical matters to enable the use of Bayesian nonparametrics in real world applications such as developing general purpose software, discussion of an objective or empirical Bayes treatment. Each focus will be given a specific session during the workshop.

Power-Law Unbounded Markov Prediction
Jan Gasthaus,
Frank Wood,
Yee Whye Teh,
One of the major problems driving current research in statistical machine learning is the search for ways to exploit highly-structured models that are both expressive and tractable. Bayesian nonparametrics provides a framework for developing robust and flexible models that can accurately represent the complex structure in the data. Model flexibility is achieved by assigning priors with unbounded capacity and overfitting is prevented by integrating out all parameters and latent variables.

We aim to bring together researchers to create a forum for discussing recent advances in Bayesian nonparametrics, to understand better the asymptotic properties of the models and to inspire research on new techniques for better models and inference algorithms. The workshop will focus mainly on two important issues. 1) Theoretical properties of complex Bayesian nonparametric models, in particular asymptotics (e.g. consistency, rates of convergence, and Bernstein von-Mises results). 2) Practical matters to enable the use of Bayesian nonparametrics in real world applications such as developing general purpose software, discussion of an objective or empirical Bayes treatment. Each focus will be given a specific session during the workshop.

Predictive computable iff posterior computable
Cameron E. Freer,
Daniel M. Roy,
One of the major problems driving current research in statistical machine learning is the search for ways to exploit highly-structured models that are both expressive and tractable. Bayesian nonparametrics provides a framework for developing robust and flexible models that can accurately represent the complex structure in the data. Model flexibility is achieved by assigning priors with unbounded capacity and overfitting is prevented by integrating out all parameters and latent variables.

We aim to bring together researchers to create a forum for discussing recent advances in Bayesian nonparametrics, to understand better the asymptotic properties of the models and to inspire research on new techniques for better models and inference algorithms. The workshop will focus mainly on two important issues. 1) Theoretical properties of complex Bayesian nonparametric models, in particular asymptotics (e.g. consistency, rates of convergence, and Bernstein von-Mises results). 2) Practical matters to enable the use of Bayesian nonparametrics in real world applications such as developing general purpose software, discussion of an objective or empirical Bayes treatment. Each focus will be given a specific session during the workshop.

System Identification of Gaussian Process Dynamic Systems
Ryan Turner,
Marc Peter Deisenroth,
Carl Edward Rasmussen,
One of the major problems driving current research in statistical machine learning is the search for ways to exploit highly-structured models that are both expressive and tractable. Bayesian nonparametrics provides a framework for developing robust and flexible models that can accurately represent the complex structure in the data. Model flexibility is achieved by assigning priors with unbounded capacity and overfitting is prevented by integrating out all parameters and latent variables.

We aim to bring together researchers to create a forum for discussing recent advances in Bayesian nonparametrics, to understand better the asymptotic properties of the models and to inspire research on new techniques for better models and inference algorithms. The workshop will focus mainly on two important issues. 1) Theoretical properties of complex Bayesian nonparametric models, in particular asymptotics (e.g. consistency, rates of convergence, and Bernstein von-Mises results). 2) Practical matters to enable the use of Bayesian nonparametrics in real world applications such as developing general purpose software, discussion of an objective or empirical Bayes treatment. Each focus will be given a specific session during the workshop.

Transfer Learning in Human Categorization
Kevin R. Canini,
Thomas L. Griffiths,
One of the major problems driving current research in statistical machine learning is the search for ways to exploit highly-structured models that are both expressive and tractable. Bayesian nonparametrics provides a framework for developing robust and flexible models that can accurately represent the complex structure in the data. Model flexibility is achieved by assigning priors with unbounded capacity and overfitting is prevented by integrating out all parameters and latent variables.

We aim to bring together researchers to create a forum for discussing recent advances in Bayesian nonparametrics, to understand better the asymptotic properties of the models and to inspire research on new techniques for better models and inference algorithms. The workshop will focus mainly on two important issues. 1) Theoretical properties of complex Bayesian nonparametric models, in particular asymptotics (e.g. consistency, rates of convergence, and Bernstein von-Mises results). 2) Practical matters to enable the use of Bayesian nonparametrics in real world applications such as developing general purpose software, discussion of an objective or empirical Bayes treatment. Each focus will be given a specific session during the workshop.

Tree-Structured Stick Breaking Processes for Hierarchical Modeling
Ryan Prescott Adams,

Zoubin Ghahramani,
Michael I. Jordan,
One of the major problems driving current research in statistical machine learning is the search for ways to exploit highly-structured models that are both expressive and tractable. Bayesian nonparametrics provides a framework for developing robust and flexible models that can accurately represent the complex structure in the data. Model flexibility is achieved by assigning priors with unbounded capacity and overfitting is prevented by integrating out all parameters and latent variables.

We aim to bring together researchers to create a forum for discussing recent advances in Bayesian nonparametrics, to understand better the asymptotic properties of the models and to inspire research on new techniques for better models and inference algorithms. The workshop will focus mainly on two important issues. 1) Theoretical properties of complex Bayesian nonparametric models, in particular asymptotics (e.g. consistency, rates of convergence, and Bernstein von-Mises results). 2) Practical matters to enable the use of Bayesian nonparametrics in real world applications such as developing general purpose software, discussion of an objective or empirical Bayes treatment. Each focus will be given a specific session during the workshop.

DECEMBER 12, 2009, 07:30-10:30 AND 15:30-18:30 WESTIN: ALPINE BC **WS8**

Bounded-rational analyses of human cognition: Bayesian models, approximate inference, and the brain

http://www.mit.edu/~ndg/NIPS09Workshop.html

Noah Goodman ndg@mit.edu
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Ed Vul evul@mit.edu
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Tom Griffiths tom_griffiths@berkeley.edu
UNIVERSITY OF CALIFORNIA AT BERKELEY
Josh Tenenbaum jbt@mit.edu
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Abstract

Bayesian accounts of human cognition have enjoyed much success in recent years; however, these computational models usually assume unbounded cognitive resources available for computation. A key challenge for the Bayesian approach to cognition is to describe the algorithms used to carry out approximate probabilistic inference using the bounded computational resources of the human brain. One appealing solution is the suggestion that instead of manipulating whole distributions, humans make inferences by drawing a small number of samples from the appropriate posterior distribution. Such Monte Carlo algorithms are attractive for boundedly optimal computation because they minimize the curse of dimensionality when scaling to complex inferences, they use resources efficiently, and they degrade gracefully when many samples cannot be obtained. The aim of this workshop is to connect Bayesian models of cognition, human cognitive processes, neural implementations of Bayesian inference, and modern inference algorithms. Thus asking: Can we make precise predictions about the dynamics of human cognition from state-of-the-art inference algorithms? Can engineering applications of machine learning improve by knowing which tradeoffs human cognition makes? Can descriptions of neural behavior be constrained by theories of human inference processes?

7:50-8:00	Introductory Remarks
8:00-8:30	**TBA** STUART RUSSELL
8:30-9:00	**TBA** NOAH GOODMAN
9:00-9:30	Coffee Break
9:30-10:00	**TBA** PAUL SCHRATER
10:00-10:30	Discussion
10:30-11:00	Poster spotlights
11:00-1:00	Posters
4:00-4:30	**TBA** ED VUL
4:30-5:00	**TBA** MATT BOTVINIK

5:00-5:30	Coffee Break
5:30-6:00	**TBA** Jerry Zhu
6:00-6:30	**TBA** Tom Griffiths
6:30-	Discussion

December 11, 2009, 07:30-10:30 and 15:30-18:30 Hilton: Mt. Currie South **WS9**

Clustering: Science or art? Towards principled approaches

http://clusteringtheory.org

Margaret Ackerman ritasemail@gmail.com
University of Waterloo
Shai Ben-David shai@cs.uwaterloo.ca
University of Waterloo
Avrim Blum avrim@cs.cmu.edu
Carnegie Mellon University
Isabelle Guyon guyon@clopinet.com
Clopinet
Ulrike von Luxburg ulrike.luxburg@tuebingen.mpg.de
Max Plank Institute for Biological Cybernetic
Robert Williamson Bob.Williamson@anu.edu.au
Australian National University
Reza Bsoagh Zadeh rezab@cmu.edu
Carnegie Mellon University

Abstract

This workshop aims at initiating a dialog between theoreticians and practitioners, aiming to bridge the theory-practice gap in this area. The workshop will be built along three main question: FROM THEORY TO PRACTICE: Which abstract theoretical characterizations / properties / statements about clustering algorithms exist that can be helpful for practitioners and should be adopted in practice? FROM PRACTICE TO THEORY: What concrete questions would practitioners like to see addressed by theoreticians? Can we identify de-facto practices in clustering in need of theoretical grounding? Which obscure (but seemingly needed or useful) practices are in need of rationalization? FROM ART TO SCIENCE: In contrast to supervised learning, where there is general consensus on how to assess the quality of an algorithm, the frameworks for analyzing clustering are only beginning to be developed and clustering is still largely an art. How can we progress towards a deeper understanding of the space of clustering problems and objectives, including the introduction of falsifiable hypotheses and properly designed experimentation? How could one set up a clustering challenge to compare different clustering algorithms? What could be scientific standards to evaluate a clustering algorithm in a paper? The workshop will also serve as a follow up meeting to the NIPS 2005 *Theoretical Foundations of clustering* workshop, a venue for the different research groups working on these issues to take stock, exchange view points and discuss the next challenges in this ambitious quest for theoretical foundations of clustering.

7:30 - 8:15	**Introduction - Presentations of different views on clustering** Shai Ben-David, Avrim Blum, Ulrike von Luxburg and Bob Williamson
	Game Theory and Physics perspectives
8:15 - 8:45	**What is a cluster: Perspectives from game theory** Marcello Pelillo
8:45 - 9:15	**Clustering with prior information** Armen E. Allahverdyan, Aram Galstyan, Greg Ver Steeg
9:15 - 9:30	Coffee break
	Evaluating clustering: the human factor and particular applications

9:30 - 10:00	**Finding a better k: a psychophysical investigation of clustering** Joshua Lewis
10:00 - 10:15	**Single data, multiple clusterings** Sajib Dasgupta and Vincent Ng
10:15 - 10:30	**Empiricial study of cluster evaluation metrics** Nima Aghaeepour, Alireza Hadj Khodabakhshi, and Ryan R. Brinkman
10:30 - 11:00	**Invited talk: Clustering applications at Yahoo!** Deepayan Chakrabarti
	Hierarchical Clustering
3:30 - 4:00	**Some ideas for formalizing clustering** Gunnar Carlsson and Facundo Memoli
4:00 - 4:30	**Characterization of linkage based clustering** Margarita Ackerman, Shai Ben-David and David Loker
	Information theoretic approaches
4:30 - 5:00	**Information theoretic model selection in clustering** Joachim Buhmann
5:00 - 5:15	Coffee break
5:15 - 5:45	**PAC-Bayesian approach to formulation of clustering** Yevgeny Seldin and Naftali Tishby
5:45 - 6:30	Panel discussion

December 12, 2009, 07:30-10:30 and 15:30-18:30 Westin: Nordic **WS10**

Connectivity Inference in Neuroimaging

http://cini2009.kyb.tuebingen.mpg.de

Karl Friston k.friston@fil.ion.ucl.ac.uk
University College London
Moritz Grosse-Wentrup moritzgw@tuebingen.mpg.de
Max Planck Institute for Biological Cybernetics
Uta Noppeney uta.noppeney@tuebingen.mpg.de
MPI Tuebingen
Bernhard Schölkopf bs@tuebingen.mpg.de
Max Planck Institute for Biological Cybernetics

Abstract

Over the past decade, brain connectivity has become a central theme in the neuroimaging community. At the same time, causal inference has recently emerged as a major research topic in machine learning. Even though the two research questions are closely related, interactions between the neuroimaging and machine-learning communities have been limited.

The aim of this workshop is to initiate productive interactions between neuroimaging and machine learning by introducing the workshop audience to the different concepts of connectivity/causal inference employed in each of the communities. Special emphasis is placed on discussing commonalities as well as distinctions between various approaches in the context of neuroimaging. Due to the increasing relevance of brain connectivity for analyzing mental states, we also highly welcome contributions discussing applications of brain connectivity measures to real-world problems such as brain-computer interfacing or mental state monitoring.

7:30-7:45	**Opening remarks** Moritz Grosse-Wentrup & Uta Noppeney, MPI for biological Cybernetics
7:45-8:15	**Keynote address: Stochastic Dynamic Causal Modelling** Jean Daunizeau, University of Zurich & University College London
08:15-8:45	**Efficient sequential inference in DBNs: Steps towards joint MEG/fMRI connectivity analysis** Sergey M. Plis, University of New Mexico
8:45-9:15	**Full stochastic differential models for fMRI, with efficient particle smoothing for state and connectivity estimation** Amos Storkey, University of Edinburgh
9:15 - 9:30	Coffee break
9:30-10:00	**Causal neural cascades during cognitive tasks** Joseph D. Ramsey, Carnegie Mellon University
10:00 - 03:00	Poster session and recreational activities
3:00-3:30	**The detection of partially directed functional networks from fMRI meta-analysis data** Jane Neumann, MPI for Human Cognitive and Brain Sciences

3:30 - 4:00	**Keynote address – The identification of causal networks from fMRI data: possibilities, limitations and subtle aspects** Alard Roebroeck, Maastricht University
4:00-4:30	**Identifying quasi-neural level task related connectivity in simultaneous EEG/fMRI a single non-stationary dynamic system** Jason F. Smith, National Institutes of Health
4:30 - 5:00	**Keynote address – Imaging human agency: Mobile brain/body imaging of cooperative dynamics** Scott Makeig, University of California, San Diego
5:00 - 5:15	Coffee break
5:15-5:45	**Supervised tract segmentation with diffusion and functional fMRI data** Emanuele Olivetti, Fondazione Bruno Kessler & University of Trento
5:45 - 6:30	Plenary discussion
6:30 -	Continuing poster session

POSTERS

A Bayesian approach for inferring neuronal connectivity from calcium fluorescent imaging data
Joshua Vogelstein, Johns Hopkins University

A functional geometry of fMRI BOLD signal interactions
Georg Langs, Massachusetts Institute of Technology

Changes in functional interactions during anaesthesia-induced loss of consciousness
Jessica Schrouff, University of Liege

Comparison of condition-specific functional connectivity networks
Svetlana V. Shinkareva, University of South Carolina

Detecting functional connectivity in networks of phase-coupled neural oscillators
Charles F. Cadieu, University of California, Berkeley

Difference-based causal models: bridging the gap between Granger causality and DCMs
Mark Voortman, University of Pittsburgh

Hierarchical mixture of classification experts uncovers interactions between brain regions
Bangpeng Yao, Stanford University

Latent causal modelling of neuroimaging data
Morten Morup, Technical University of Denmark

Learning brain fMRI structure through sparseness and local constancy
Jean Honorio, Stony Brook University

Modeling (sparsely) connected sources of the EEG
Stefan Haufe, Fraunhofer FIRST

Multivariate dynamical systems method for estimating causal interactions in fMRI data
Srikanth Ryali, Stanford University

Neuronal systems involved in letter search in high-frequency and low-frequency words
Jagath C. Rajapakse, Nanyang Technological University

Predictive network models of schizophrenia
Irina Rish, IBM T.J. Watson Research Center

Quantifying neuronal synchrony using copulas
Satish G. Iyengar, Syracuse University

December 11, 2009, 07:30-10:30 and 15:30-18:30 Hilton: Black Tusk **WS11**

The Curse of Dimensionality Problem: How Can the Brain Solve It?

http://cnl.salk.edu/~terry/NIPS-Workshop/2009

Simon Haykin haykin@mcmaster.ca
McMaster University
Terrence Sejnowski terry@salk.edu
Salk Institute
Steven Zucker steven.zucker@yale.edu
Yale University

Abstract

The notion of "Curse of Dimensionality" was coined by Richard Bellman (1961). It refers to the exponential increase in computing a task of interest when extra dimensions are added to an associated mathematical space. For example, it arises in solving dynamic programming and optimal control problems when the dimension of the state vector is large. It also arises in solving learning problems when a finite number of data samples is used to learn a "state of nature, the distribution of which is infinitely large." Much has been written on the curse of dimensionality problem in the mathematics and engineering literature. In contrast, little is known on how the human brain solves problems of this kind with relative ease. The key question is: How does the brain do it? To address this basic problem, it may be that we can learn from the mathematics and engineering literature, reformulated in the context of neuroscience.

7:30 **Tutorial: Scaling principles and brain architecture**
Terry Sejnowski (Salk Institute)

7:45 **Tutorial: The Curse of Dimensionality and How to mitigate it in Dynamic Programming Applications**
Simon Haykin (McMaster University)

8:15 Break

8:30 **Did Temporal Difference Reinforcement Learning of Games Break the Curse of Dimensionality?**
Gerry Tesauro (IBM Yorktown Heights)

9:15 Break

9:30 **How the Brain Deals with the Computational Complexity of Vision: A Different Kind of Dimensionality Curse**
John Tsotsos (York University)

4.00 **Predictive Information Bottleneck: Why Simple Organisms Can Cope with Complex Environments**
Naftali Tishby (Weizmann Institute)

4:45 Break

5:00 **Experience-Induced Neural Circuits That Achieve High Capacity**
Les Valiant (Harvard University)

5:45 Break

6:00 **Hyper-dimensional Computing: Computing in Distributed Representation with High-dimensional Random Vectors**
Pentti Kanerva (Stanford University)

December 12, 2009, 07:30-10:30 and 15:30-18:30 Hilton: Cheakamus **WS12**

Deep Learning for Speech Recognition and Related Applications

http://research.microsoft.com/en-us/um/people/dongyu/NIPS2009

Li Deng deng@microsoft.com
Microsoft Corporation
Dong Yu dongyu@microsoft.com
Microsoft Research
Geoffrey Hinton hinton@cs.toronto.edu
University of Toronto

Abstract

Over the past 25 years or so, speech recognition technology has been dominated by a "shallow" architecture — hidden Markov models (HMMs). Significant technological success has been achieved using complex and carefully engineered variants of HMMs. The next generation of the technology requires solutions to remaining technical challenges under diversified deployment environments. These challenges, not adequately addressed in the past, arise from the many types of variability present in the speech generation process. Overcoming these challenges is likely to require "deep" architectures with efficient learning algorithms.

For speech recognition and related sequential pattern recognition applications, some attempts have been made in the past to develop computational architectures that are "deeper" than conventional HMMs, such as hierarchical HMMs, hierarchical point-process models, hidden dynamic models, and multi-level detection-based architectures, etc. While positive recognition results have been reported, there has been a conspicuous lack of systematic learning techniques and theoretical guidance to facilitate the development of these deep architectures. Further, there has been virtually no effective communication between machine learning researchers and speech recognition researchers who are both advocating the use of deep architecture and learning. One goal of the proposed workshop is to bring together these two groups of researchers to review the progress in both fields and to identify promising and synergistic research directions for potential future cross-fertilization and collaboration.

7:30 - 9:00	**Recent developments in deep learning: Architectures and Algorithms** Geoff Hinton
9:00 - 9:30	Cofee break and posters
9:30 - 11:00	**Overview of speech processing using deep architectures beyond HMMs** Li Deng and Dong Yu
11:00 - 3:30	Lunch break
3:30 - 5:00	Panel position talks
5:00 - 5:30	Coffee break and posters
5:30 - 6:30	Panel discussion

Recent developments in deep learning: Architectures and Algorithms
Geoffrey Hinton, University of Toronto
I will start by explaining how deep belief nets can be learned one layer at a time without using any label information. I will then present evidence that this type of "pre-training" creates excellent features for the hidden layers of deep, feedforward neural networks that are then fine-tuned with backpropagation. The

pre-training greatly reduces overfitting especially when additional, unlabeled data is available. It also makes the optimization much easier. I will then describe several different types of units that can be used in deep belief nets and several different learning algorithms that can be used for the pre-training and fine-tuning. Finally I will briefly describe a variety of applications, including phone recognition, in which deep belief nets have outperformed the other methods.

Overview of speech processing using deep architectures beyond HMMs
Li Deng, Microsoft Research
Dong Yu, Microsoft Research
The current dominant technology in speech recognition is based on the hidden Markov model (HMM), a shallow, two-layer architecture that has been carefully engineered over nearly 30 years, with the performance nevertheless far lower than human speech recognition. Researchers have recognized fundamental limitations of such an architecture, and have made a multitude of attempts to develop "deeper" computational architectures for acoustic models in speech recognition aimed to overcome the limitations. These research efforts have been largely isolated in the past, and in this overview talk, we intend to provide a fresh look that this rich body of work and analyze them within a common machine learning framework. The topics to be covered include: 1) multi-level, detection-based framework; 2) Structured speech models (super-segmental or hidden dynamic models); 3) tandem neural network architecture; 4) layered neural network architecture; 5) hierarchical conditional random field; and 6) deep-structured conditional random field. Based on the analysis of the above "beyond-HMM" architectures, we discuss future directions in speech recognition.

Panel discussion
Yoshua Bengio, University of Montreal
Jeff Bilmes, University of Washington
Li Deng, Microsoft Research, Redmond
Geoffrey Hinton, University of Toronto
Helen Meng, Chinese Univerisity of Hong Kong
Larry Saul, University of California at San Diego
Fei Sha, University of Southern California
Dong Yu, Microsoft Research, Redmond

December 11, 2009, 07:30-10:30 and 15:30-18:30 Hilton: Cheakamus **WS13**

Discrete Optimization in Machine Learning: Submodularity, Polyhedra and Sparsity

http://www.discml.cc

Andreas Krause krausea@caltech.edu
California Institute of Technology
Pradeep Ravikumar pradeep.ravikumar@gmail.com
University of Texas, Austin
Jeff Bilmes bilmes@ee.washington.edu
University of Washington

Abstract

Solving optimization problems with ultimately discretely 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.

Format Broadly, this workshop aims at exploring the current challenges in discrete optimization in machine learning. It will explore these topics in tutorials and invited talks. In addition, we will have a poster session with spotlight presentations to provide a platform for presenting new contributions.

Invited Speakers

Alekh Agarwal, University of California, Berkeley
Nina Balcan, Georgia Institute of Technology
Yuri Boykov, University of Western Ontario
Daniel Golovin, California Institute of Technology
Carlos Guestrin, Carnegie Mellon University
Pawan Kumar, Stanford University
Yoram Singer, Hebrew University
David Sontag, Massachusetts Institute of Technology
Ben Taskar, University of Pennsylvania

December 12, 2009, 07:30-10:45 and 15:50-18:45 Westin: Alpine DE **WS14**

The Generative and Discriminative Learning Interface

http://gen-disc2009.wikidot.com/

Simon Lacoste-Julien sl522@cam.ac.uk
University of Cambridge
Percy Liang pliang@cs.berkeley.edu
University of California, Berkeley
Guillaume Bouchard guillaume.Bouchard@xrce.xerox.com
Xerox Research Centre Europe

Abstract

Schedule subject to change - please double check the website for the most up-to-date schedule as well as updated abstracts.

Generative and discriminative learning are two of the major paradigms for solving prediction problems in machine learning, each offering important distinct advantages. They have often been studied in different sub-communities, but over the past decade, there has been increasing interest in trying to understand and leverage the advantages of both approaches. The aim of this workshop is to provide a platform for both theoretical and applied researchers from different communities to discuss the status of our understanding on the interplay between generative and discriminative learning, as well as to identify forward-looking open problems of interest to the NIPS community. Examples of topics of interest to the workshop are as follows:

- Theoretical analysis of generative vs. discriminative learning
- Techniques for combining generative and discriminative approaches
- Successful applications of hybrids
- Empirical comparison of generative vs. discriminative learning
- Inclusion of prior knowledge in discriminative methods (semi-supervised approaches, generalized expectation criteria, posterior regularization, etc.)
- Insights into the role of generative/discriminative interface for deep learning
- Computational issues in discriminatively trained generative models/hybrid models
- Map of possible generative/discriminative approaches and combinations
- Bayesian approaches optimized for predictive performance
- Comparison of model-free and model-based approaches in statistics or reinforcement learning

07:30-07:50 **Opening address: themes of the workshop, terminology, open questions**
Simon Lacoste-Julien, Percy Liang, Guillaume Bouchard

07:50-08:20 **Invited talk: Generative and Discriminative Models in Statistical Parsing**
Michael Collins (MIT)

08:20-08:40 **Generative and Discriminative Latent Variable Grammars**
Slav Petrov (Google Research)

08:40-09:00 **Discriminative and Generative Views of Binary Experiments**
Mark D. Reid, Robert C. Williamson (Australian National University)

09:00-09:30 Coffee Break

09:30-10:00 **Invited talk: Multi-Task Discriminative Estimation for Generative Models and Probabilities**
Tony Jebara (Columbia University)

10:00- **Poster Session** (see below for abstracts)

SKI / DISCUSSION BREAK

15:50-16:20 **Invited talk: Generative and Discriminative Image Models**
John Winn (Microsoft Research Cambridge)

16:20-16:40 **Learning Feature Hierarchies by Learning Deep Generative Models**
Ruslan Salakhutdinov (MIT)

16:40-17:00 **Why does Unsupervised Pre-training Help Deep Discriminant Learning?**
Dumitru Erhan, Yoshua Bengio, Aaron Courville Pierre-Antoine Manzagol, Pascal Vincent (Université de Montréal)

17:00-17:30 Coffee Break

17:30-17:50 **Unsupervised Learning by Discriminating Data from Artificial Noise**
Michael Gutmann, Aapo Hyvärinen (University of Helsinki)

17:50-18:45 **Panel Discussion** - Panelists:

- Michael Collins, MIT
- Dan Klein, UC Berkeley
- Tony Jebara, Columbia University
- Ben Taskar, University of Pennsylvania
- John Winn, Microsoft Research Cambridge

INVITED TALKS

Generative and Discriminative Models in Statistical Parsing
Michael Collins, MIT
Since the earliest work on statistical parsing, a constant theme has been the development of discriminative and generative models with complementary strengths. In this work I'll give a brief history of discriminative and generative models in statistical parsing, focusing on strengths and weaknesses of the various models. I'll start with early work on discriminative history-based models (in particular, the SPATTER parser), moving through early discriminative and generative models based on lexicalized (dependency) representations, through to recent work on conditional-random-field based models. Finally, I'll describe research on semi-supervised approaches that combine discriminative and generative models.

Multi-Task Discriminative Estimation for Generative Models and Probabilities
Tony Jebara, Columbia University
Maximum entropy discrimination is a method for estimating distributions such that they meet classification constraints and perform accurate prediction. These distributions are over parameters of a classifier, for instance, log-linear prediction models or log-likelihood ratios of generative models. Many of the resulting

optimization problems are convex programs and sometimes just simple quadratic programs. In multi-task settings, several discrimination constraints are available from many tasks which potentially produce even better discrimination. This advantage manifests itself if some parameter tying is involved, for instance, via multi-task sparsity assumptions. Using new variational bounds, it is possible to implement the multi-task variants as (sequential) quadratic programs or sequential versions of the independent discrimination problems. In these settings, it is possible to show that multi-task discrimination requires no more than a constant increase in computation over independent single-task discrimination.

Generative and Discriminative Image Models
John Winn, Microsoft Research Cambridge
Creating a good probabilistic model for images is a challenging task, due to the large variability in natural images. For general photographs, an ideal generative model would have to cope with scene layout, occlusion, variability in object appearance, variability in object position and 3D rotation and illumination effects like shading and shadows. The formidable challenges in creating such a model have led many researchers to pursue discriminative models, which instead use image features that are largely invariant to many of these sources of variability. In this talk, I will compare both approaches and describe some strengths and weaknesses of each and suggest some directions in which the best aspects of both can be combined.

CONTRIBUTED TALKS

Generative and Discriminative Latent Variable Grammars
Slav Petrov, Google Research
Latent variable grammars take an observed (coarse) treebank and induce more fine-grained grammar categories, that are better suited for modeling the syntax of natural languages. Estimation can be done in a generative or a discriminative framework, and results in the best published parsing accuracies over a wide range of syntactically divergent languages and domains. In this paper we highlight the commonalities and the differences between the two learning paradigms and speculate that a hybrid approach might outperform either respectively.

Discriminative and Generative Views of Binary Experiments
Mark D. Reid, Australian National University
Robert C. Williamson, Australian National University and NICTA
We consider Binary experiments (supervised learning problems where there are two different labels) and explore formal relationships between two views of them, which we call "generative" and "discriminative". The discriminative perspective involves an expected loss. The generative perspective (in our sense) involves the distances between class-conditional distributions. We extend known results to the class of all proper losses (scoring rules) and all f-divergences as distances between distributions. We also sketch how one can derive the SVM and MMD algorithms from the generative perspective.

Learning Feature Hierarchies by Learning Deep Generative Models
Ruslan Salakhutdinov, MIT
In this paper we present several ideas based on learning deep generative models from high-dimensional, richly structured sensory input. We will exploit the following two key properties: First, we show that deep generative models can be learned efficiently from large amounts of unlabeled data. Second, they can be discriminatively fine-tuned using the standard backpropagation algorithm. Our results reveal that the learned high-level feature representations capture a lot of structure in the unlabeled input data, which is useful for subsequent discriminative tasks, such as classification or regression, even though these tasks are unknown when the deep generative model is being trained.

Why does Unsupervised Pre-training Help Deep Discriminant Learning?
Dumitru Erhan, Université de Montréal
Yoshua Bengio, Université de Montréal
Aaron Courville, Université de Montréal
Pierre-Antoine Manzagol, Université de Montréal
Pascal Vincent, Université de Montréal

Recent research has been devoted to learning algorithms for deep architectures such as Deep Belief Networks and stacks of auto-encoder variants, with impressive results obtained in several areas. The best results obtained on supervised learning tasks involve an unsupervised learning component, usually in an unsupervised pre-training phase, with a generative model. Even though these new algorithms have enabled training deep models fine-tuned with a discriminant criterion, many questions remain as to the nature of this difficult learning problem. The main question investigated here is the following: why does unsupervised pre-training work and why does it work so well? Answering these questions is important if learning in deep architectures is to be further improved. We propose several explanatory hypotheses and test them through extensive simulations. We empirically show the influence of unsupervised pre-training with respect to architecture depth, model capacity, and number of training examples. The experiments confirm and clarify the advantage of unsupervised pre-training. The results suggest that unsupervised pre-training guides the learning towards basins of attraction of minima that are better in terms of the underlying data distribution; the evidence from these results supports an unusual regularization explanation for the effect of pre-training.

Unsupervised Learning by Discriminating Data from Artificial Noise

Michael Gutmann, University of Helsinki
Aapo Hyvärinen, University of Helsinki

We present a new estimation principle for parameterized statistical models. The idea is to train a classifier to discriminate between the observed data and some artificially generated noise, using the model log-density function in a logistic regression function. It can be proven that this leads to a consistent (convergent) estimator of the parameters. In particular, the method is shown to directly work for unnormalized models, i.e.models where the density function does not integrate to unity. The normalization constant (partition function) can be estimated just like any other parameter. We compare the method with other estimation methods that can be used to learn unnormalized models, including score matching, contrastive divergence, and maximum-likelihood where the correct normalization is estimated with importance sampling. The method is then applied to the estimation of two and three-layer models of natural images.

POSTERS

Integrations of Generative and Discriminative Models

Zhuowen Tu, University of California, Los Angeles
Yi Hong, University of California, Los Angeles
Cheng-Yi Liu, University of California, Los Angeles
Jiayan Jiang, University of California, Los Angeles

Discriminative models (DM) are often focusing on the decision boundary and have strong classification power; generative models (GM) try to explain the data and often have good representational capability. There are many possible ways to combine the two types of models to make effective and efficient inference. In this paper, we summarize a few integrations we have developed in the past and outline two new algorithms, all in the spirit of taking the complementariness of GM and DM. These methods include data-driven Markov chain Monte Carlo (using DM as proposals for inference in GM), GM learning via a series of DM, hybrid DM and GM (on appearance and shape prior respectively) for segmentation, information fusion using DM on augmented features with GM, and locally generative and globally discriminative models. We have achieved state-of-the-art results for various tasks by taking different aspects of hybrid approaches.

Inferring Meta-covariates Via an Integrated Generative and Discriminative Model

Keith J. Harris, University of Glasgow
Lisa Hopcroft, University of Glasgow
Mark Girolami, University of Glasgow

This paper develops an alternative method for analysing high dimensional data sets that combines model based clustering and multiclass classification. By averaging the covariates within the clusters obtained from model based clustering, we define "meta-covariates" and use them to build a multinomial probit regression model, thereby selecting clusters of similarly behaving covariates, aiding interpretation. This simultaneous learning task is accomplished by a variational EM algorithm that optimises a joint distribution which rewards good performance at both classification and clustering. We explore the performance of our methodology on a well known leukaemia dataset and use the Gene Ontology to interpret our results.

Naïve Bayes vs. Logistic Regression: An Assessment of the Impact of the Misclassification Cost

Vidit Jain, University of Massachusetts Amherst

Recent advances in the asymptotic characterization of generative and discriminative learning have suggested several ways to develop more effective hybrid models. An application of these suggested approaches to a practical problem domain remains non-trivial, perhaps due to the violation of various underlying assumptions. One common assumption corresponds to the choice of equal misclassification cost or the ability to estimate such cost. Here, we investigate the effect of this misclassification cost on the comparison between nïve Bayes and logistic regression. To assess the utility of this comparison for practical domains, we include a comparison of mean average precision values for our experiments. We present the empirical comparison patterns on the LETOR data set to solicit the support from related theoretical results.

Hybrid model of Conditional Random Field and Support Vector Machine

Qinfeng Shi, Australia National University and NICTA
Mark Reid, Australia National University
Tiberio Caetano, Australia National University and NICTA

It is known that probabilistic models often converge to the true distribution asymptotically (i.e. fisher consistent). However, the consistency is often useless in practice, since in real world it is impossible to fit the models with infinite many data in a finite time. SVM is fisher inconsistent in multiclass and structured label case, however, it does provide a PAC bound on the true error (known as generalization bound). Is there a model that is fisher consistent for classification and has a generalization bound? We use a naive combination of two models by simply weighted summing up the losses of two. It turns out a surprising theoretical result — the hybrid loss could be fisher consistent in some circumstance and it has a PAC-bayes bound on its true error.

Weighting Priors for Hybrid Learning Principles

Jens Keilwagen, IPK Gatersleben
Jan Grau, University of Halle-Wittenberg
Stefan Posch, University of Halle-Wittenberg
Marc Strickert, University of Halle-Wittenberg
Ivo Grosse, University of Halle-Wittenberg

Nevertheless, it is not obvious how the different learning principles are related each other and how this relation can be interpreted. Here, we define a unified generative-discriminative learning principle containing ML, MAP, MCL, MSP, GDT, and PGDT as limiting cases, we discuss its interpretation, and we investigate the utility of this learning principle on four data sets of transcription factor binding sites (TFBSs).

Parameter Estimation in a Hierarchical Model for Species Occupancy

Rebecca A. Hutchinson, Oregon State University
Thomas G. Dietterich, Oregon State University

The recognition of functional binding sites in genomic DNA remains one of the fundamental challenges of genome research. During the last decades, a plethora of different and well-adapted models has been developed, but only little attention has be payed to the development of different and similarly well-adapted learning principles. Only recently it was noticed that discriminative learning principles can be superior over generative ones in diverse bioinformatics applications, too. Here, we propose a generalization of generative and discriminative learning principles containing the maximum likelihood, maximum a-posteriori, maximum conditional likelihood, maximum supervised posterior, generative-discriminative trade-off, and penalized generative-discriminative trade-off learning principles as special cases. We present an interesting interpretation of this learning principle in case of a special class of priors, and we illustrate its efficacy for the recognition of vertebrate transcription factor binding sites.

December 11, 2009, 07:30-10:30 and 15:30-18:30 Hilton: Sutcliffe A **WS15**

Grammar Induction, Representation of Language and Language Learning

http://www.cs.ucl.ac.uk/staff/rmartin/grll09

Alex Clark alexc@cs.rhul.ac.uk
Royal Holloway University of London
Dorota Glowacka d.glowacka@cs.ucl.ac.uk
University College London
John Shawe-Taylor jst@cs.ucl.ac.uk
University College London
Yee Whye Teh yeewhye@gmail.com
Gatsby Computational Neuroscience Unit, UCL
Chris Watkins chrisw@cs.rhul.ac.uk
Royal Holloway University of London

Abstract

Now is the time to revisit some of the fundamental grammar/language learning tasks such as grammar acquisition, language acquisition, language change, and the general problem of automatically inferring generic representations of language structure in a data driven manner.

Though the underlying problems have been known to be computationally intractable for the standard representations of the Chomsky hierarchy, such as regular grammars and context free grammars, progress has been made by modifying or restricting these classes to make them more observable. Generalisations of distributional learning have shown promise in unsupervised learning of linguistic structure using tree based representations, or using non-parametric approaches to inference. More radically, significant advances in this domain have been made by switching to different representations such as the work in Clark, Eyrand & Habrard (2008) that addresses the issue of language acquisition, but has the potential to cross-fertilise a wide range of problems that require data driven representations of language. Such approaches are starting to make inroads into one of the fundamental problems of cognitive science: that of learning complex representations that encode meaning. This adds a further motivation for returning to this topic at this point.

Grammar induction was the subject of an intense study in the early days of Computational Learning Theory, with the theory of query learning largely developing out of this research. More recently the study of new methods of representing language and grammars through complex kernels and probabilistic modelling together with algorithms such as structured output learning has enabled machine learning methods to be applied successfully to a range of language related tasks from simple topic classification through parts of speech tagging to statistical machine translation. These methods typically rely on more fluid structures than those derived from formal grammars and yet are able to compete favourably with classical grammatical approaches that require significant input from domain experts, often in the form of annotated data.

7:30 - 8:15 **Inference for PCFGs and Adaptor Grammars**
Mark Johnson

8:15 - 8:35 **Learning to Disambiguate Natural Language Using World Knowledge**
Antoine Bordes, Nicolas Usunier, Jason Weston, Ronan Collobert

8:35 - 8:55	**Language Modeling with Tree Substitution Grammars** Matt Post, Daniel Gildea
8:55 - 9:15	**A preliminary evaluation of word representations for named-entity recognition** Joseph Turian, Lev Ratinov, Yoshua Bengio, Dan Roth
9:15 - 9:30	Coffee Break
9:30 - 10:30	**Tutorial: Learnable representations for natural language** Alexander Clark
1:30 - 3:30	Ski Break
3:30 - 4:15	**Learning Languages and Rational Kernels** Mehryar Mohri
4:15 - 4:35	**Sparsity in Grammar Induction** Jennifer Gillenwater, Kuzman Ganchev, Joao Graca, Ben Taskar, Fernando Pereira
4:35 - 5:05	Poster Spotlights
5:05 - 6:00	Poster session
6:00 - 6:30	Discussion and Future Directions

Inference for PCFGs and Adaptor Grammars
Mark Johnson, Brown University
This talk describes the procedures we've developed for adaptor grammarinference. Adaptor grammars are a non-parametric extension to PCFGs that can be used to describe a variety of phonological and morphological languagelearning tasks. We start by reviewing an MCMC sampler for ProbabilisticContext-Free Grammars that serves as the basis for adaptor grammar inference,and then explain how samples from a PCFG whose rules depend on the othersampled trees can be used as a proposal distribution in an MCMC procedure forestimating adaptor grammars. Finally we describe several optimizations thatdramatically speed inference of complex adaptor grammars.

Learning to Disambiguate Natural Language Using World Knowledge
Antoine Bordes, Universite Paris
Nicolas Usunier, Universite Paris
Jason Weston, NEC Labs
Ronan Collobert, NEC Labs
We present a general framework and learning algorithm for the task of concept labeling: each word in a given sentence has to be tagged with the unique physical entity (e.g. person, object or location) or abstract concept it refers to. Our method allows both world knowledge and linguistic information to be used during learning and prediction. We show experimentally that we can handle natural language and learn to use world knowledge to resolve ambiguities in language, such as word senses or coreference, without the use of hand-crafted rules or features.

Language Modeling with Tree Substitution Grammars
Matt Post, University of Rochester
Daniel Gildea, University of Rochester
We show that a tree substitution grammar (TSG) induced with a collapsed Gibbs sampler results in lower perplexity on test data than both a standard context-free grammar and other heuristically trained TSGs, suggesting that it is better suited to language modeling. Training a more complicated bilexical parsing model across TSG derivations shows further (though nuanced) improvement. We conduct analysis and point to future areas of research using TSGs as language models.

A preliminary evaluation of word representations for named-entity recognition

Joseph Turian, Universite de Montreal
Lev Ratinov, University of Illinois at Urbana-Champaign
Yoshua Bengio, Universite de Montreal
Dan Roth, University of Illinois at Urbana-Champaign

We use different word representations as word features for a named-entity recognition (NER) system with a linear model. This work is part of a larger empirical survey, evaluating different word representations on different NLP tasks. We evaluate Brown clusters, Collobert and Weston (2008) embeddings, and HLBL (Mnih & Hinton, 2009) embeddings of words. All three representations improve accuracy on NER, with the Brown clusters providing a larger improvement than the two embeddings, and the HLBL embeddings more than the Collobert and Weston (2008) embeddings. We also discuss some of the practical issues in using embeddings as features. Brown clusters are simpler than embeddings because they require less hyperparameter tuning.

Learnable representations for natural language

Alexander Clark,

The Chomsky hierarchy was explicitly intended to represent the hypotheses from distributional learning algorithms; yet these standard representations are well known to be hard to learn, even under quite benign learning paradigms, because of the computationally complexity of inferring rich hidden structures like trees.

There is a lot of interest in unsupervised learning of natural language – current approaches (e.g. Klein and Manning, Johnson's Adaptor Grammars) use modifications of existing models such as tree or dependency structures together with sophisticated statistical models in order to recover structures that are as close as possible to gold standard manual annotations.

This tutorial will cover a different approach: recent algorithms for the unsupervised learning of representations of natural language based on distributional learning (Clark & Eyraud 2007; Clark, Eyraud and Habrard, 2008; Clark 2009). This research direction involves abandoning the standard models and designing new representation classes for formal languages that are richly structured but where the structure is not hidden but based on observable structures of the language – the syntactic monoid or a lattice derived from that monoid. These representation classes are as a result easy to learn.

We will look briefly at algorithms for learning deterministic automata, and then move on to algorithms for learning context free and context sensitive languages. These algorithms explicitly model the distribution of substrings of the language: they are efficient (polynomial update time) and provably correct for a class of languages that includes all regular languages, many context free languages and a few context sensitive languages. This class may be rich enough to represent natural language syntax.

Learning Languages and Rational Kernels

Mehryar Mohri, NYU

This talk will discuss several topics related to learning automata and learning languages with rational kernels.

POSTERS

An Empirical Study of Hierarchical Dirichlet Process Priors for Grammar Induction

Kewei Tu,
Vasant Honavar,

Using PCA for Probabilistic Grammatical Inference on Trees

Raphaël Bailly,
François Denis,

Èdouard Gilbert,
Amaury Habrard,

Building Bilingual Parallel Corpora based on Wikipedia
Mehdi Mohammadi,
Naser QasemAghaee,

Unsupervised Part Speech Tagging Without a Lexicon
Adam R. Teichert,
Hal Daume III,

A Constructionist Approach to Grammar Inference
Oskar Kohonen,
Sami Virpioja,
Krista Lagus,

Unsupervised Morphological Disambiguation using Statistical Language Models
Mehmet Ali Yatbaz,
Deniz Yuret,

Baby Steps: How "Less is More" in Unsupervised Dependency Parsing
Valentin I. Spitkovsky,
Hiyan Alshawi,
Daniel Jurafsky,

Grammatical Inference with Tree Kernels
Armin Buch,

Probabilistic Languag
Dorota Glowacka,
John Shawe-Taylor,

December 12, 2009, 07:30-10:30 and 15:30-18:30 Westin: Alpine A **WS16**

Kernels for Multiple Outputs and Multi-task Learning: Frequentist and Bayesian Points of View

http://intranet.cs.man.ac.uk/mlo/mock09

Mauricio Alvarez alvarezm@cs.man.ac.uk
University of Manchester
Lorenzo Rosasco lrosasco@mit.edu
MIT
Neil D. Lawrence neill@cs.man.ac.uk
University of Manchester

Abstract

TOPIC

Accounting for dependencies between outputs has important applications in several areas. In sensor networks, for example, missing signals from temporal failing sensors may be predicted due to correlations with signals acquired from other sensors. In geostatistics, prediction of the concentration of heavy pollutant metals (for example, Copper concentration), that require expensive procedures to be measured, can be done using inexpensive and oversampled variables (for example, pH data). Within machine learning, this framework is known as multitask learning. Multi-task learning is a general learning framework in which it is assumed that learning multiple tasks simultaneously leads to better modeling results and performance that learning the same tasks individually. Exploiting correlations and dependencies among tasks, it becomes possible to handle common practical situations such as missing data or to increase the amount of potential data when only few amount of data per task is available. Although there are several approaches to multi-task learning out there, in this workshop we focus our attention to methods based on constructing covariance functions (kernels) for multiple outputs, to be employed, for example, together with Gaussian processes or regularization networks.

AIMS

In the last few years there has been an increased amount of work on Multi-taskLearning. From the Bayesian perspective, this problem has been tackled using hierarchical Bayesian together with neural networks. More recently, the Gaussian Processes framework has been considered, where the correlations among tasks can be captured by appropriate choices of covariance functions. Many of these choices have been inspired by the geo-statistics literature, in which a similar area is known as cokriging. From the frequentist perspective, regularization theory has provided a natural framework to deal with multi-task problems: assumptions on the relation of the different tasks translate into the design of suitable regularizers. Despite the common traits of the proposed approaches, so far different communities have worked independently. For example it is natural to ask whether the proposed choices of the covariance function can be interpreted from a regularization perspective. Or, in turn, if each regularizer induces a specific form of the covariance/kernel function. By bringing together the latest advances from both communities, we aim at establishing what is the state of the art and the possible future challenges in the context of multiple-task learning.

07:30 - 07:45 **Introduction**
Organizers

07:45 - 08:25 **Geostatistics for Gaussian processes**
Hans Wackernagel

08:25 - 09:05 **Borrowing strength, learning vector valued functions, andsupervised dimension reduction**
Sayan Mukherjee

09:05 - 09:30	Coffee Break
09:30 - 10:10	**Invited Talk (TBA)** Dave Higdon
10:10 - 10:30	Discussion session
15:30 - 16:10	**Multitask/Multiple Output Kernels from underlying Mechanistic Models** Neil D. Lawrence
16:10-16:50	**Multi-Task Learning and Matrix Regularization** Andreas Argyriou
17:00-17:30	Coffee Break
17:30-18:10	**Learning Vector Fields with Spectral Filtering** Lorenzo Rosasco
18:10-18:30	Final Discussion

Geostatistics for Gaussian processes
Hans Wackernagel,
Gaussian process methodology has inspired a number of stimulating new ideas in the area of machine learning. Kriging has been introduced as a statistical interpolation method for the design of computer experiments twenty years ago. However, some aspects of the geostatistical methodology originally developed for natural resource estimation have been ignored when switching to this new context. This talk reviews concepts of geostatistics and in particular the estimation of components of spatial variation in the context of multiple correlated outputs.

Borrowing strength, learning vector valued functions, and supervised dimension reduction
Sayan Mukherjee,
We study the problem of supervised dimension reduction from the perspective of learning vector valued functions and multi-task or hierarchical modeling in ai regularization framework. An algorithm is specified and empirical results are provided. In the second part of the talk the same problem of supervised dimension reduction for a hierarchical model is revisted from a non-parametric Bayesian perspective.

Invited Talk (TBA)
David Higdon,

Multitask/Multiple Output Kernels from underlying Mechanistic Models
Neil D. Lawrence,
From the Gaussian process perspective a multiple output kernel is a multiple output covariance function. The interpretation given in Gaussian processes is that the kernel function can express correlations between different outputs. This is an excellent approach to dealing with multivariate data. If the data is time series, we can use time as an input and use the multiple output kernel methods to provide us with an approach for jointly modeling the entire multivariate time series. A key question is how to express the cross-kernels (or cross-covariances) between the different outputs.

In this talk we will review the latent force model approach to generating these cross covariances. In essence the idea is to assume that a simple physical system (expressed through differential equations) underlies our data. By assuming that the system is driven by a set of unobserved forces (which are marginalized through Gaussian process priors) we recover multiple output covariance functions which can express a great deal of structure in the data.

Multi-Task Learning and Matrix Regularization
Andreas Argyriou,

Multi-task learning extends the standard paradigm of supervised learning. In multi-task learning, samples for multiple related tasks are given and the goal is to learn a function for each task and also to generalize well (transfer learned knowledge) on new tasks. The applications of this paradigm are numerous and range from computer vision to collaborative filtering to bioinformatics while it also relates to vector valued problems, multiclass, multiview learning etc. I will present a framework for multi-task learning which is based on learning a common kernel for all tasks. I will also show how this formulation connects to the trace norm and group Lasso approaches. Moreover, the proposed optimization problemcan be solved using an alternating minimization algorithm which is simple and efficient. It can also be "kernelized" by virtue of a multi-task representer theorem, which holds for a large family of matrix regularization problems and includes the classical representer theorem as a special case.

Learning Vector Fields with Spectral Filtering
Lorenzo Rosasco,
We present a class of regularized kernel methods for vector valued learning, which are based on filtering the spectrum of the kernel matrix. The considered methods include Tikhonov regularization as a special case, as well as interesting alternatives such as vector valued extensions of L2 boosting. While preserving the good statistical properties of Tikhonov regularization, some of the new algorithms allows for a much faster implementation since they require only matrix vector multiplications. We discuss the computational complexity of the different methods, taking into account the regularization parameter choice step. The results of our analysis are supported by numerical experiments.

December 11, 2009, 07:30-10:30 and 15:30-18:30 Hilton: Mt. Currie North **WS17**

Large-Scale Machine Learning: Parallelism and Massive Datasets

http://www.select.cs.cmu.edu/meetings/biglearn09

Alex Gray agray@cc.gatech.edu
Georgia Tech
Alex Smola alex@smola.org
Yahoo!
Arthur Gretton arthur.gretton@gmail.com
Carnegie Mellon
Joseph Gonzalez jegonzal@cs.cmu.edu
Carnegie Mellon
Carlos Guestrin guestrin@cs.cmu.edu
Carnegie Mellon

Abstract

Physical and economic limitations have forced computer architecture towards parallelism and away from exponential frequency scaling. Meanwhile, increased access to ubiquitous sensing and the web has resulted in an explosion in the size of machine learning tasks. In order to benefit from current and future trends in processor technology we must discover, understand, and exploit the available parallelism in machine learning. This workshop will achieve four key goals: *Bring together people with varying approaches to parallelism in machine learning to identify tools, techniques, and algorithmic ideas which have lead to successful parallel learning. *Invite researchers from related fields, including parallel algorithms, computer architecture, scientific computing, and distributed systems, who will provide new perspectives to the NIPS community on these problems, and may also benefit from future collaborations with the NIPS audience. *Identify the next key challenges and opportunities to parallel learning. *Discuss large-scale applications, e.g., those with real time demands, that might benefit from parallel learning. Prior NIPS workshops have focused on the topic of scaling machine learning, which remains an important developing area. We introduce a new perspective by focusing on how large-scale machine learning algorithms should be informed by future parallel architectures.

7:30 - 7:40	**Brief Introduction** Workshop Chairs
7:40 - 8:05	**Parallel Inference in Large Probabilistic Graphical Models** Viktor Prasanna
8:05 - 8:30	**Parallel Online Learning** John Langford
8:30 - 8:55	**Probabilistic Machine Learning in Computational Advertising** Joaquin Quiñero Candela
8:55 - 9:15	Coffee Break
9:15 - 9:40	**Tuning GPUs for Fast Multipole Methods** Rich Vuduc
9:40 - 10:05	**Scalable Learning in Computer Vision** Adam Coates
10:05 - 10:30	**Hadoop-ML: An Infrastructure for Rapid Implementation of Parallel Reusable Analytics** Amol Ghoting

15:30 - 15:55	**TBD** Andrew Moore (Tentative)
15:55 - 16:20	**1 Billion Instances, 1 Thousand Machines, and 3.5 Hours** Gideon Mann
16:20 - 16:45	**FPGA-Based MapReduce Framework for Machine Learning** Ningyi Xu
16:45 - 17:30	Poster Session & Coffee Break
17:35 - 18:00	**Large-Scale Graph-based Transductive Inference** Jeff Bilmes
18:00 - 18:25	**Splash Belief Propagation: Efficient Parallelization Through Asynchronous Scheduling** Joseph Gonzalez

Parallel exact inference on multi-core processors
Viktor K. Prasanna,
Exact inference in Bayesian networks is a fundamental AI technique that has numerous applications including medical diagnosis, consumer help desk, pattern recognition, credit assessment, data mining, genetics, and others. Inference is NP-hard and in many applications real time performance is required. In this talk we show task and data parallel techniques to achieve scalable performance on general purpose multi-core and heterogeneous multi-core architectures. We develop collaborative schedulers to dynamically map the junction tree tasks leading to highly optimized implementations. We design lock-free structures to reduce thread coordination overheads in scheduling, while balancing the load across the threads. For the Cell BE, we develop a light weight centralized scheduler that coordinates the activities of the synergistic processing elements (SPEs). Our scheduler is further optimized to run on throughput oriented architectures such as SUN Niagara processors. We demonstrate scalable and efficient implementations using Pthreads for a wide class of Bayesian networks with various topologies, clique widths, and number of states of random variables. Our implementations show improved performance compared with OpenMP and complier based optimizations.

Parallel Online Learning
John Langford,
A fundamental limit on the speed of training and prediction is imposed by bandwidth: there is a finite amount of data that a computer can access in a fixed amount of time. Somewhat surprisingly, we can build an online learning algorithm fully capable of hitting this limit. I will discuss approaches for breaking the bandwidth limit, including empirical results.

Probabilistic Machine Learning in Computational Advertising
Joaquin Candela,
In the past years online advertising has grown at least an order of magnitude faster than advertising on all other media. This talk focuses on advertising on search engines, where accurate predictions of the probability that a user clicks on an advertisement crucially benefit all three parties involved: the user, the advertiser, and the search engine. We present a Bayesian probabilistic classification model that has the ability to learn 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 should be larger. We also present a scheme for approximate parallel inference that allows efficient training of the algorithm on a distributed data architecture.

Parallel n-body Solvers: Lessons Learned in the Multicore/Manycore Era
Aparna Chandramowlishwaran,
Aashay Shringarpure,
Ryan Riegel,
Sam Williams,

Lenny Oliker,
Alex Gray,
George Biros,
Richard Vuduc,
Generalized n-body problems (GNPs; Gray & Moore NIPS 2000) constitute an important class of computations in both the physical sciences and in massive-scale data analysis. This talk describes some of the latest results and "lessons-learned" in parallelizing, implementing, and tuning an important example of this class—the fast multipole method—on state-of-the-art multicore- and GPU-based systems. Our lessons include careful data layouts, vectorization, mixed precision, and automated algorithmic and code tuning, and a surprising finding in the debate on performance and energy-efficiency of multicore vs. GPU processors.

Scalable Learning in Computer Vision
Adam Coates,
Honglak Lee,
Andrew Y. Ng,
Computer vision is a challenging application area of machine learning. Recent work has shown that large training sets may yield higher performance in vision tasks like object detection. We overview our work in object detection using a scalable, distributed training system capable of training on more than 100 million examples in just a few hours. We also briefly describe recent work with deep learning algorithms that may allow us to apply these architectures to large datasets as well.

Hadoop-ML: An Infrastructure for the Rapid Implementation of Parallel Reusable Analytics
Amol Ghoting,
Edwin Pednault,
Hadoop is an open-source implementation of Google's Map-Reduce programming model. Over the past few years, it has evolved into a popular platform for parallelization in industry and academia. Furthermore, trends suggest that Hadoop will likely be the analytics platform of choice on forthcoming Cloud-based systems. Unfortunately, implementing parallel machine learning/data mining (ML/DM) algorithms on Hadoop is complex and time consuming. To address this challenge, we present Hadoop-ML, an infrastructure to facilitate the implementation of parallel ML/DM algorithms on Hadoop. Hadoop-ML has been designed to allow for the specification of both task-parallel and data-parallel ML/DM algorithms. Furthermore, it supports the composition of parallel ML/DM algorithms using both serial as well as parallel building blocks – this allows one to write reusable parallel code. The proposed abstraction eases the implementation process by requiring the user to only specify computations and their dependencies, without worrying about scheduling, data management, and communication. As a consequence, the codes are portable in that the user never needs to write Hadoop-specific code. This potentially allows one to leverage future parallelization platforms without rewriting one's code.

1 Billion Instances, 1 Thousand Machines and 3.5 Hours
Gideon Mann,
Ryan McDonald,
Mehryar Mohri,
Nathan Silberman,
Daniel Walker,
Training conditional maximum entropy models on massive data sets requires significant computational resources, but by distributing the computation, training time can be significant reduced. Recent theoretical results have demonstrated conditional maximum entropy models trained by weight mixtures of independently trained models converge at the same rate as traditional distributed schemes, but significantly faster. This efficiency is achieved primarily by reducing network communication costs, a cost not usually considered but actually quite crucial.

FPGA-based MapReduce Framework for Machine Learning
Bo Wang,
Yi Shan,
Jing Yan,

Yu Wang,
Ningyi Xu,
Huangzhong Yang,
Machine learning algorithms are becoming increasingly important in our daily life. However, training on very large scale datasets is usually very slow. FPGA is a reconfigurable platform that can achieve high parallelism and data throughput. Many works have been done on accelerating machine learning algorithms on FPGA. In this paper, we adapt Google's MapReduce model to FPGA by realizing an on-chip MapReduce framework for machine learning algorithms. A processor scheduler is implemented for the maximum computation resource utilization and load balancing. In accordance with the characteristics of many machine learning algorithms, a common data access scheme is carefully designed to maximize data throughput for large scale dataset. This framework hides the task control, synchronization and communication away from designers to shorten development cycles. In a case study of RankBoost acceleration, up to 31.8x speedup is achieved versus CPU-based design, which is comparable with a fully manually designed version. We also discuss the implementations of two other machine learning algorithms, SVM and PageRank, to demonstrate the capability of the framework.

Large-Scale Graph-based Transductive Inference
Amarnag Subramanya,
Jeff Bilmes,
We consider the issue of scalability of graph-based semi-supervised learning (SSL) algorithms. In this context, we propose a fast graph node ordering algorithm that improves parallel spatial locality by being cache cognizant. This approach allows for a linear speedup on a shared-memory parallel machine to be achievable, and thus means that graph-based SSL can scale to very large data sets. We use the above algorithm an a multi-threaded implementation to solve a SSL problem on a 120 million node graph in a reasonable amount of time.

Splash Belief Propagation: Efficient Parallelization Through Asynchronous Scheduling
Joseph Gonzalez,
Yucheng Low,
Carlos Guestrin,
David O'Hallaron,
In this work we focus on approximate parallel inference in loopy graphical models using loopy belief propagation. We demonstrate that the natural, fully synchronous parallelization of belief propagation is highly inefficient. By bounding the achievable parallel performance of loopy belief propagation on chain graphical models we develop a theoretical understanding of the parallel limitations of belief propagation. We then introduce Splash belief propagation, a parallel asynchronous approach which achieves the optimal bounds and demonstrates linear to super-linear scaling on large graphical models. Finally we discuss how these ideas may be generalized to parallel iterative graph algorithms in the context of our new GraphLab framework.

December 12, 2009, 07:30-10:30 and 15:30-18:30 Hilton: Sutcliffe A **WS18**

Learning from Multiple Sources with Applications to Robotics

http://www.dcs.gla.ac.uk/~srogers/lms09/index.htm

Barbara Caputo bcaputo@idiap.ch
Idiap Research Institute
Nicoló Cesa-Bianchi cesa-bianchi@dsi.unimi.it
Università degli Studi di Milan
David R. Hardoon drhardoon@i2r.a-star.edu.sg
Institute for Infocomm Research (I^2R)
Gayle Leen gayle.leen@tkk.fi
Helsinki University of Technology
Francesco Orabona francesco.orabona@idiap.ch
Idiap Research Institute
Jaakko Peltonen jaakko.peltonen@tkk.fi
Helsinki University of Technology
Simon Rogers srogers@dcs.gla.ac.uk
University of Glasgow

Abstract

Learning from multiple sources denotes the problem of jointly learning from a set of (partially) related learning problems / views / tasks. This general concept underlies several subfields receiving increasing interest from the machine learning community, which differ in terms of the assumptions made about the dependency structure between learning problems. In particular, the concept includes topics such as data fusion, transfer learning, multitask learning, multiview learning, and learning under covariate shift. Several approaches for inferring and exploiting complex relationships between data sources have been presented, including both generative and discriminative approaches.

The workshop will provide a unified forum for cutting edge research on learning from multiple sources; the workshop will examine the general concept, theory and methods, and will also examine robotics as a natural application domain for learning from multiple sources. The workshop will address methodological challenges in the different subtopics and further interaction between them. The intended audience is researchers working in fields of multi-modal learning, data fusion, and robotics.

07:30	**Invited talk 1** Chris Williams, University of Edinburgh
08:20	**Regular talk 1** speakers TBA
08:40	**Regular talk 2** speakers TBA
09:00	Coffee break
09:30	**Regular talk 3** speakers TBA
09:50	**Poster spotlights** speakers TBA
10:00	Poster session
10:30	Session break

15:30	**Invited talk 2** Ingmar Posner, University of Oxford
16:20	**Regular talk 4** speakers TBA
16:40	**Regular talk 5** speakers TBA
17:00	Coffee break
17:30	**Regular talk 6** speakers TBA
17:50	**Regular talk 7** speakers TBA
18:10	Discussion and Future Directions

Invited talk 1
Chris Williams, University of Edinburgh
TBA

Invited talk 2
Ingmar Posner, University of Oxford
TBA

December 12, 2009, 07:30-10:30 and 15:30-18:30 Hilton: Diamond Head **WS19**

Learning with Orderings

http://www.select.cs.cmu.edu/meetings/nips09perm

Tiberio Caetano tiberiocaetano@gmail.com
NICTA - Australian National University
Carlos Guestrin guestrin@cs.cmu.edu
Carnegie Mellon University
Jonathan Huang jch1@cs.cmu.edu
Carnegie Mellon University
Guy Lebanon lebanon@cc.gatech.edu
Georgia Tech
Risi Kondor risi@gatsby.ucl.ac.uk
Caltech
Marina Meila mmp@stat.washington.edu
University of Washington

Abstract

Motivation and Goals

Permutations and partial orders as input and output data are ubiquitous. Examples include:

- Preference elicitation
- Social choice and voting theory
- Ranking and search.
- Tracking and identity management
- Structure learning for Bayesian networks
- Multi-way classification and other categorization tasks
- Natural language processing
- Graph matching
- ... and many others!

A general and effective way to handle ordered sets of items is to assign each item a score computed from its features. Scoring effectively maps the items onto the real line or another Euclidean space, where standard learning algorithms and other operations apply. While this approach has often worked, we believe that much can be gained by directly building statistical models and learning algorithms on the discrete combinatorial spaces themselves. It is to forward this direction of research that we propose this workshop.

We propose to accomplish:

- Compact representations: What parametrizations for permutations and partial orders exist and what are their respective advantages? What are 'good' parametrizations for subclasses of partial orders that are common in practice (e.g posets of bounded height, ratings, top-t orderings)?

- Algorithms: Poset classes are rich classes (superexponential) which do not naturally admit low dimensional embeddings. Most estimation algorithms over the space of permutation have theoretically exponential or larger running times; other classes of algorithms are polynomial, but of high degree (see e.g. group theoretical algorithms). What structure can one impose on posets in order to achieve compact representations that can be learned efficiently? What are the most effective approximate and heuristic algorithms?
- Learning theory for discrete parameters: Statistical models on permutations and posets have both continuous and discrete parameters, like the central permutation in consensus ranking. How can one formulate and obtain standard theoretical guarantees of learnability: like generalization bounds, consistency, confidence intervals in such combined discrete-continuous spaces?
- Dissemination: Inform the larger NIPS audience on the progress and possibilities in learning with orderings: what data can be advantageously modeled as posets? What algorithmic solutions exist? What software implementations are available?

7:30 - 7:55	**Ranking in the algebra of the symmetric group** Risi Kondor
7:55 - 8:20	**Spectral Analysis for Partially Ranked Data** Dan Rockmore
8:20 - 8:45	**Clustering ranked preference data using sociodemographic covariates** Brendan Murphy
8:45 - 9:10	**TBA** Marina Meila
9:10 - 9:30	Coffee Break
9:30 - 9:55	**Discovering and exploiting riffled independence relations in ranked data** Jonathan Huang
9:55 - 10:45	**Projection Pursuit for Discrete Data** Persi Diaconis
15:30 - 15:35	**Stable Identification of Cliques with Restricted Sensing** Xiaoye Jiang
15:55 - 16:20	**Content Modeling Using Latent Permutations** Regina Barzilay/Harr Chen
16:20 - 16:45	**Learning permutations with exponential weights** Manfred Warmuth
16:45 - 17:30	Coffee Break
17:30 - 18:30	Poster/Demonstration session

Ranking in the algebra of the symmetric group

Risi Kondor,
Marconi Soares Barbosa,

Ranking is hard because implicitly it involves manipulating n!-dimensional vectors. We show that if each training example only involves k out of the n objects to be ranked, then by Fourier analysis on the ranking vectors we can reduce the dimensionality of the problem to $O(n^{2k})$. Moreover, with respect to a natural class of kernels on permutations the inner product between two ranking vectors can be computed in $O((2k)^{2k+2})$ time. We demonstrate these results by experiments using "SnOB" on real-world data.

Spectral Analysis for Partially Ranked Data
Dan Rockmore,
Martin Malandro,
In this talk we present some recent work on the use of the rook monoid, an inverse semigroup, for the as a framework for analyzing partially ranked data. Algorithmic aspects are discussed.

Clustering ranked preference data using sociodemographic covariates
Brendan Murphy,
Claire Gormley,
Ranked preference data arise when a set of judges rank, in order of their preference, some or all of a set of objects. Such data arise in a wide range of contexts: in preferential voting systems, in market research surveys and in university application procedures. Modelling preference data in an appropriate manner is imperative when examining the behaviour of the set of judges who gave rise to the data. Additionally, it is often the case that covariate data associated with the set of judges is recorded when a survey of their preferences is taken. Such covariate data should be used in conjunction with preference data when drawing inferences about a set of judges. In order to cluster a population of judges, the population is modelled as a collection of homogeneous groups of judges. The Plackett-Luce (ex- ploded logit) model for rank data is employed to model a judge's ranked preferences within a group. Thus, a mixture of Plackett-Luce models is employed as an appropriate statistical model for the population of judges, where each component in the mixture represents a group of judges with a specific parameterisation of the Plackett-Luce model. Mixture of experts models provide a framework in which covariates are included in mixture models. In these models, covariates are included through the mixing proportions and through the parameters of component densities using generalized linear model theory. A mixture of experts model for preference data is developed by combining a mixture of experts model and a mixture of Plackett-Luce models. Particular attention is given to the manner in which covariates enter the model. Both the mixing proportions and the group specific parameters are potentially dependent on the covariates. Model selection procedures are employed to select both the manner in which covariates enter the model and to select the optimal number of groups within the population. The model parameters are estimated via the EMM algorithm, a hy- brid of the EM and MM algorithms. Illustrative examples are provided through the 1996 Menu Census Survey conducted by the Market Research Corporation of America and through Irish election data where voters rank electoral candidates in order of their preference. Results indicate mixture modelling using covariates is insightful when examining a population of judges who express preferences.

Discovering and exploiting riffled independence relations in ranked data
Jonathan Huang,
Carlos Guestrin,
Representing distributions over permutations can be a daunting task due to the fact that the number of permutations of n objects scales factorially in n. One recent way that has been used to reduce storage complexity has been to exploit probabilistic independence, but full independence assumptions impose strong sparsity constraints on distributions and are unsuitable for modeling rankings. I will discuss a novel class of independence structures, called riffled independence, encompassing a more expressive family of distributions while retaining many of the properties necessary for performing efficient inference and reducing sample complexity. In riffled independence, one draws two permutations independently, then performs the riffle shuffle, common in card games, to combine the two permutations to form a single permutation. In ranking, riffled independence corresponds to ranking disjoint sets of objects independently, then interleaving those rankings. In addition to pointing out ways in which riffled independence assumptions can be leveraged during learning and inference, I will discuss strategies for efficiently finding riffle independent subsets of objects from ranked data and show some examples in real ranked datasets.

Projection Pursuit for Discrete Data
Persi Diaconis,
Projection Pursuit was developed as a graphical method for "looking at" high-dimensional data by Kruskal, Friedman, and Tukey. In joint work with Julia Saltzman, we have develope notions of projection that work for discrete data, such as rankings and partial rankings. I will review the classical methods and illustrate the new and some very recent advances.

Stable Identification of Cliques with Restricted Sensing
Xiaoye Jiang,
We study the identification of common interest groups from low order interactive observations. A new algebraic approach based on the Radon basis pursuit on homogeneous spaces is proposed. We prove that if the common interest groups satisfy the condition that overlaps between different common interest groups are small, then such common interest groups can be recovered in a robust way by solving a linear programming problem. We demonstrate the applicability of our approach with examples in identifying social communities in the social network of Les Miserables and in inferring the most popular top 5-jokes in the Jester dataset.

Content Modeling Using Latent Permutations
Harr Chen,
Regina Barzilay,
Ordering plays an important role in natural language processing, arising in contexts as disparate as language modeling, parsing, and discourse. In this talk, we present a Bayesian topic model for learning discourse-level document structure based on latent permutations. Our model leverages insights from discourse theory to constrain topic assignments in a way that reflects the underlying organization of document topics. In particular, both topic selection and ordering are biased to be similar across a collection of related documents. We show that this space of orderings can be effectively represented using the Generalized Mallows Model. Embedding this permutation distribution into a generative model for text gives rise to interesting modeling and inference challenges, particularly in how the permutation space can be explored jointly with other hidden parameters. We apply our method to three complementary discourse-level tasks, cross-document alignment, document segmentation, and information ordering, yielding improved performance in each over previous methods.

Learning Permutations with Exponential Weights
Manfred Warmuth,
David Helmbold,
We give an algorithm for the on-line learning of permutations. The algorithm maintains its uncertainty about the target permutation as a doubly stochastic weight matrix (i.e. first order information), and makes predictions using an efficient method for decomposing the weight matrix into a convex combination of permutations. The weight matrix is updated by multiplying the current matrix entries by exponential factors, and an iterative procedure is needed to restore double stochasticity. Even though the result of this procedure does not have a closed form, a new analysis approach allows us to prove optimal regret bounds (up to small constant factors) for the case when the loss is linear in the doubly stochastic matrix.

We conclude with a discussion of what happens to our method when higher order information about permutations needs to be maintained.

December 11, 2009, 07:30-10:30 and 15:30-18:30 Westin: Alpine DE **WS20**

Machine Learning in Computational Biology

http://www.fml.tuebingen.mpg.de/nipscompbio

Gal Chechik gal.chechik@gmail.com
Google Research
Tomer Hertz thertz@fhcrc.org
Fred Hutchinson Cancer Research Center
William Stafford Noble noble@gs.washington.edu
University of Washington
Yanjun Qi yanjun@nec-labs.com
NEC Labs America
Jean-Philippe Vert Jean-Philippe.Vert@ensmp.fr
Mines ParisTech
Alexander Zien Alexander.Zien@tuebingen.mpg.de
LIFE Biosystems

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 are 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. Furthermore, next generation sequencing technologies are yielding terabyte scale data sets that require novel algorithmic solutions. This workshop focuses on machine learning solutions to emerging problems in computational biology.

7:45-8:10	**Direct maximization of protein identifications from tandem mass spectra** Marina Spivak, Jason Weston, Michael J. MacCoss and William Stafford Noble
8:10-8:35	**Exploiting physico-chemical properties in string kernels** Nora Toussaint, Oliver Kohlbacher and Gunnar Rätsch
8:35-9:00	**A Bayesian method for 3D reconstruction of macromolecular structure using class averages from single particle electron microscopy** Navdeep Jaitly, Marcus Brubaker, John Rubinstein and Ryan Lilien
09:00-09:15	Coffee
9:15-9:40	**vbFRET: a Bayesian approach to single-molecule forster resonance energy transfer analysis** Jonathan Bronson and Chris Wiggins
9:40-10:05	**Leveraging joint test status distribution for an optimal significance testing** Buhm Han, Chun Ye, Ted Choi and Eleazar Eskin

10:05-10:30	**Statistical methods for ultra-deep pyrosequencing of fast evolving viruses** David Knowles and Susan Holmes
3:45-4:10	**A machine learning pipeline for phenotype prediction from genotype data** Giorgio Guzzetta, Giuseppe Jurman and Cesare Furlanello
4:10-4:35	**Association mapping of traits over time using Gaussian processes** Oliver Stegle and Karsten Borgwardt
4:353-5:00	**Learning graphical model structure with sparse Bayesian factor models and process priors** Ricardo Henao and Ole Winther
05:00-05:15	Coffee
5:15-5:40	**Scalable hierarchical multitask learning in sequence biology** Christian Widmer, Jose Leiva, Yasemin Altun and Gunnar Rätsch
5:40-6:05	**Abstraction augmented Markov models** Cornelia Caragea, Adrian Silvescu, Doina Caragea and Vasant Honavar
6:05-6:30	Discussion

December 12, 2009, 07:30-10:30 and 15:30-18:30 Hilton: Mt. Currie South **WS21**

Manifolds, sparsity, and structured models: When can low-dimensional geometry really help?

http://dsp.rice.edu/nips-2009

Richard Baraniuk richb@rice.edu
Rice University
Volkan Cevher volkan@rice.edu
Rice University
Mark Davenport md@rice.edu
Rice University
Michael Wakin mwakin@mines.edu
Colorado School of Mines
Piotr Indyk indyk@mit.edu
MIT
Bruno Olshausen baolshausen@berkeley.edu
UC Berkeley

Abstract

Manifolds, sparsity, and other low-dimensional geometric models have long been studied and exploited in machine learning, signal processing and computer science. For instance, manifold models lie at the heart of a variety of nonlinear dimensionality reduction techniques. Similarly, sparsity has made an impact in the problems of compression, linear regression, subset selection, graphical model learning, and compressive sensing. Moreover, often motivated by evidence that various neural systems are performing sparse coding, sparse representations have been exploited as an efficient and robust method for encoding a variety of natural signals. In all of these cases the key idea is to exploit low-dimensional models to obtain compact representations of the data. The goal of this workshop is to find commonalities and forge connections between these different fields and to examine how we can we exploit low-dimensional geometric models to help solve common problems in machine learning and beyond.

7:30-7:40	Opening Remarks
7:40-8:00	**Talk 1** Volkan Cevher
8:00-8:20	**Talk 2** Mark Davenport
8:20-8:40	**Talk 3** Martin Wainwright
8:40-9:00	**Talk 4** Rob Nowak
9:00-9:30	Coffee Break
9:30-9:40	Poster Spotlights
9:40-10:30	Poster Session
15:40-16:00	**Talk 5** Ken Clarkson

16:00-16:20	**Talk 6** Larry Carin
16:20-16:40	**Talk 7** Mikhail Belkin
16:40-17:00	**Talk 8** Bruno Olshausen
17:00-17:30	Coffee Break
17:30-17:50	**Talk 9** TBA
17:50-18:30	Panel Discussion

December 11, 2009, 07:30-10:30 and 15:30-18:30 Westin: Alpine BC **WS22**

Probabilistic Approaches for Robotics and Control

http://mlg.eng.cam.ac.uk/marc/nipsWS09

Marc Peter Deisenroth mpd37@cam.ac.uk
University of Cambridge
Bert Kappen b.kappen@science.ru.nl
Radeboud University Nijmegen
Emanuel Todorov todorov@cs.washington.edu
University of Washington
Duy Nguyen-Tuong duy@tuebingen.mpg.de
Max Planck Institute for Biological Cybernetics
Carl Edward Rasmussen cer54@cam.ac.uk
University of Cambridge
Jan Peters jan.peters@tuebingen.mpg.de
Max Planck Institute for Biological Cybernetics

Abstract

During the last decade, many areas of Bayesian machine learning have reached a high level of maturity. This has resulted in a variety of theoretically sound and efficient algorithms for learning and inference in the presence of uncertainty. However, in the context of control, robotics, and reinforcement learning, uncertainty has not yet been treated with comparable rigor despite its central role in risk-sensitive control, sensorimotor control, robust control, and cautious control. A consistent treatment of uncertainty is also essential when dealing with stochastic policies, incomplete state information, and exploration strategies.

The use of probabilistic approaches requires (approximate) inference algorithms, where Bayesian machine learning can come into play. Although probabilistic modeling and inference conceptually fit into this context, they are not widespread in robotics, control, and reinforcement learning. Hence, this workshop aims to bring researchers together to discuss the need, the theoretical properties, and the practical implications of probabilistic methods in control, robotics, and reinforcement learning.

07:30–07:40	**Opening** Organizers
07:40–08:00	**Planning under Uncertainty using Distributions over Posteriors** Nicholas Roy
08:00–08:20	**GP-BayesFilters: Gaussian Process Regression for Bayesian Filtering** Dieter Fox
08:20–08:40	**Imitation Learning and Purposeful Prediction: Probabilistic and Non-probabilistic Methods** Drew Bagnell
08:40–09:00	**Probabilistic Control in Human Computer Interaction** Roderick Murray-Smith
09:00–09:30	Coffee Break
09:30–09:50	**Estimating the Sources of Motor Errors** Konrad Körding
09:50–10:30	**Poster Session** Contributed Papers

15:30–15:50	**Linear Bellman Equations: Theory and Applications** Emanuel Todorov
15:50–16:10	**KL Control Theory and Decision Making under Uncertainty** Bert Kappen
16:10–16:30	**Linear Bellman Combination for Simulation of Human Motion¡** Jovan Popovic
16:30–16:50	**Reinforcement Learning in High Dimensional State Spaces: A Path Integral Approach** Evangelos Theodorou
17:00–17:30	Coffee Break
17:30–17:50	**Probabilistic Design: Promises and Prospects** Miroslav Kárný
17:50–18:10	**Approximate Inference Control** Marc Toussaint

Planning under Uncertainty using Distributions over Posteriors
Nicholas Roy, Massachusetts Institute of Technology
Modern control theory has provided a large number of tools for dealing with probabilistic systems. However, most of these tools solve for local policies; there are relatively few tools for solving for complex plans that, for instance, gather information. In contrast, the planning community has provided ways to compute plans that handle complex probabilistic uncertainty, but these often don't work for large or continuous problems. Recently, our group has developed techniques for planners that can efficiently search for complex plans in probabilistic domains by taking advantage of local solutions provided by feedback and open-loop controllers, and predicting a distribution over the posteriors. This approach of planning over distributions of posteriors can incorporate a surprisingly wide variety of sensor models and objective functions. I will show some results in a couple of domains including helicopter flight in GPS-denied environments.

GP-BayesFilters: Gaussian Process Regression for Bayesian Filtering
Dieter Fox, University of Washington
Bayes filters recursively estimate the state of dynamical systems from streams of sensor data. Key components of each Bayes filter are probabilistic prediction and observation models. In robotics, these models are typically based on parametric descriptions of the physical process generating the data. In this talk I will show how non-parametric Gaussian process prediction and observation models can be integrated into different versions of Bayes filters, namely particle filters and extended and unscented Kalman filters. The resulting GP-BayesFilters can have several advantages over standard filters. Most importantly, GP-BayesFilters do not require an accurate, parametric model of the system. Given enough training data, they enable improved tracking accuracy compared to parametric models, and they degrade gracefully with increased model uncertainty. We extend Gaussian Process Latent Variable Models to train GP-BayesFilters from partially or fully unlabeled training data. The techniques are evaluated in the context of visual tracking of a micro blimp and IMU-based tracking of a slotcar.

Imitation Learning and Purposeful Prediction: Probabilistic and Non-probabilistic Methods
Drew Bagnell, Carnegie Mellon University
Programming robot behavior remains a challenging task. While it is often easy to abstractly define or even demonstrate a desired behavior, designing a controller that embodies the same behavior is difficult, time consuming, and ultimately expensive. The machine learning paradigm offers the promise of enabling "programming by demonstration" for developing high-performance robotic systems. Unfortunately, many "behavioral cloning" approaches that utilize the classical tools of supervised learning (e.g. decision trees, neural networks, or support vector machines) do not fit the needs of modern robotic systems. Classical statistics and supervised machine learning exist in a vacuum: predictions made by these algorithms are explicitly assumed to not affect the world in which they operate.

In practice, robotic systems are often built atop sophisticated planning algorithms that efficiently reason far into the future; consequently, ignoring these planning algorithms in lieu of a supervised learning approach often leads to myopic and poor-quality robot performance. While planning algorithms have shown success in many real-world applications ranging from legged locomotion to outdoor unstructured navigation, such algorithms rely on fully specified cost functions that map sensor readings and environment models to quantifiable costs. Such cost functions are usually manually designed and programmed. Recently, our group has developed a set of techniques that learn these functions from human demonstration. These algorithms apply an Inverse Optimal Control (IOC) approach to find a cost function for which planned behavior mimics an expert's demonstration.

I'll discuss these methodologies, both probabilistic and otherwise, for imitation learning. I'll focus on the Principle of **Causal** Maximum Entropy that generalizes the classical Maximum Entropy Principle, widely used in many fields including physics, statistics, and computer vision, to problems of decision making and control. This generalization enables MaxEnt to apply to a new class of problems including Inverse Optimal Control and activity forecasting. This approach further elucidates the intimate connections between probabilistic inference and optimal control.

I'll consider case studies in activity forecasting of drivers and pedestrians as well as the imitation learning of robotic locomotion and rough-terrain navigation. These case-studies highlight key challenges in applying the algorithms in practical settings that utilize state-of-the-art planners and are constrained by efficiency requirements and imperfect expert demonstration.

Probabilistic Control in Human Computer Interaction
Roderick Murray-Smith, University of Glasgow
Continuous interaction with computers can be treated as a control problem subject to various sources of uncertainty. We present examples of interaction based on multiple noisy sensors (capacitive sensing, location- and bearing sensing and EEG), in domains which rely on inference about user intention, and where the use of particle filters can improve performance. We use the "H-metaphor" for automated, flexibly handover of level of autonomy in control, as a function of the certainty of control actions from the user, in an analogous fashion to 'loosening the reins' when horse-riding. Integration of the inference mechanisms with probabilistic feedback designs can have a significant effect on behaviour, and some examples are presented. (Joint work with John Williamson, Simon Rogers and Steven Strachan).

Estimating the Sources of Motor Errors
Konrad Körding, Northwestern University
Motor adaptation is usually defined as the process by which our nervous system produces accurate movements while the properties of our bodies and our environment continuously change. Many experimental and theoretical studies have characterized this process by assuming that the nervous system uses internal models to compensate for motor errors. Here we extend these approaches and construct a probabilistic model that not only compensates for motor errors but estimates the sources of these errors. These estimates dictate how the nervous system should generalize. For example, estimated changes of limb properties will affect movements across the workspace but not movements with the other limb. We extend previous studies in that area to account for temporal and context effects. This extended model explains aspects of savings along with aspects of generalization.

Linear Bellman Equations: Theory and Applications
Emanuel Todorov, University of Washington
I will provide a brief overview of a class stochastic optimal control problems recently developed by our group as well as by Bert Kappen's group. This problem class is quite general and yet has a number of unique properties, including linearity of the exponentially-transformed (Hamilton-Jacobi) Bellman equation, duality with Bayesian inference, convexity of the inverse optimal control problem, compositionality of optimal control laws, path-integral representation of the exponentially-transformed value function. I will then focus on function approximation methods that exploit the linearity of the Bellman equation, and illustrate how such methods scale to high-dimensional continuous dynamical systems. Computing the weights for a fixed set of basis functions can be done very efficiently by solving a large but sparse linear problem. This enables us to work with hundreds of millions of (localized) bases. Still, the volume of a high-dimensional state space

is too large to be filled with localized bases, forcing us to consider adaptive methods for positioning and shaping those bases. Several such methods will be compared.

KL control theory and decision making under uncertainty
Bert Kappen, Radeboud University Nijmegen
KL control theory consists of a class of control problems for which the control computation can be solved as a graphical model inference problem. In this talk, we show how to apply this theory in the context of a delayed choice task and for collaborating agents. We first introduce the KL control framework. Then we show that in a delayed reward task when the future is uncertain it is optimal to delay the timing of your decision. We show preliminary results on human subjects that confirm this prediction. Subsequently, we discuss two player games, such as the stag-hunt game, where collaboration can improve or worsten as a result of recursive reasoning about the opponents actions. The Nash equilibria appear as local minima of the optimal cost to go, but may disappear when monetary gain decreases. This behaviour is in agreement with experimental findings in humans.

Linear Bellman Combination for Simulation of Human Motion
Jovan Popovic, University of Washington
Simulation of natural human motion is challenging because the relevant system dynamics is high-dimensional, underactuated—no direct control over global position and orientation—and non-smooth—frequent and intermittent ground contacts. In order to succeed, control policy must look ahead to determine stabilizing actions and it must optimize to generate lifelike motion. In this talk, we will review recently developed control systems that yield high-quality agile movements for three-dimensional human simulations. Creating such controllers requires intensive computer optimization and reveals a need for reusing as many control policies as possible. We will answer this problem partially with an efficient combination that creates a new optimal control policy by reusing a set of optimal controls for related tasks. It remains to be seen if the same approach can also be applied to control systems needed to generate lifelike human motion.

Reinforcement Learning in High Dimensional State Spaces: A Path Integral Approach
Evangelos Theodorou, University of Southern California
With the goal to generate more scalable algorithms with higher efficiency and fewer open parameters, reinforcement learning (RL) has recently moved towards combining classical techniques from optimal control and dynamic programming with modern learning techniques from statistical estimation theory. In this vein, this paper suggests to use the framework of stochastic optimal control with path integrals to derive a novel approach to RL with parameterized policies. While solidly grounded in value function estimation and optimal control based on the stochastic Hamilton-Jacobi-Bellman (HJB) equations, policy improvements can be transformed into an approximation problem of a path integral which has no open parameters other than the exploration noise. The resulting algorithm can be conceived of as model-based, semi-model-based, or even model free, depending on how the learning problem is structured. The update equations have no danger of numerical instabilites as neither matrix inversions nor gradient learning rates are required. Our new algorithm demonstrates interesting similarities with previous RL research in the framework of probability matching and provides intuition why the slightly heuristically motivated probability matching approach can actually perform well. Empirical evaluations demonstrate significant performance improvements over gradient-based policy learning and scalability to high-dimensional control problems. Finally, a learning experiment on a simulated 12 degree-of-freedom robot dog illustrates the functionality of our algorithm in a comoplex robot learning scenario. We believe that Policy Improvement with Path Integrals or PI^2 offers currently one of the most efficient, numerically robust, and easy to implement algorithms for RL based on trajectory roll-outs.

Probabilistic Design: Promises and Prospects
Miroslav Kárný, Academy of Sciences of the Czech Republic
The Fully Probabilistic Design (FPD) suggests a probabilistic description of the closed control loop behaviour as well as *desired* closed-loop behaviour. The optimal control strategy is selected as the minimiser of the Kullback-Leibler divergence of these distributions. The approach yields: (i) an explicit minimiser with the evaluation reduced to a conceptually feasible solution of integral equations; (ii) a randomised optimal strategy; (iii) a proper subset of FPDs formed via standard Bayesian designs; (iv) uncertain knowledge,

multiple control goals, and optimisation constrains be expressed in the common probabilistic language. It implies: (i) an easier approximation of the dynamic programming counterpart; (ii) the optimal strategy is naturally explorative; (iii) the goals-expressing ideal distribution can be, even recursively, tailored to the observed closed-loop behavior; (iv) an opportunity to automatically harmonise knowledge and goals within a flat cooperation structure of *decentralised* task.

An importance of the last point has been confirmed by a huge amount of societal/industrial problems that cannot be governed in a centralised way. The anticipated decentralised solution based on the FPD may concern either a number of interacting, locally independent elements, which have their local goals, but have to collaborate to reach a common group goal (e.g. cooperative robots, multi-agent systems, etc.); or a set of independent elements with own goals that need to coordinate their activities (e.g. transportation).
The talk will recall the basic properties of FPD and discusses the promises of an exploitation of the FPD potential.

References

[1] T.V. Guy and M. Kárný. Stationary fully probabilistic control design. In *Proc. of the 2nd Int. Conference on Informatics in Control, Automation and Robotics*, pp. 109–112, Barcelona, 2005.

[2] M. Kárný. Towards fully probabilistic control design. *Automatica*, 32(12):–1722, 1996.

[3] M. Kárný and T. V. Guy. Fully probabilistic control design. *Systems & Control Letters*, 55(4):259–265, 2006.

[4] M. Kárný, J. Böhm, T. V. Guy, L. Jirsa, I. Nagy, P. Nedoma, and L. Tesař. *Optimized Bayesian Dynamic Advising: Theory and Algorithms.* Springer, London, 2005.

Acknowledgement This work was partially supported by GA ČR 102/08/0567.

Approximate Inference Control
Marc Toussaint, Technical University of Berlin
Approximate Inference Control (AICO) is a method for solving Stochastic Optimal Control (SOC) problems. The general idea is to think of control as the problem of computing a posterior over trajectories and control signals conditioned on constraints and goals. Since exact inference is infeasible in realistic scenarios, the key for high-speed planning and control algorithms is the choice of approximations. In this talk I will introduce to the general approach, discuss its intimate relations to DDP and the current research on Kalman's duality, and discuss the approximations that we use to get towards real-time planning in high-dimensional robotic systems. I will also mention recent work on using Expectation Propagation and truncated Gaussians for inference under hard constraints and limits as they typically arise in robotics (collision and joint limit constraints).

December 11, 2009, 07:30-10:30 and 16:00-18:30 Westin: Alpine A **WS23**

Normative electrophysiology: explaining cellular properties of neurons from first principles

http://learning.eng.cam.ac.uk/Public/Lengyel/EventNips09

Jean-Pascal Pfister jean-pascal.pfister@eng.cam.ac.uk
University of Cambridge
Máté Lengyel m.lengyel@eng.cam.ac.uk
University of Cambridge

Abstract

Motto: "A neuron would be a most mystifying system if one did not know that brains computed."

"A wing would be a most mystifying structure if one did not know that birds flew." Nearly 50 years ago, Barlow (1961) used this opening sentence in his landmark paper about the organization of the (sensory) nervous system to motivate why the properties of neurons and neural circuits should be studied in a normative way. According to this approach, one should start by thinking about the functions neurons ought to serve, and then derive their properties from the premise that they serve those functions well – rather than amassing all the teeny-weeny details known about those properties and then search for some function that might explain some subset of them. Indeed, both neuronal and synaptic dynamics can be so complex that it is important to have in mind their potential functional roles, otherwise "one will get lost in a mass of irrelevant detail and fail to make the crucial observations" (Barlow 1961).

In the past decades, computational neuroscience has seen a burgeoning of normative approaches. These studies made significant advances in formulating formal theories of optimality, and optimal computations, identifying relevant physical and computational constraints under which those computations need to be implemented, developing analytical methods and numerical algorithms to solve the resulting constrained optimization problems, and relating these solutions to biological substrates. However, only a relatively small fraction of these studies attempted to make specific predictions about, and thus interpret in normative terms, the cellular-level electrophysiological properties of individual neurons or synapses. Small in numbers it may be, the potential impact of this particular line of research cannot be ignored as such theories may provide a way to bridge the gap between the cellular-molecular and the systems-level branches of neuroscience by connecting low-level properties of the nervous system to its high-level functions. Our workshop aims to highlight and discuss recent work in this field.

In this workshop we will discuss three different, though not necessarily unrelated, organizational principles that have been pursued in explaining cellular properties of neurons and synapses and assess their predictive and interpretive power:

- **The redundancy reduction hypothesis**, which was later formulated as the infomax principle, assumes that the amount of transmitted or stored (Shannon) information should be maximized in neural circuits. Modern reincarnations of this idea seek to show that neural adaptation and long-term synaptic plasticity have key properties which are optimal for this function.
- **The Bayesian approach** assumes that neurons need to perform statistical inferences on their inputs for efficient computation, and recent studies show how neural spiking dynamics and short-term synaptic plasticity may contribute to such computations.
- **The constraint-based approach** assumes that the basic biophysical constraints (energy, space, intrinsic noise, etc) profoundly affect signalling in neurons and synapses and derive show how properties of spike generation and synaptic transmission reflect these constraints.

The non-exhaustive list above aptly illustrates that each of these principles can be applied to study both neurons and synapses, and conversely, the same neuronal-synaptic properties may feature in several functional theories. Identifying these overlaps, conflicts, and alternative interpretations in lively discussions and debates will be a central aspect of the workshop.

Since much of the theoretical background in this field has been adopted from information theory, machine learning, and related fields, we expect that not only experimental and computational neuroscientists, but also machine learning researchers will be interested in the general topic and the specific talks.

07:30-07:40	welcome notes
07:40-08:00	**A beginner's guide to constructing neural networks that remember** Máté Lengyel
08:00-08:30	**Optimal change detection in sensory circuits** Sophie Denève
08:30-09:00	**Single neuron implementations of adaptive coding** Adrienne Fairhall
09:00-09:30	coffee break
09:30-10:00	**Predictable irregularities in retinal receptive fields based on normative approaches** Tatyana Sharpee
10:00-10:30	**Randomly connected networks maximize Fisher information at the edge of chaos** Taro Toyoizumi
10:30	break – poster session
16:00-16:30	**Synaptic plasticity from first principles** Jean-Pascal Pfister
16:30-17:00	**Normative neurophysiology from first-principle biophysical constraints** Aldo Faisal
17:00-17:30	coffee break
17:30:18:00	**What determines the shape of neuronal arbors?** Dmitri Chklovskii
18:00-18:30	discussion

A beginner's guide to constructing neural networks that remember
Máté Lengyel, University of Cambridge
One of the classical computational tasks faced by the brain is that of autoassociative memory: memories stored in the past need to be recalled from fragmented and noisy information in the present. I will analyse auto-associative memory from an information theoretic perspective and treat it at the computational level as a problem of Bayesian inference. Unlike most previous approaches to autoassociative memory, this approach is fairly agnostic at the level of representation: it can be applied to cases when memories are represented by firing rates, or spike timings, or a combination of these. Therefore, the resulting theories have the potential to provide general guidelines for constructing optimal autoassociative networks, or to interpret properties of neural circuits in normative terms.

First, I will show how to optimise recall dynamics in a network of neurons. At the level of implementation, we predict how the synaptic plasticity rule that stores memories and the form of network dynamics that recalls them need to be matched to each other. We applied this theory to a case when memories are (at least partially) represented by spike times, and obtained experimental evidence confirming such a match in hippocampal area CA3. Second, I will present recent work about optimising synaptic plasticity for storing memories, treating it initially 'just' as a problem of information maximization. However, the theory points out a fundamental incompatibility between local learning rules for storage and local network dynamics for recall, which implies that the 'formatting' of information is just as relevant as its maximization. These results suggest normative interpretations for heterosynaptic learning rules and for wiring rules in sparsely connected networks.

Optimal change detection in sensory circuits
Sophie Denève, École Normale Supérieure
Models of sensory coding and computation usually consider sensory cells as representing static stimuli in their receptive fields. In particular, this view pervades theories of visual perception where neurons are primarily seen as responding to stereotyped "patterns" which may include a temporal dimension (i.e. spatio-temporal receptive fields) but are otherwise largely independent of stimulus history.

However, all sensory systems, including the visual system, respond more strongly and precisely to dynamic stimuli than to steady ones. For instance, a hallmark of visual receptors is that they adapt so quickly that visual perception requires the constant retinal motion induced by small eye movements. Thus, it is likely that one of the major roles of sensory processing is to detect and signal sudden changes in the sensory environment.

To test this hypothesis, we explored the idea that sensory circuits are tuned to respond as quickly and reliably as possible to sudden transients in their inputs. To this end, we developed a Bayesian model of change detection under the assumption that appearance (or disappearance) of stimuli are unpredictable and cause rapid changes in firing rates of noisy input spike trains. From this "ideal observer" (normative) model, we derived a minimal neural circuit estimating on-line the probability of stimulus appearance. This minimal circuit couples an excitatory synapse exhibiting short term synaptic depression (STP) and an inhibitory synapse with short term facilitation (STF). This mechanism has anatomical correlates in the neocortex, e.g. through Martinotti inhibitory interneurons and feed-forward inhibition.

We next explored the implication of this simple mechanism for sensory coding and adaptation, in particular in early stages of visual processing. A neural structure tuned to detect binary changes (i.e. "ON" and "OFF" transitions) will respond very differently from a system signalling continuous levels of stimulation (i.e. local luminance or motion energy). Assuming a simple firing mechanism corresponding to a decision threshold, we found properties analogous to the bi-phasic temporal receptive fields (tRFs) reported in the retina, LGN and V1. However, the predicted responses to time varying stimuli are much sparser and temporally precise than would be predicted by the tRF alone. Moreover, response gain and tRF shapes adapt to stimulus variance, also as reported experimentally. This invites us to revise current theories of the computational role and mechanism of this form of adaptation.

Our models predicts how biophysical parameters such as time constant of synaptic plasticity should be tuned to the assumed statistics of the stimulus, i.e. its probability of appearance, duration and levels of input noise. We derived on-line learning rules for these parameters based on the expectation maximization algorithm.

Single neuron implementations of adaptive coding
Adrienne Fairhall, University of Washington
Neural systems use strategies on several timescales for efficiently encoding different stimulus components. We review evidence for some of these strategies in several sensory systems. We will show how the biophysics of single neurons leads to dynamics and input/output transformations that can be seen to optimally track stimulus statistics and to implement efficient coding.

Predictable irregularities in retinal receptive fields based on normative approaches
Tatyana Sharpee, Salk Institute
Neural variability is present throughout the nervous system. Yet, in some tasks, motor output variability can be almost exclusively attributed to sensory input variability, despite the many levels of neural circuitry that are involved in sensory-motor transformations. This suggests that mechanisms for noise reduction exist.

Here, we find one such mechanism by studying the properties of retinal cells. In vision, retinal ganglion cells partition visual space into approximately circular regions termed receptive fields (RFs). Average RF shapes are such that they would provide maximal spatial resolution if they were centered on a perfect lattice. However, individual shapes have fine-scale irregularities. Here, we find that irregular RF shapes increase the spatial resolution in the presence of lattice irregularities from 60% to 92% of that possible for a perfect lattice. Optimization of RF boundaries around their fixed center positions reproduced experimental observations neuron-by-neuron. Our results suggest that lattice irregularities determine the shapes of retinal RFs and that similar algorithms can improve the performance of retinal prosthetics where substantial irregularities arise at their interface with neural tissue.

Randomly connected networks maximize Fisher information at the edge of chaos
Taro Toyoizumi, Columbia University
A randomly connected network is known to show a transition from non-chaotic to chaotic behavior as the strengths of connections increase. Although this chaotic state has been argued as the origin of the irregular activity seen in the cortex, its functional significance is largely unknown. In this study, I analytically derived the Fisher information of a recurrently connected network about its external input. I found that the Fisher information is maximized at the edge of chaos where the system is most sensitive to the external input. Moreover, with observation noise, the chaotic state is more informative than the non-chaotic state around the critical point. The analytical expression of the Fisher information provides an intuitive picture of the trade-off between increasing signal and decreasing noise and shows how the input-output nonlinearity influences information coding. The optimal variation in synaptic strengths is predicted based on the input-output nonlinearity of neurons.

Synaptic plasticity from first principles
Jean-Pascal Pfister, University of Cambridge
Far from being static relays, synapses are complex dynamical elements. The effect of a presynaptic spike on the postsynaptic neuron depends on the history of the activity of both pre- and postsynaptic neurons, and thus the efficacy of a synapse undergoes perpetual modification. These changes in efficacy can last from hundreds of milliseconds or minutes (short-term plasticity) to hours or months (long-term plasticity). In order to regulate their efficacy over these different time scales, synapses use more than 1000 different proteins. In the face of this complexity, it seems reasonable to study synaptic plasticity by starting from first principles rather than by modelling every biophysical detail.

In this talk, I will present two normative models of synaptic plasticity: one for long-term plasticity and one for short-term plasticity. The first model considers a synaptic learning rule that maximises, under some constraints, the mutual information between the pre- and postsynaptic spike trains. This type of learning rule is consistent with data about spike timing-dependent plasticity and can also be mapped to the well-known BCM learning rule.

The second model focuses on short-term plasticity and views it in a Bayesian framework. It starts from the commonplace observation that the spiking of a neuron is an incomplete, digital, report of the analog quantity that contains all the critical information, namely its membrane potential. We therefore suggests that a synapse solves the inverse problem of estimating the pre-synaptic membrane potential from the spikes it receives, acting as a recursive filter. I will show that the dynamics of short-term synaptic depression closely resemble those required for optimal filtering.

Normative neurophysiology from first-principle biophysical constraints
Aldo Faisal, Imperial College
Do hard physical limits constrain the structure and function of neural circuits? We studied this problem from first-principle biophysics looking at three fundamental constraints (noise, energy and time) and how the basic properties of a neuron's components set up a trade-off between these. We focus on the action potentials as the fundamental signal used by neurons to transmit information rapidly and reliably to other neurons along neural circuits.

What determines the shape of neuronal arbors?
Dmitri Chklovskii, HHMI Janelia Farm
We have developed a theory of dendritic and axonal shape based on two principles: minimization of the

wiring cost and maximization of the connectivity repertoire. These two principles can be expressed mathematically as an optimization problem. We solved this optimization problem using the methods of statistical physics and found good agreement with experimental measurements. Remaining discrepancies between theory and experiment should point to other factors affecting neuronal shape such as non-linear computations in dendrites.

December 12, 2009, 07:30-10:30 and 15:30-18:30 | Hilton: Sutcliffe B **WS24**

Optimization for Machine Learning

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

Suvrit Sra suvrit.sra@tuebingen.mpg.de
Max Planck Institute for Biological Cybernetics, Tübingen
Sebastian Nowozin sebastian.nowozin@tuebingen.mpg.de
Microsoft Research Cambridge
Stephen Wright swright@cs.wisc.edu
University of Wisconsin-Madison
S.V.N. Vishwanthan vishy@stat.purdue.edu
Purdue University

Abstract

It is fair to say that at the heart of every machine learning algorithm is an optimization problem. It is only recently that this viewpoint has gained significant following. Classical optimization techniques based on convex optimization have occupied center-stage due to their attractive theoretical properties. But, new non-smooth and non-convex problems are being posed by machine learning paradigms such as structured learning and semi-supervised learning. Moreover, machine learning is now very important for real-world problems which often have massive datasets, streaming inputs, and complex models that also pose significant algorithmic and engineering challenges. In summary, machine learning not only provides interesting applications but also challenges the underlying assumptions of most existing optimization algorithms. Therefore, there is a pressing need for optimization "tuned" to the machine learning context. For example, techniques such as non-convex optimization (for semi-supervised learning), combinatorial optimization and relaxations (structured learning), non-smooth optimization (sparsity constraints, L1, Lasso, structure learning), stochastic optimization (massive datasets, noisy data), decomposition techniques (parallel and distributed computation), and online learning (streaming inputs) are relevant in this setting. These techniques naturally draw inspiration from other fields, such as operations research, theoretical computer science, and the optimization community. Motivated by these concerns, we would like to address these issues in the framework of this workshop.

07:30–08:20	**Invited Talk I: Chordal Sparsity in Semidefinite Programming and Machine Learning** Lieven Vandenberghe
08:20–08:40	**A Pathwise Algorithm for Covariance Selection** Alexandre d'Aspremont
08:40–09:00	**Active Set Algorithm for Structured Sparsity-Inducing Norms** Rodolphe Jenatton
09:00–09:25	COFFEE BREAK
09:25–10:15	**Invited Talk II: On Recent Trends in Extremely Large-Scale Convex Optimization** Arkadi Nemirovski
10:15–10:30	Poster Spotlights
10:30–15:30	Long Break; Poster Session Begins
15:30–16:20	**Invited Talk III: TBA** Nathan Srebro

16:20–16:40	**Tree Based Ensemble Models Regularization by Convex Optimization** Bertrand Cornélusse
16:40–17:00	**On the Convergence of the Convex-Concave Procedure** Bharath Sriperumbudur
17:00–17:20	COFFEE BREAK
17:20–17:40	**SINCO - an Efficient Greedy Method for Learning Sparse INverse COvariance Matrix** Irina Rish
17:40–18:00	**Super-Linear Convergence of Dual Augmented Lagrangian Algorithm for Sparse Learning** Ryota Tomioka
18:00–19:00	Poster Session Continues

INVITED TALKS

On recent trends in extremely large-scale convex optimization
Arkadi Nemirovski, Georgia Institute of Technology
In the talk, we focus on algorithms for solving well-structured large-scale convex programs in the case where huge problem's sizes prevent processing it by polynomial time algorithms and thus make computationally cheap first order optimization methods the methods of choice. We overview significant recent progress in utilizing problem's structure within the first order framework, with emphasis on algorithms with dimension-independent (and optimal in the large-scale case) iteration complexity $O(1/\epsilon)$, ϵ being the target accuracy. We then discuss the possibility to further accelerate the first order algorithms by randomization, specifically, by passing from expensive in the extremely large scale case precise deterministic first order oracles to their computationally cheap stochastic counterparts. Applications to be discussed include SVM's, ℓ_1 minimization, testing sensing matrices for "goodness' in the Compressed Sensing context, low-dimensional approximation of high-dimensional samples, and some others.

Chordal sparsity in semidefinite programming and machine learning
Lieven Vandenberghe, University of California, Los Angeles
Chordal graphs play a fundamental role in algorithms for sparse matrix factorization, graphical models, and matrix completion problems. In matrix optimization chordal sparsity patterns can be exploited in fast algorithms for evaluating the logarithmic barrier function of the cone of positive definite matrices with a given sparsity pattern and of the corresponding dual cone. We will give a survey of chordal sparse matrix methods and discuss two applications in more detail: linear optimization with sparse matrix cone constraints, and the approximate solution of dense quadratic programs arising in support vector machine training.

TBA
Nathan Srebro, Toyota Technical Institute, Chicago

CONTRIBUTED WORKS — * indicates talk; otherwise a poster

Large Margin Classification with the Progressive Hedging Algorithm
Boris Defourny,
Louis Wehenkel,

Bandit-Aided Boosting
Robert Busa-Fekete,

Balazs Kegl,

***A Pathwise Algorithm for Covariance Selection**
Vijay Krishnamurthy,
Selin Ahipasaoglu,
Alexandre d'Aspremont,

***Super-Linear Convergence of Dual Augmented Lagrangian Algorithm for Sparse Learning**
Ryota Tomioka,
Taiji Suzuki,
Masashi Sugiyama,

***Active Set Algorithm for Structured Sparsity-Inducing Norms**
Rodolphe Jenatton,
Jean-Yves Audibert,
Francis Bach,

Mixed-Integer Support Vector Machine
Wei Guan,
Alexander Gray,
Sven Leyffer,

***On the convergence of the concave-convex procedure**
Bharath Sriperumbudur,
Gert Lanckriet,

***Tree based ensemble models regularization by convex optimization**
Bertrand Cornalusse,
Pierre Geurts,
Louis Wehenke,

Sampling-based optimization with mixtures
Remi Bardenet,
Balazs Kegl,

Variable Selection and Grouping with Multiple Graph Priors
Marco Signoretto,
Anneleen Daemen,
Carlo Savorgnan,
Johan Suykens,

Feature Selection as a one-player game
Romaric Gaudel,
Michele Sebag,

A D.C. Programming Approach to the Sparse Generalized Eigenvalue Problem
Bharath Sriperumbudur,
David Torres,

Gert Lanckriet,

***SINCO - an Efficient Greedy Method for Learning Sparse INverse COvariance Matrix**
Katya Scheinberg,
Irina Rish,

December 11, 2009, 07:30-10:30 and 15:30-18:45 Westin: Glacier **WS25**

Statistical Machine Learning for Visual Analytics

http://www.cc.gatech.edu/~lebanon/smlva

Guy Lebanon lebanon@cc.gatech.edu
Georgia Institute of Technology
Fei Sha feisha@usc.edu
University of Southern California

Abstract

As the amount and complexity of available information grows, traditional data analysis methods are insufficient. In particular, the data analysis process becomes inherently iterative and interactive: i) users start analysis with a vague modeling assumption (expressed often as a form of domain knowledge) about the data; ii) data are analyzed and the intermediate results are visually presented to the users; iii) users revise modeling assumptions and the process iterates. This process has emerged as a prominent framework in many data analysis application areas including business, homeland security, and health care. This framework, known succinctly as visual analytics, combines visualization, human computer interaction, and statistical data analysis in order to derive insight from massive high dimensional data.

Many statistical learning techniques, for instance, dimensionality reduction for information visualization and navigation, are fundamental tools in visual analytics. Addressing new challenges – being iterative and interactive – has potential to go beyond the limits of traditional techniques. However, to realize its potential, there is a need to develop new theory and methodology that bridges visualization, interaction, and statistical learning.

The purpose of this workshop is to expose the NIPS audience to this new and exciting interdisciplinary area and to foster the creation of a new specialization within the machine learning community: machine learning for visual analytics.

7:30-8:00	**New Directions in Text Visualization** Guy Lebanon
8:00-8:30	**Compressed Counting** Ping Li
8:30-9:00	**Visualization using probabilistic models** Amir Globerson
9:00-9:30	Coffee Break
9:30-10:00	**Three New Ideas in Manifold Learning based on Semidefinite Programming for High-Dimensional Visualization** Alex Gray
10:00-10:30	**Future Challenges and Open Problems in Machine Learning for Visual Analytics** Panel Discussion
15:30-16:00	**Dimension Reduction: Some Tutorial Notes** Chris Burges
16:00-16:30	**Visual Analytics for Networks** Jure Leskovec

16:30-17:00	**Dimension Reduction with Kernel Dependency Test** Fei Sha
17:00-17:20	Coffee Break
17:20-17:50	**Visualization of hyper-graphs multiscale factors in molecular networks** Sayan Mukherjee
17:50-18:20	**Visual Analytics for Audio** Mark Hasegawa-Johnson
18:20-18:45	Poster Session

New Directions in Text Visualization

Guy Lebanon, Georgia Institute of Technology

We will cover the current state-of-the-art in text visualization, existing challenges, and promising new directions.

Compressed Counting

Ping Li, Cornell University

Many dynamic data such as network traffic can be modeled as data streams. Efficiently summarizing streaming data (in real-time) is a critical task, for visualizing the traffic distribution and detecting anomaly events such as DDoS (distributed denial of service) attacks or network failures. Useful summary statistics include the moments (e.g., L2, L1, L0, and Lp) and (Shannon) entropies. Exactly computing the moments (or entropies) is often infeasible because the data are massive, highly frequently updated, and we usually need the answers in real-time. Randomized algorithms could provide approximate answers to moments and entropies efficiently, using a small space, in one pass of the data. One popular algorithm is the Symmetric Stable Random Projections (Piotr Indyk, JACM06; Ping Li, SODA08).

The Lp moment with $p \to 1$ is extremely useful for computing the Shannon entropy of data streams. The previous algorithms, however, did not capture the interesting observation that the first moment of the data (i.e, the sum, L1 moment) should be trivial to compute (i.e., using only one simple counter). Based on the idea of Skewed Stable Random Projections, Compressed Counting is proposed to dramatically improve the estimates of Lp moments near p=1. Indeed, Comrpessed Counting provides a theoretically rigorous algorithm to estimate the Shannon entropy using only a very small storage space.

In terms of the sample complexity, previous algorithms such as Symmetric Stable Random Projections required $O(1/\epsilon^2)$ space, where epsilon is the specified (relative) accuracy. The first estimation algorithm for Compressed Counting achieved a space bound of $O(1/\epsilon)$, which was a very significant improvement. More recently, we provide a new estimation algorithm to achieve a space bound of $O(1/[\log(1/(1-p)-\log(1/\epsilon)])$

Visualization using probabilistic models

Amir Globerson, The Hebrew University

Visualization is often needed as a means of representing the relations between objects. These relations are also captured by the joint probability. I will describe various visualization methods whose goal is to reflect the underlying distributions. I will also discuss information theoretic principles governing the dimensionality of such representations and their use in supervised learning.

Three New Ideas in Manifold Learning based on Semidefinite Programming, for High-Dimensional Visualization

Alex Gray, Georgia Institute of Technology

I will describe three new manifold learning methods representing alternate paradigms for visualizing high-dimensional data, all based on semidefinite programming: 1) Rank-based manifold learning, which can visualize ordinal data, where only rankings on dissimilarities between points are given as input; 2) Isometric separation maps, which can visualize data with class labels, and can be seen as learning the kernel of an SVM such that linear separability is guaranteed; 3) Density-preserving maps, which preserve the densities of the original points rather than their distances to other points. We will focus on the third idea in particular, as it holds the promise of the ability to effectively reduce high-dimensional data to lower dimensionalities than possible with current popular methods, both empirically and theoretically (based on a theorem rooted in Riemmanian geometry stating that reduction to a particular low dimension is not always possible under distance preservation, but is always possible under density (volume) preservation). I will then describe recent computational methods for performing these and other manifold learning methods on massive modern datasets, treating the difficult problems of nearest-neighbor graph construction, nonparametric density estimation, and semidefinite programming in high dimensionalities.

Dimension Reduction with Kernel Dependency Test
Fei Sha, University of Southern California

Visual Analytics for Networks
Jure Leskovec, Stanford University

Dimension Reduction: Some Tutorial Notes
Chris Burges, Microsoft Research
I will give a brief guided tour of some interesting and useful methods for dimension reduction, some of which are not widely known in the machine learning community. I will also point out the underlying mathematical connections between several dimension reduction methods.

Visualization of hyper-graphs and visulaization of multiscale factors in molecular networks
Sayan Mukherjee, Duke University
Two probabilistic models that highlight geometry and visualization are developed.

The first is in the graphical modeling framework and uses configuration of points in Euclidean space to encode hypergraph or simplicial complex models. This allows for a very natural visualization of higher-order interactions in graphs.

The second develops a multi-scale factor modeling framework that allows for the visualization of molecular networks at varying scales. The is both predictive of phenotypic or response variation and the inferred factors offer insight with respect to underlying physical or biological processes. Illustrations of the utility of this method is shown for modeling cancer phenotypes.

Visual Analytics for Audio
Mark Hasegawa-Johnson, University of Illinois, Urbana-Champaign
Audio browsing of audio content is possible at rates of up to twice or three times real-time; more rapid browsing of audio content requires some type of visualization. Useful visualization requires two components: (1) a representation of the channel capacity of the human visual system, and (2) a representation of the information density of the audio signal. The channel capacity of the human visual system is constrained by cognitive rather than optical factors. Optical factors include blurring caused by the finite density of rods and cones, and neuron thermal noise manifested as a non-zero intensity and color JND. Cognitive factors include the human preference for sparse rather than dense encodings, manifested, for example, in the comparative visual saliency of sparse vs. dense texture maps. Information density of an audio signal includes task-independent perceptual factors and task-dependent logical factors. Perceptual factors in audio are similar to cognitive factors in video: sounds whose texture changes suddenly are more salient than those with slowly varying texture. Logical factors include class labels, label sets, and the detection of anomaly. Four partially-implemented prototype-in-progress visualization systems will be described: (1) a timeliner system that uses multiscale FFT to rapidly portray spectrograms at time scales ranging from microseconds to days, (2) a timeliner system that maps audio salience signals into visual salience signals in order to visually emphasize sounds that would be salient if audited, (3) a milliphone (thousand-microphone) browser that maps audio color (lowpass, bandpass, highpass) to visual hue of a thread hanging in space, and (4) a milliphone browser that maps logical audio content (speech, vehicle, percussion) to visual hue, and loudness to value.

December 12, 2009, 07:30-10:30 and 15:30-18:30 Westin: Glacier **WS26**

Temporal Segmentation: Perspectives from Statistics, Machine Learning, and Signal Processing

http://www.harchaoui.eu/zaid/workshops/nips09/index.html

Zaid Harchaoui zaidh@andrew.cmu.edu
Carnegie Mellon University
Stephane Canu scanu@insa-rouen.fr
INSA Rouen
Olivier Cappe cappe@telecom-paristech.fr
TELECOM ParisTech
Arthur Gretton arthur.gretton@googlemail.com
Carnegie Mellon University
Alain Rakotomamonjy alain.rakoto@insa-rouen.fr
University of Rouen
Jean-Philippe Vert jean-philippe.vert@mines-paristech.fr
Institut Curie - Centre de recherche

Abstract

Data with temporal (or sequential) structure arise in several applications, such as speaker diarization, human action segmentation, network intrusion detection, DNA copy number analysis, and neuron activity modelling, to name a few. A particularly recurrent temporal structure in real applications is the so-called change-point model, where the data may be temporally partitioned into a sequence of segments delimited by change-points, such that a single model holds within each segment whereas different models hold accross segments. Change-point problems may be tackled from two points of view, corresponding to the practical problem at hand: retrospective (or "a posteriori"), aka multiple change-point estimation, where the whole signal is taken at once and the goal is to estimate the change-point locations, and online (or sequential), aka quickest detection, where data are observed sequentially and the goal is to quickly detect change-points. The purpose of this workshop is to bring together experts from the statistics, machine learning, signal processing communities, to address a broad range of applications from robotics to neuroscience, to discuss and cross-fertilize ideas, and to define the current challenges in temporal segmentation.

7:30-8:15	**Stochastic Segmentation Models** Tse Leung Lai
8:15-8:30	**contributed talk** TBD
8:30-9:15	**Product Partition Models for Modelling Changing Dependency Structure in Time Series** Kevin Murhpy
9:15-9:35	Discussion and posters
9:35-9:50	**contributed talk** TBD
9:50-10:35	**Hierarchical-Dirichlet-Process-based Hidden Markov Models** Erik Sudderth
10:35-10:55	Posters and discussion

15:30-16:15	**Sequential Change-point Detection** Olympia Hadjiliadis
16:15-16:30	**contributed talk** TBD
16:30-17:15	**Change Detection in Autonomous Systems** Brian Williams
17:15-17:35	Posters and discussion
17:35-17:50	**contributed talk** TBD
17:50-18:35	**Invited speaker** TBD
18:35-18:45	Discussion and wrap-up

Stochastic Segmentation Models
Tse Leung Lai, Stanford University
We propose for the analysis of array-CGH data, a new stochastic segmentation model and an associated estimation procedure that has attractive statistical and computational properties. An important benefit of this Bayesian segmentation model is that it yields explicit formulas for posterior means, which can be used to estimate the signal directly without performing segmentation. Other quantities relating to the posterior distribution that are useful for providing confidence assessments of any given segmentation can also be estimated by using our method. We propose an approximation method whose computation time is linear in sequence length which makes our method practically applicable to the new higher density arrays. Simulation studies and applications to real array-CGH data illustrate the advantages of the proposed approach

Product Partition Models for Modelling Changing Dependency Structure in Time Series
Kevin Murphy, University of British Columbia
We show how to apply the efficient Bayesian changepoint detection techniques of Fearnhead in the multivariate setting. We model the joint density of vector-valued observations using undirected Gaussian graphical models, whose structure we estimate. We show how we can exactly compute the MAP segmentation, as well as how to draw perfect samples from the posterior over segmentations, simultaneously accounting for uncertainty about the number and location of changepoints, as well as uncertainty about the covariance structure. We illustrate the technique by applying it to financial data and to bee tracking data.

Hierarchical-Dirichlet-Process-based Hidden Markov Models
Erik Sudderth, Brown University
We consider the problem of speaker diarization, the problem of segmenting an audio recording of a meeting into temporal segments corresponding to individual speakers. The problem is rendered particularly difficult by the fact that we are not allowed to assume knowledge of the number of people participating in the meeting. To address this problem, we take a Bayesian nonparametric approach to speaker diarization that builds on the hierarchical Dirichlet process hidden Markov model (HDP-HMM) of Teh et al. (2006). Although the basic HDP-HMM tends to over-segment the audio data–creating redundant states and rapidly switching among them–we describe an augmented HDP-HMM that provides effective control over the switching rate. We also show that this augmentation makes it possible to treat emission distributions nonparametrically. To scale the resulting architecture to realistic diarization problems, we develop a sampling algorithm that employs a truncated approximation of the Dirichlet process to jointly resample the full state sequence, greatly improving mixing rates. Working with a benchmark NIST data set, we show that our Bayesian nonparametric architecture yields state-of-the-art speaker diarization results.

Quickest Change Detection
Olympia Hadjiliadis, City University of New York
This work examines the problem of sequential change detection in the constant drift of a Brownian motion

in the case of multiple alternatives. As a performance measure an extended Lordenas criterion is proposed. When the possible drifts, assumed after the change, have the same sign, the CUSUM rule, designed to detect the smallest in absolute value drift, is proven to be the optimum. If the drifts have opposite signs, then a specific 2-CUSUM rule is shown to be asymptotically optimal as the frequency of false alarms tends to infinity.

Change Detection in Autonomous Systems
Brian Williams, MIT

December 12, 2009, 07:30-10:30 and 15:30-18:30 Westin: Callaghan **WS27**

Transfer Learning for Structured Data

http://www.cse.ust.hk/~sinnopan/nips09tlsd/index.html

Sinno Jialin Pan sinnopan@cse.ust.hk
Hong Kong University of Science and Technology
Ivor Tsang IvorTsang@ntu.edu.sg
Nanyang Technological University
Le Song lesong@cs.cmu.edu
Carnegie Mellon University
Karsten Borgwardt karsten.borgwardt@tuebingen.mpg.de
MPIs Tuebingen
Qiang Yang qyang@cse.ust.hk
Hong Kong University of Science and Technology

Abstract

Recently, transfer learning (TL) has gained much popularity as an approach to reduce the training-data calibration effort as well as improve generalization performance of learning tasks. Unlike traditional learning, transfer learning methods make the best use of data from one or more source tasks in order to learn a target task. Many previous works on transfer learning have focused on transferring the knowledge across domains where the data are assumed to be i.i.d. In many real-world applications, such as identifying entities in social networks or classifying Web pages, data are often intrinsically non i.i.d., which present a major challenge to transfer learning. In this workshop, we call for papers on the topic of transfer learning for structured data. Structured data are those that have certain intrinsic structures such as network topology, and present several challenges to knowledge transfer. A first challenge is how to judge the relatedness between tasks and avoid negative transfer. Since data are non i.i.d., standard methods for measuring the distance between data distributions, such as KL divergence, Maximum Mean Discrepancy (MMD) and A-distance, may not be applicable. A second challenge is that the target and source data may be heterogeneous. For example, a source domain is a bioinformatics network, while a target domain may be a network of webpage. In this case, deep transfer or heterogeneous transfer approaches are required. Heterogeneous transfer learning for structured data is a new area of research, which concerns transferring knowledge between different tasks where the data are non-i.i.d. and may be even heterogeneous. This area has emerged as one of the most promising areas in machine learning. In this workshop, we wish to boost the research activities of knowledge transfer across structured data in the machine learning community. We welcome theoretical and applied disseminations that make efforts (1) to expose novel knowledge transfer methodology and frameworks for transfer mining across structured data. (2) to investigate effective (automated, human-machined-cooperated) principles and techniques for acquiring, representing, modeling and engaging transfer learning on structured data in real-world applications. This workshop on Transfer learning for structured data will bring active researchers in artificial intelligence, machine learning and data mining together toward developing methods or systems together, to explore methods for solving real-world problems associated with learning on structured data. The workshop invites researchers interested in transfer learning, statistical relational learning and structured data mining to contribute their recent works on the topic of interest.

Check website for schedule
Speakers listed below

Arthur Gretton, Carnegie Mellon University
Shai Ben-David, University of Waterloo

Neil Lawrence, University of Manchester
Gunnar Raetsch, FML of the Max Planck Societ

December 11, 2009, 07:30-10:30 and 15:30-18:30 Hilton: Sutcliffe B **WS28**

Understanding Multiple Kernel Learning Methods

http://mkl.ucsd.edu/workshop

Gert Lanckriet gert@ece.ucsd.edu
University of California, San Diego
Brian McFee bmcfee@ucsd.edu
University of California, San Diego
Francis Bach francis.bach@ens.fr
INRIA - Ecole Normale Superieure
Nathan Srebro nati@uchicago.edu
TTI-Chicago

Abstract

Multiple kernel learning has been the subject of nearly a decade of research. Designing and integrating kernels has proven to be an appealing approach to address several, challenging real world applications, involving multiple, heterogeneous data sources in computer vision, bioinformatics, audio processing problems, etc.

The goal of this workshop is to step back and evaluate the achievements of multiple kernel learning in the past decade, covering a variety of applications. In short, this workshop seeks to understand where and how kernel learning is relevant (with respect to accuracy, interpretability, feature selection, etc.), rather than exploring the latest optimization techniques and extension formulations. More specifically, the workshop envisions to discuss the following two questions:

1. Kernel learning vs. kernel design: Does kernel learning offer a practical advantage over the manual design of kernels?
2. Given a set of kernels, what is the optimal way, if any, to combine them (sums, products, learned or non learned, with or without cross-validation)?

Check website for schedule

7:30–7:45 **Multiple kernel learning introduction and workshop goals**
Gert Lanckriet

7:45–8:15 **Formulations and basic methods and theory**
Francis Bach and Nathan Srebro

8:15–8:45 **Survey of MKL use**
Corinna Cortes

8:45–9:00 **Preliminary analysis of multiple kernel learning: flat maxima, diversity, and Fisher information**
Theodoros Damoulas, Mark Girolami and Simon Rogers

9:00–9:15 Discussion

9:15–9:50 Coffee break

9:50–10:20 **Designing and combining kernels: some lessons learned from bioinformatics**
Jean-Philippe Vert

10:20–10:30	**Poster spotlights**
	Fold recognition using convex Ccmbinations of multiple kernels Huzefa Rangwala
	Multiple kernel learning on imbalanced data: creating a receptor-ligand classifier Ernesto Iacucci, Shi Yu, Fabian Ojeda and Yves Moreau
	Kernel-based inductive transfer Ulrich Ruckert
10:30–15:30	Break
15:30–16:00	**Multiple kernel learning for feature selection** Manik Varma
16:00–16:30	**Multiple lernel learning approaches for image classification** Peter Gehler
16:30–16:45	Discussion: MKL in vision
16:45–16:55	**Poster spotlights**
	On the algorithmics and applications of a mixed-norm based kernel learning formulation Saketha Nath Jagarlapudi, dinesh govindaraj, Raman S, Chiranjib Bhattacharyya, Aharon Ben-Tal and K. R. Ramakrishnan
	Localized multiple kernel machines for image recognition Mehmet Gonen and Ethem Alpaydin
	Comparison of sparse and nonsparse multiple kernel methods on VOC2009 challenge data Alexander Binder and Motoaki Kawanabe
16:55–17:25	Coffee break
17:25–17:40	**Sparsity-accuracy trade-off in MKL** Ryota Tomioka and Taiji Suzuki
17:40–18:30	Panel Discussion

Fold recognition using convex combinations of multiple kernels
Huzefa Rangwala, George Mason University

Kernel-based inductive transfer
Ulrich Rückert, UC Berkeley

An attention-based approach for learning how to fuse decisions of local experts
Maryam S. Mirian, University of Tehran
Majid Nili Ahmadabadi, University of Tehran
Babak N. Araabi, University of Tehran
Mohammed H. Zokaei A., University of Tehran

On the Algorithmics and Applications of a Mixed-norm based Kernel Learning Formulation

Saketha Nath Jagarlapudi, IIT Bombay
Dinesh Dovindaraj, Indian Institute of Science
Raman S,
Chiranjib Bhattacharyya, Indian Institute of Science
Aharon Ben-Tal,
K. R. Ramakrishnan, Indian Institute of Science

Localized multiple kernel machines for image recognition
Mehmet Gönen, Boğaziçi University
Ethem Alpaydin, Boğaziçi University

Multiple kernel learning on imbalanced data: creating a receptor-ligand classifier
Ernesto Iacucci, ESAT/SCD KULeuven
Shi Yu, ESAT/SCD KULeuven
Fabian Ojeda, ESAT/SCD KULeuven
Yves Moreau, ESAT/SCD KULeuven

Detecting anomalies in multivariate data sets with switching sequences and continuous streams
Santanu Das, NASA Ames Research Center
Bryan Matthews, NASA Ames Research Center
Kanishka Bhaduri, NASA Ames Research Center
Nikunj Oza, NASA Ames Research Center
Ashok Srivastava, NASA Ames Research Center

Multi-kernel gaussian processes
Arman Melkumyan, The University of Sydney
Fabio Ramos, The University of Sydney

Comparison of sparse and nonsparse multiple kernel methods on VOC2009 challenge data
Alexander Binder, Fraunhofer FIRST
Motoaki Kawanabe, Fraunhofer FIRST

Index

Abbeel, Pieter, 33
Ackerman, Margareta, 62
Adams, Ryan Prescott, 58
Agarwal, Alekh, 72
Agarwal, Shivani, 35
Ahipasaoglu, Selin, 111
Ahmadabadi, Majid Nili, 122
Ahmed, Amr, 46, 54
Airoldi, Edoardo, 41
Ali Yatbaz, Mehmet, 81
Alpaydin, Ethem, 123
Alshawi, Hiyan, 81
Alvarez, Mauricio, 82
Amershi, Saleema, 39
Amin, Saurabh, 33
Anderson, Kenneth, 46
Araabi, Babak N., 122
Argyriou, Andreas, 83
Audibert, Jean-Yves, 56, 111
Awasthi, Pranjal, 39

Bach, Francis, 111, 121
Bagnell, Drew, 100
Bagnell, J. Andrew, 40
Bailly, Raphaël, 80
Balasubramanian, Vineeth, 39
Balcan, Maria-Florina, 32
Balcan, Nina, 72
Baldi, Pierre, 45
Baraniuk, Richard, 97
Bardenet, Remi, 111
Bart, Evgeniy, 47
Barzilay, Regina, 94
Basu, Sumit, 38
Bayen, Alexandre, 34
Ben-David, Shai, 62, 119
Ben-Tal, Aharon, 123
Bengio, Yoshua, 70, 75, 80
Bergstrom, Tony, 39
Beygelzimer, Alina, 33
Bhaduri, Kanishka, 123
Bhardwaj, Anurag, 46
Bhattacharyya, Chiranjib, 123
Bilgic, Mustafa, 34
Bilmes, Jeff, 70, 71, 88
Binder, Alexander, 123
Biros, George, 87
Blei, David, 42, 44, 46
Blum, Avrim, 39, 62
Blunsden, Scott, 39
Bordes, Antoine, 79
Borgwardt, Karsten, 119
Bouchard, Guillaume, 73
Boyd-Graber, Jordan, 42
Boykov, Yuri, 72
Broniatowski, David, 45
Bsoagh Zadeh, Reza, 62
Buch, Armin, 81
Bundschus, Markus, 47
Burges, Chris, 35, 115
Busa-Fekete, Robert, 110

Cadieu, Charles F., 65
Caetano, Tiberio, 77, 91
Cai, Chenghui, 57
Cai, Rui, 47
Candela, Joaquin, 86
Canini, Kevin R., 58
Canu, Stephane, 116
Cappe, Olivier, 116
Caputo, Barbara, 89
Carenini, Giuseppe, 46
Carin, Lawrence, 45, 57
Caron, Francois, 51
Castro, Rui, 31
Celikyilmaz, Asli, 46
Cesa-Bianchi, Nicoló, 89
Cevher, Volkan, 97
Chakraborty, Shayok, 39
Chandramowlishwaran, Aparna, 86
Chang, Jonathan, 42
Chang, Ying-Lang, 45
Chayes, Jennifer, 41
Chechik, Gal, 25, 95
Chen, Daozheng, 34
Chen, Harr, 94
Chien, Jen-Tzung, 45, 46
Chklovskii, Dmitri, 107
Chu, Wei, 40
Chuang, Jason, 47
Chueh, Chuang-Hua, 46
Clark, Alex, 78
Clark, Alexander, 80
Coates, Adam, 87
Collins, Michael, 74
Collobert, Ronan, 79
Cornalusse, Bertrand, 111
Courville, Aaron, 75
Crammer, Koby, 35
Cronin, Beau, 55

d'Aspremont, Alexandre, 111

Daemen, Anneleen, 111
Dai, Andrew, 47
Dai, Andrew M., 56
Das, Pradipto, 46
Das, Santanu, 123
Daume III, Hal, 55, 81
Davenport, Mark, 97
de Freitas, Nando, 31
Defourny, Boris, 110
Deisenroth, Marc Peter, 58, 99
Dekel, Ofer, 34
Denève, Sophie, 106
Deng, Li, 69, 70
Denis, François, 80
Dey, Dipak K., 56
Diaconis, Persi, 36, 93
Dietterich, Thomas G., 23, 77
Dietz, Laura, 45
Ding, Bailu, 47
Domeniconi, Carlotta, 57
Domingos, Pedro, 49
Dovindaraj, Dinesh, 123
Doyle, Gabriel, 44
Dunagan, John, 40
Dunson, David, 51

Elkan, Charles, 44
Erhan, Dumitru, 75

Fairhall, Adrienne, 106
Faisal, Aldo, 107
Florina-Balcan, Maria, 39
Fogarty, James, 39
Folz, Peter, 47
Fox, Dieter, 100
Freeman, Bill, 50
Freer, Cameron E., 58
Friston, Karl, 64
Fritz, Doug, 47

Gönen, Mehmet, 123
Gajos, Krzysztof, 39
Garnett, Roman, 33
Gasthaus, Jan, 57
Gaudel, Romaric, 111
Gentile, Claudio, 34
Georgiou, Panayiotis, 45
Gerrish, Sean, 44
Getoor, Lise, 34
Geurts, Pierre, 111
Ghahramani, Zoubin, 51, 52, 59
Ghosh, Joydeep, 39
Ghoshal, Subhashis, 52
Ghoting, Amol, 87
Gilbert, Èdouard, 81
Gildea, Daniel, 79
Girolami, Mark, 45, 76
Globerson, Amir, 48, 114
Glowacka, Dorota, 78, 81
Goebel, Rainer, 21
Golovin, Daniel, 72
Gonzalez, Joseph, 85, 88
Goodman, Noah, 28, 60
Gormley, Claire, 93
Gorur, Dilan, 51
Govindaraju, Venu, 47
Grau, Jan, 77
Gray, Alex, 85, 87, 114
Gray, Alexander, 111
Gretton, Arthur, 85, 116, 119
Griffiths, Thomas L., 53, 58
Griffiths, Tom, 60
Grlder, Steffen, 56
Grosse, Ivo, 77
Grosse-Wentrup, Moritz, 64
Guan, Wei, 111
Guestrin, Carlos, 72, 85, 88, 91, 93
Gutmann, Michael, 76
Guyon, Isabelle, 20, 21, 62

Habrard, Amaury, 81
Hadjiliadis, Olympia, 117
Halloran, Elizabeth, 26
Hanneke, Steve, 34
Harchaoui, Zaid, 116
Hardoon, David R., 17, 89
Harris, Keith J., 76
Hasegawa-Johnson, Mark, 115
Haufe, Stefan, 66
Haykin, Simon, 67
Heller, Katherine, 42, 46
Helmbold, David, 94
Herbrich, Ralf, 37
Herring, Ryan, 33
Hertz, Tomer, 25, 95
Higdon, David, 83
Hill, Jeremy, 18
Hinton, Geoffrey, 45, 50, 69, 70
Hofleitner, Aude, 33
Holloway, America, 47
Honavar, Vasant, 80
Hong, Yi, 76
Honorio, Jean, 65
Hopcroft, Lisa, 76
Hsu, Daniel, 33
Hu, Diane, 44
Huang, Jonathan, 91, 93
Hughes, Lindsey, 45

Hutchinson, Rebecca A., 77
Hutter, Marcus, 27, 28
Hyvärinen, Aapo, 76

Iacucci, Ernesto, 123
Indyk, Piotr, 97
Iyengar, Satish G., 66

Jackson, Matthew, 41
Jacobs, David, 34
Jagarlapudi, Saketha Nath, 123
Jain, Vidit, 77
Jara, Alejandro, 53
Jebara, Tony, 74
Jenatton, Rodolphe, 111
Jiang, Jiayan, 76
Jiang, Xiaoye, 94
Johnson, Mark, 43, 79
Jonas, Eric, 55
Jordan, Michael I., 51, 59
Joshi, Ajay J., 39
Joty, Shafiq, 46
Jun, Goo, 39
Jurafsky, Daniel, 81

Körding, Konrad, 101
Kárný, Miroslav, 102
Kakade, Sham M, 34
Kapoor, Ashish, 38–40
Kappen, Bert, 99, 102
Karahalios, Karrie, 39
Kawanabe, Motoaki, 123
Kegl, Balazs, 111
Keilwagen, Jens, 77
Kim, Samuel, 45
Kireyev, Kirill, 46
Kleinberg, Jon, 41
Knox, W. Bradley, 39
Kohonen, Oskar, 81
Koller, Daphne, 25
Kolter, J. Zico, 23
Kondor, Risi, 91, 92
Krause, Andreas, 32, 34, 71
Kriegel, Hans-Peter, 47
Krishnamurthy, Vijay, 111
Kumar, Pawan, 49, 72
Kumar, Ravi, 41

Lacoste-Julien, Simon, 73
Lagus, Krista, 81
Lanckriet, Gert, 111, 112, 121
Landauer, Thomas, 44
Langford, John, 33, 40, 86
Langs, Georg, 65
Laskey, Kathryn B., 57
Lawrence, Neil, 120
Lawrence, Neil D., 82, 83
Lebanon, Guy, 91, 113, 114
Lee, Honglak, 87
Leen, Gayle, 89
Lengyel, Máté, 104, 105
Leskovec, Jure, 41, 115
Leung Lai, Tse, 117
Lewi, Jeremy, 32
Lewis, Joshua M., 40
Leyffer, Sven, 111
Li, Fei-Fei, 43
Li, Lihong, 40
Li, Ping, 114
Li, Zhiwei, 47
Liang, Percy, 73
Liao, Xuejun, 57
Lim, Lek-Heng, 36
Lima, Pedro U., 33
Linstead, Erik, 45
Liu, Alexander, 39
Liu, Cheng-Yi, 76
Lizotte, Daniel, 32
Lopes, Cristina, 45
Low, Yucheng, 88
Lu, Xun, 21

Madhavan, C.E. Veni, 46
Madrid, Jaisiel, 18
Makino, Takaki, 54
Malandro, Martin, 93
Malgireddy, Manavender, 46
Mann, Gideon, 87
Manning, Chris, 47
Mansinghka, Vikash, 55
Manzagol, Pierre-Antoine, 75
Martinez-Cantin, Ruben, 31
Matthews, Bryan, 123
McAllester, David, 50
McDonald, Ryan, 87
McFarland, Daniel, 47
McFee, Brian, 121
Meila, Marina, 91
Melkumyan, Arman, 123
Meng, Helen, 70
Miao, Xu, 40
Mimno, David, 44
Mirian, Maryam S., 122
Mochihashi, Daichi, 55
Mohammadi, Mehdi, 81
Mohri, Mehryar, 80, 87
Moneta, Alessio, 21
Montesano, Luis, 32

Mooney, Raymond, 45
Moreau, Yves, 123
Morris, Martina, 41
Morup, Morten, 65
Mukherjee, Sayan, 83, 115
Muller, Klaus-Robert, 18
Murphy, Brendan, 93
Murphy, Kevin, 117
Murray, Gabriel, 46
Murray-Smith, Roderick, 101
Murty, M. Narasimha, 46

Narayanan, Shrikanth, 45
Nemirovski, Arkadi, 110
Ng, Andrew Y., 23, 87
Ng, Kee Siong, 27
Ng, Raymond, 46
Nguyen-Tuong, Duy, 99
Noble, William, 25
Noble, William Stafford, 95
Nolte, Guido, 20, 21
Noppeney, Uta, 64
Norets, Andriy, 52
Nowozin, Sebastian, 109

O'Hallaron, David, 88
Ojeda, Fabian, 123
Oliker, Lenny, 87
Olshausen, Bruno, 97
Opper, Manfred, 56
Orabona, Francesco, 89
Orbanz, Peter, 52
Osborne, Michael A., 33
Oza, Nikunj, 123

Pérez-Cruz, Fernando, 17
Palen, Leysia, 46
Pan, Sinno Jialin, 119
Panaccione, Charles, 47
Panchanathan, Sethuraman, 39
Paninski, Liam, 32
Papanikolopoulos, Nikolaos, 39
Parrado-Hernández, Emilio, 17
Pasca, Marius, 46
Pednault, Edwin, 87
Peltonen, Jaakko, 89
Peters, Jan, 99
Petrov, Slav, 75
Petschulat, Cap, 55
Pfister, Jean-Pascal, 104, 107
Polajnar, Tamara, 45
Popescu, Florin, 20, 21
Popovic, Jovan, 102
Porikli, Fatih, 39
Posch, Stefan, 77
Posner, Ingmar, 90
Post, Matt, 79
Poupart, Pascal, 27, 28
Prasanna, Viktor K., 86

QasemAghaee, Naser, 81
Qi, Yanjun, 25, 95

R., Arun, 45
R., Saradha, 45
Rückert, Ulrich, 122
Raetsch, Gunnar, 120
Rai, Piyush, 55
Raj, Bhiksha, 46
Rajapakse, Jagath C., 66
Rakotomamonjy, Alain, 116
Ramage, Daniel, 47
Ramakrishnan, K. R., 123
Ramos, Fabio, 123
Rangwala, Huzefa, 122
Rao, Rajesh P. N., 40
Rasmussen, Carl Edward, 58, 99
Ratinov, Lev, 80
Ravikumar, Pradeep, 71
Reid, Mark, 77
Reid, Mark D., 75
Reisinger, Joseph, 45, 46
Riegel, Ryan, 86
Rish, Irina, 66, 112
Roberts, Stephen J., 33
Rockmore, Dan, 93
Rodriguez, Abel, 53
Roebroeck, Alard, 21
Rogers, Simon, 89
Rosasco, Lorenzo, 82, 84
Rosen, Evan, 47
Ross, Stephane, 40
Roth, Dan, 40, 80
Roy, Daniel, 44
Roy, Daniel M., 58
Roy, Nicholas, 100
Rubin, Timothy, 47
Ryali, Srikanth, 66

S, Raman, 123
Sánchez-Madrid, Jaisiel, 17
Saal, Hannes P., 33
Salakhutdinov, Ruslan, 45, 48, 75
Sato, Issei, 55
Satoh, Shin'ichi, 40
Saul, Larry, 70
Saul, Lawrence, 44
Savorgnan, Carlo, 111

Schölkopf, Bernhard, 64
Schapire, Robert E., 40
Scheinberg, Katya, 112
Schrouff, Jessica, 65
Sebag, Michele, 111
Seeger, Matthias, 33, 34
Sejnowski, Terrence, 67
Sha, Fei, 70, 113, 115
Shafto, Patrick, 55
Shalizi, Cosma, 41
Shan, Yi, 87
Sharan, Roded, 26
Sharpee, Tatyana, 106
Shashanka, Madhusudana, 46
Shawe-Taylor, John, 56, 78, 81
Shi, Qinfeng, 77
Shinkareva, Svetlana V., 65
Shringarpure, Aashay, 86
Signoretto, Marco, 111
Silberman, Nathan, 87
Silverthorn, Bryan, 45
Singer, Yoram, 72
Singh, Satinder, 28
Small, Kevin, 40
Smaragdis, Paris, 46
Smith, Noah, 49
Smola, Alex, 85
Smyth, Padhraic, 47
Soares Barbosa, Marconi, 92
Song, Le, 119
Sontag, David, 44, 48, 72
Spaan, Matthijs T.J., 33
Spitkovsky, Valentin I., 81
Sra, Suvrit, 109
Srebro, Nathan, 110, 121
Sridharan, Karthik, 34
Srihari, Rohini, 46
Srinivas, Niranjan, 34
Sriperumbudur, Bharath, 111
Srivastava, Ashok, 123
Steyvers, Mark, 47
Stone, Peter, 39
Storkey, Amos, 47
Storkey, Amos J., 56
Strickert, Marc, 77
Subramanya, Amarnag, 88
Sudderth, Erik, 117
Sugiyama, Masashi, 111
Sundaram, Shiva, 45
Suresh, V., 45
Suykens, Johan, 111
Suzuki, Taiji, 111

Takagi, Toshihisa, 55
Takei, Shunsuke, 54
Tan, Desney, 39
Taskar, Ben, 49, 72
Tchumtchoua, Sylvie, 56
Teh, Yee Whye, 51, 57, 78
Teichert, Adam R., 81
Tenenbaum, Josh, 28, 41, 60
Tenenbaum, Joshua, 55
Theodorou, Evangelos, 102
Ting, Jo-Anne, 33
Todorov, Emanuel, 99, 101
Tokdar, Surya, 54
Tomioka, Ryota, 111
Torres, David, 111
Toussaint, Marc, 103
Toyoizumi, Taro, 107
Tresp, Volker, 47
Tsang, Ivor, 119
Tu, Kewei, 80
Tu, Zhuowen, 76
Turian, Joseph, 80
Turner, Ryan, 58

Usunier, Nicolas, 79
Uther, William, 27

Vandenberghe, Lieven, 110
Veness, Joel, 29
Versino, Cristina, 39
Vert, Jean-Philippe, 95, 116
Vert, Jean-Phillipe, 25
Vijayakumar, Sethu, 33
Vincent, Pascal, 75
Virpioja, Sami, 81
Vishwanthan, S.V.N., 109
Vogelstein, Joshua, 65
von Luxburg, Ulrike, 62
Voortman, Mark, 65
Vuduc, Richard, 87
Vul, Ed, 60

Wackernagel, Hans, 83
Wakin, Michael, 97
Walker, Daniel, 87
Wallach, Hanna, 42
Wang, Bo, 87
Wang, Chong, 46, 47
Wang, Eric, 45
Wang, Pu, 57
Wang, Yu, 88
Warmuth, Manfred, 94
Waters, Austin, 45
Watkins, Chris, 78
Wehenke, Louis, 111

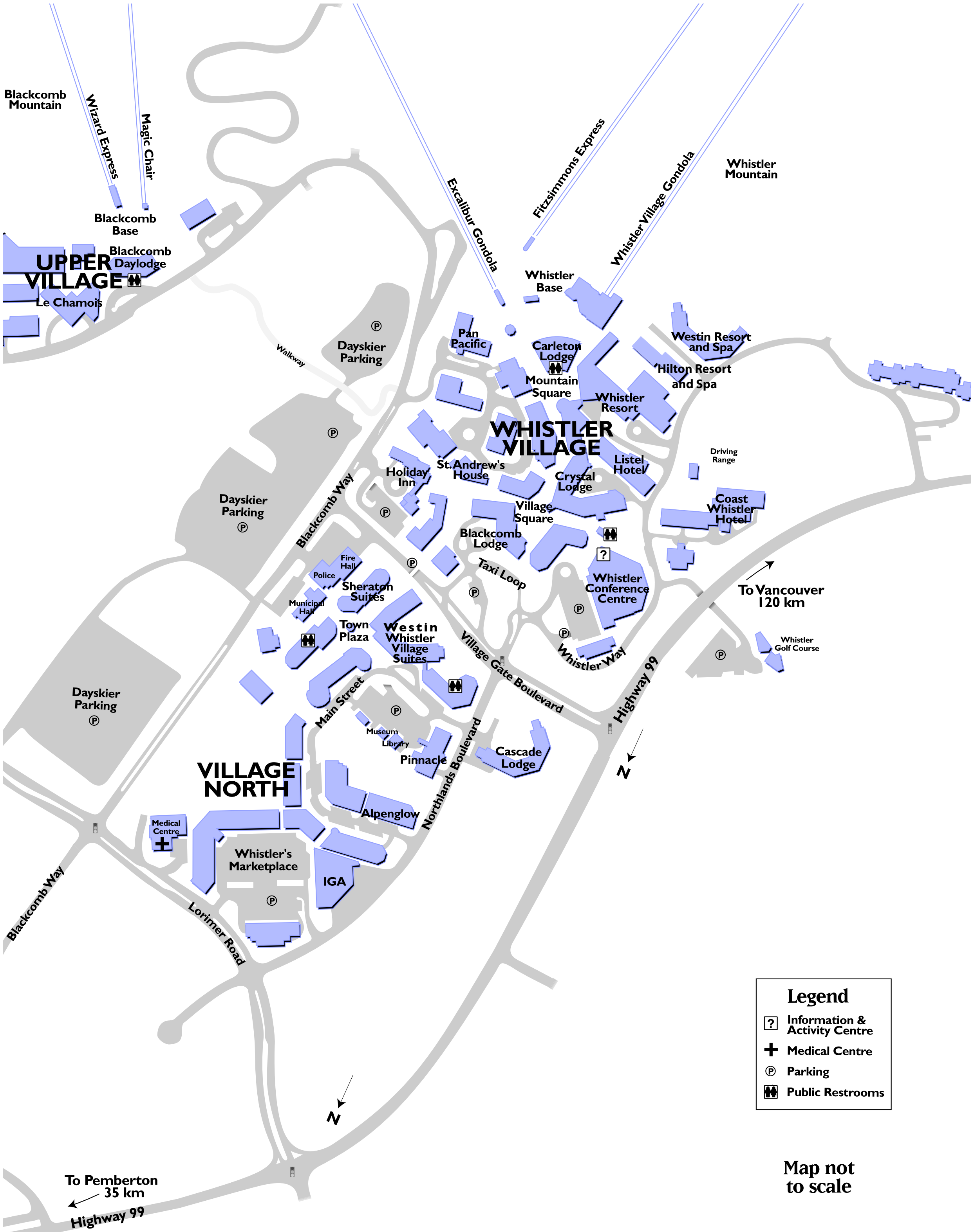
Wehenkel, Louis, 110
Welsh, Noel, 54
Weston, Jason, 79
White, Halbert, 21
Williams, Brian, 118
Williams, Chris, 90
Williams, Sam, 86
Williamson, Robert, 62
Williamson, Robert C., 75
Williamson, Sinead, 46
Wingate, David, 28
Winn, John, 75
Wood, Frank, 57
Wright, Stephen, 109
Wyatt, Jeremy, 54

Xing, Eric, 43, 46, 49, 54
Xu, Ningyi, 88

Yan, Jing, 87
Yang, Huangzhong, 88
Yang, Jiang-Ming, 47
Yang, Qiang, 119
Yao, Bangpeng, 65
Yu, Dong, 69, 70
Yu, Shi, 123
Yue, Yisong, 40
Yuret, Deniz, 81

Zhang, Lei, 47
Zhang, Tong, 33
Zheng, Alice X., 40
Zien, Alexander, 25, 95
Zinman, Aaron, 47
Zokaei A., Mohammed H., 122
Zucker, Steven, 67

Whistler Map

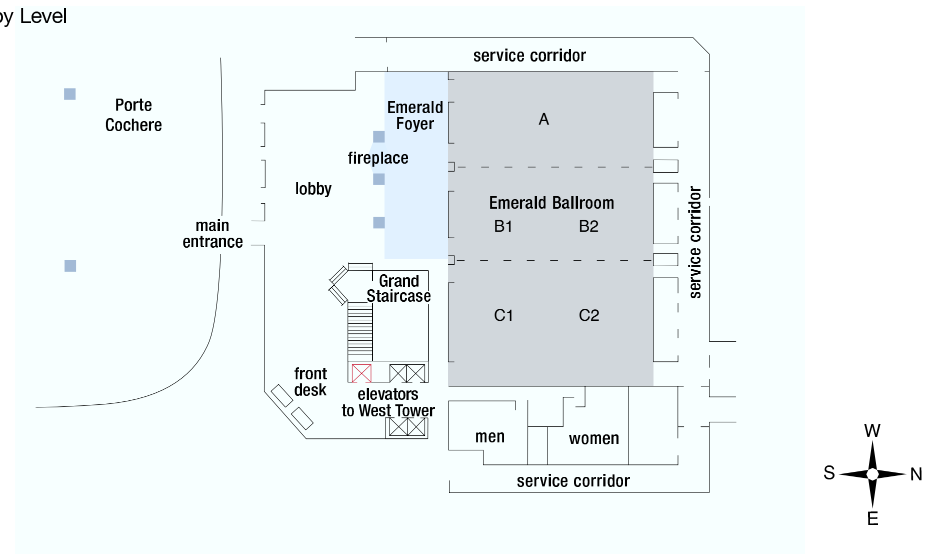


Legend
- ? Information & Activity Centre
- + Medical Centre
- Ⓟ Parking
- Public Restrooms

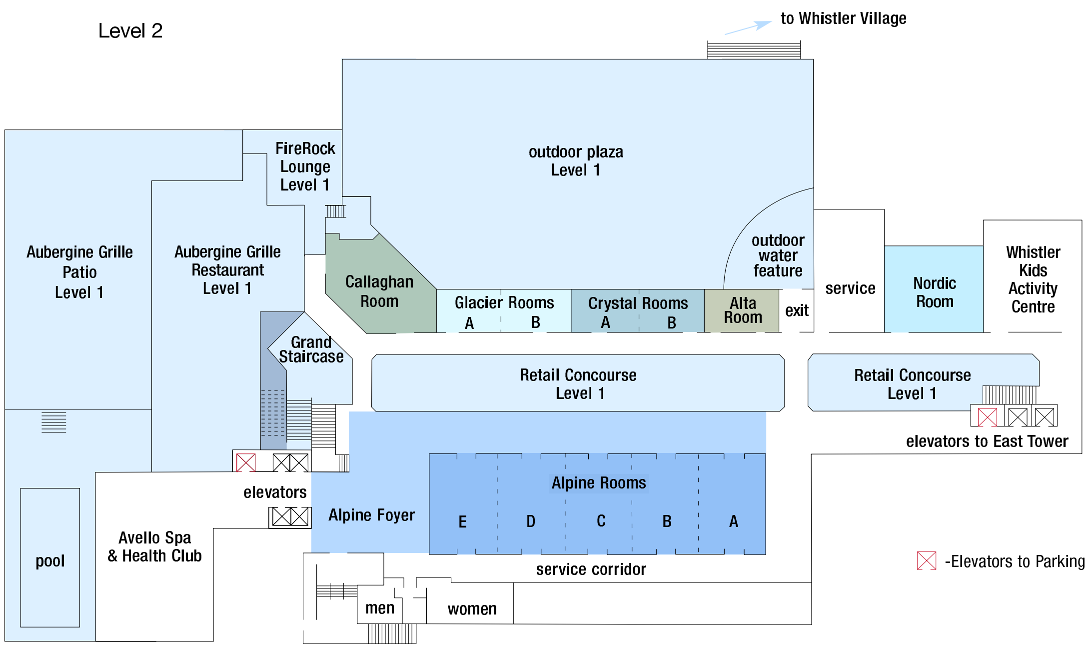
Map not to scale

Westin Resort Workshop Rooms

Lobby Level



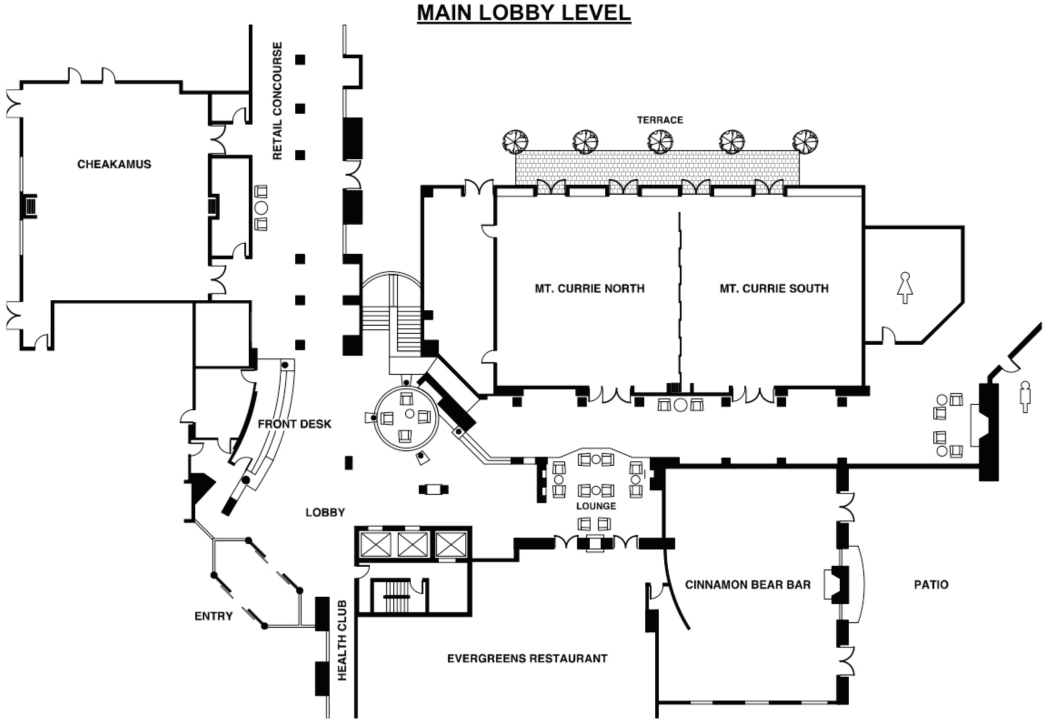
Level 2



Hilton Resort Workshop Rooms

Hilton
Whistler Resort

MAIN LOBBY LEVEL



LOWER LEVEL