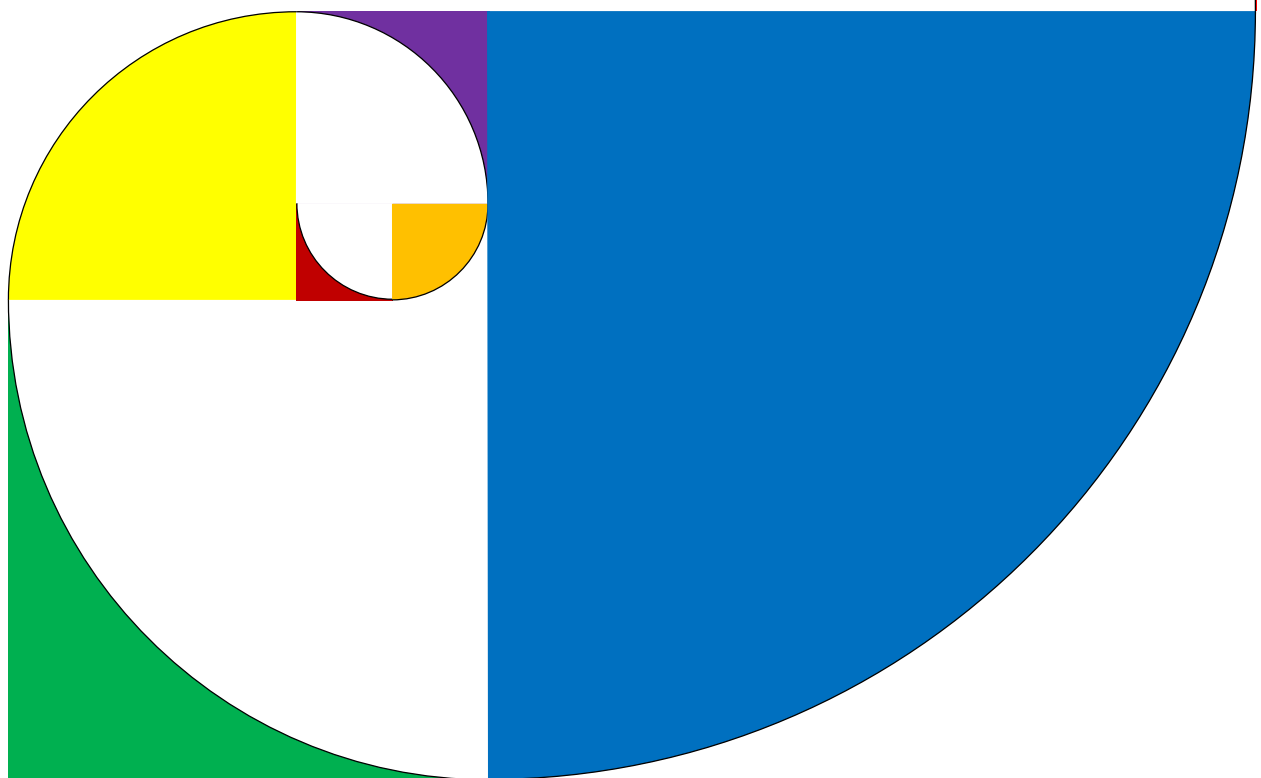# Uncertainty In Artificial Intelligence 2013

Bellevue, Washington, USA

Uncertainty in Artificial Intelligence

Proceedings of the Twenty-Eighth Conference (2013)

July 11-15, 2013, Bellevue, Washington, United States

Edited by
Ann Nicholson, Monash University, Australia
Padhraic Smyth, University of California, Irvine, United States

General Chairs
Nando De Freitas, University of British Columbia, Canada, and
        Oxford University, United Kingdom
Max Chickering, Microsoft Research, Redmond, United States

Local Arrangements Chair
Marina Meila, University of Washington, United States

Sponsored by
Microsoft Research, Artificial Intelligence Journal, Amazon.com,
Google Inc., Charles River Analytics, Toyota InfoTechnology
Center, U.S.A., Inc., IBM Research

Contents

Preface

This volume contains the schedule and abstracts of all papers that were accepted for the 29th Conference on Uncertainty in Artificial Intelligence (UAI), held in Bellevue, Washington, USA, from July 12 to 14th 2013. 233 papers were submitted to the conference and each was peer-reviewed by 3 or more reviewers. From the 233 papers, a total of 73 papers were accepted, 26 for oral presentation and 47 for poster presentation, for an acceptance rate of 31%. We are very grateful to the senior program committee and program committee members for their diligence in generating over 700 reviews for the 233 submitted papers.

In addition to the presentation of technical papers, the following invited speakers were also scheduled to give keynote talks at UAI 2013: Tom Mitchell (Carnegie Mellon University), Ralf Herbrich (Amazon), and Josh Tenenbaum (MIT).

A day of UAI 2013 tutorials is organized by tutorials chair David Sontag, on topics such as Computational Advertising and Causality (Leon Bottou), Large-Scale Distributed Learning with GraphLab (Carlos Guestrin), Statistical Methods in Genomics (Lior Pachter), and Polynomial Methods in Learning and Statistics (Ankur Moitra). UAI 2013 also features a day of workshops, organized by workshops chair John Mark Agosta, on topics such as Approaches to Causal Structure Learning, Big Data meet Complex Models, New Challenges in E-Commerice Recommendations, and Models for Spatial, Temporal, and Network Data.

Ann Nicholson and Padhraic Smyth (Program Co-Chairs)
Nando De Freitas and Max Chickering (General Co-Chairs)

Organizing Committee

Program Chairs

Ann Nicholson, Monash University, Australia
Padhraic Smyth, University of California, Irvine, USA

General Chairs

Nando De Freitas, University of British Columbia, Canada, and Oxford University, UK
Max Chickering, Microsoft Research, Redmond, USA

Local Arrangements Chair

Marina Meila, University of Washington, USA

Proceedings Chair

Alice Zheng, Microsoft Research, Redmond, USA

Tutorials Chair

David Sontag, New York University, USA

Workshops Chair

John Mark Agosta, Toyota Information Technology Center, Mountain View, USA

Acknowledgments

The success of a conference such as UAI depends greatly on the efforts of many individuals who volunteer their time to provide expert and detailed reviews of submitted papers. In particular, the Program Committee and Senior Program Committee for UAI 2013 were responsible for generating reviews and recommendations for the 233 submissions to the conference. Each submitted paper was reviewed by at least 3 members of the Program Committee. The Senior Program Committee then assessed the individual reviews for each paper, moderated discussion among Program Committee members if needed, and generated meta-reviews and recommendations for the program chairs. We are extremely grateful for the efforts of all of the individuals listed below.

Senior Program Committee

Adnan Darwiche	UCLA, USA
Denver Dash	Intel Labs, Pittsburgh, USA
Rina Dechter	University of California, Irvine, USA
Francisco J. Diez	UNED, Spain
Norman Fenton	Queen Mary University, London, UK
Alan Fern	Oregon State University, USA
Emily Fox	University of Washington, USA
Lise Getoor	University of Maryland, USA
Alexander Ihler	University of California, Irvine, USA
Tommi Jaakkola	MIT, USA
Dominik Janzing	Max Planck Institute for Biological Cybernetics, Germany
Kristian Kersting	Fraunhofer IAIS, University of Bonn, Germany
Helge Langseth	The Norwegian University of Science and Technology, Norway
Kathryn Laskey	George Mason University, USA
Guy Lebanon	Georgia Tech/Amazon, USA
Tze Yun Leong	National University of Singapore
Kevin Leyton-Brown	University of British Columbia, Canada
Chris Meek	Microsoft Research, USA
Brian Milch	Google, USA
Remi Munos	INRIA Lille, France
Petri Myllymaki	Helsinki Institute for Information Technology, Finland
Thomas D. Nielsen	Aalborg University, Denmark
Nuria Oliver	Telefonica, Spain
Ronald Parr	Duke University, USA
Avi Pfeffer	Charles River Analytics, USA
David Poole	University of British Columbia, Canada
Greg Provan	University College Cork, Ireland
Thomas Richardson	University of Washington, USA
Dale Schuurmans	University of Alberta, Canada
Christian Shelton	University of California, Riverside, USA
Prakash Shenoy	University of Kansas, USA
Ricardo Silva	University College London, UK
Alex Smola	Carnegie Mellon University, USA
Peter Spirtes	Carnegie Mellon University, USA
Umberto Straccia	ISTI-CNR, Italy
Jin Tian	Iowa State University, USA
Yair Weiss	Hebrew University, Israel

Michael Wellman	University of Michigan, USA
Dit-Yan Yeung	The Hong Kong University of Science and Technology, Hong Kong
Nevin Zhang	The Hong Kong Unversity of Science and Technology, Hong Kong

Program Committee

Ryan P. Adams	Harvard University, USA
Arvind Agarwal	University of Maryland, USA
Babak Ahmadi	Fraunhofer Institute, Germany
Amr Ahmed	Google, USA
Ayesha Ali	University of Guelph, Canada
Russell Almond	Florida State University, USA
Animashree Anandkumar	University of California, Irvine, USA
Dragomir Anguelov	Google, USA
Alessandro Antonucci	IDSIA, Switzerland
Cedric Archambeau	Xerox Research Centre Europe, France
Nimar Arora	Bayesian Logic, Inc., USA
Arthur Asuncion	Google, USA
Elias Bareinboim	UCLA, USA
Dhruv Batra	Virgina Tech, USA
Salem Benferhat	CRIL, University of Artois, France
Carlo Berzuini	University of Manchester, UK
Alina Beygelzimer	IBM Research, USA
Bozhena Bidyuk	Google, USA
Charles Blundell	Gatsby Unit, UCL, UK
Fernando Bobillo	Universidad de Zaragoza, Spain
Guillaume Bouchard	Xerox Research Centre Europe, France
Alexandre Bouchard-Cote	University of British Columbia, Canada
Abdeslam Boularias	Max Planck Institute, Germany
Darius Braziunas	Kobo, Canada
Emma Brunskill	Carnegie Mellon University, USA
Wray Buntine	NICTA, Canberra, Australia
Tiberio Caetano	NICTA, Canberra, Australia
Joaquin Candela	Facebook, USA
Peter Carbonetto	University of Chicago, USA
Lawrence Carin	Duke, USA
Francois Caron	INRIA Bordeaux, France
Robert Castelo	Universitat Pompeu Fabra, Spain
Hei Chan	National Institute of Informatics, Japan
Laurent Charlin	University of Toronto, Canada
Tianjiao Chu	University of Pittsburgh, USA
Tom Claassen	Radboud University Nijmegen, The Netherlands
Adam Coates	Stanford University, USA
Barry Cobb	Virginia Military Institute, USA
Ira Cohen	Hewlett Packard Labs, USA
Diego Colombo	ETH, Switzerland
Paulo Costa	George Mason University, USA
Fabio Cozman	University of Sao Paolo, Brazil
Mark Craven	University of Wisconsin, Madison, USA
James Cussens	University of York, UK
Cassio de Campos	IDSIA, Switzerland
Luis M. De Campos	University of Granada, Spain
Frank Dellaert	Georgia Tech, USA
Sebastien Destercke	CNRS, France
Marek J. Druzdzel	University of Pittsburgh, USA

David Dunson	Duke University, USA
Jennifer Dy	Northeastern University, USA
Gal Elidan	Hebrew University of Jerusalem, Israel
Helene Fargier	IRIT-CNRS, FRance
M. Julia Flores	University of Castilla—La Mancha (UCLM), Spain
James Foulds	University of California, Irvine, USA
Charless Fowlkes	University of California, Irvine, USA
Minos Garofalakis	Technical University of Crete, Greece
Thomas Gartner	University of Bonn, Germany
Jan Gasthaus	University College London, UK
Joydeep Ghosh	University of Texas, Austin, USA
Mark Girolami	University College London, UK
Amir Globerson	Hebrew University, Israel
Vibhav Gogate	University of Texas at Dallas, USA
Moises Goldszmidt	Microsoft Research, USA
Manuel Gomez-Olmedo	Universidad de Granada, Spain
Christophe Gonzales	LIP6-UPMC, France
Mihajlo Grbovic	Yahoo! Labs, USA
Shengbo Guo	Samsung Information Systems America, USA
Yuhong Guo	Temple University, USA
Maya Gupta	University of Washington, USA
Hannaneh Hajishirzi	University of Washington, USAS
David Heckerman	Microsoft Research, USA
Philipp Hennig	Max Planck Institute, Germany
Ralf Herbrich	Amazon, USA
Jesse Hoey	University of Waterloo, Canada
Antti Honkela	University of Helsinki, Finland
Eric Horvitz	Microsoft Research, USA
William Hsu	Kansas State University, USA
Bert Huang	University of Maryland, USA
David Jensen	University of Massachussetts, USA
Albert Xin Jiang	USC, USA
Nebojsa Jojic	Microsoft Research
Kshitij Judah	Oregon State University
Ece Kamar	Microsoft Research
Ashish Kapoor	Microsoft Research
Kalev Kask	University of California, Irvine, USA
Seyoung Kim	Carnegie Mellon University, USA
Sergey Kirshner	Purdue University, USA
David Knowles	Stanford University, USA
Mikko Koivisto	Helsinki Institute for Information Technology, Finland
Vladimir Kolmogorov	Institute of Science and Technology, Austria
Petri Kontkanen	HIIT/CoSCo, Finland
Kevin Korb	Monash University, Australia
Timo Koski	KTH, Sweden
Wojciech Kotlowski	University of Poznan, Poland
Alex Kulesza	University of Michigan, USA
Branislav Kveton	Technicolor Labs, USA
Jerome Lang	LAMSADE, CNRS/Universite Paris-Dauphine, France
Tobias Lang	FU Berlin, Germany
Hugo Larochelle	Universite de Sherbrooke, Canada
Erik Learned-Miller	University of Massachusetts Amherst, USA
Anthony Lee	University of Warwick, USA
Jan Lemeire	Vrije Universiteit Brussel, Belgium
Philippe Leray	University of Nantes, France

Jennifer Listgarten	Microsoft Research, USA
Qiang Liu	University of California, Irvine, USA
Weiru Liu	Queen's University Belfast, UK
Dan Lizotte	University of Waterloo, Canada
Ben London	University of Maryland, USA
Daniel Lowd	University of Oregon, USA
Peter Lucas	Radboud University Nijmegen, The Netherlands
Michael Lyu	The Chinese University of Hong Kong, Hong Kong
Marloes Maathuis	ETH Zurich, Switzerland
Malik Magdon-Ismail	Rensselear Polytechnic Institute, USA
Subramani mani	University of New Mexico, USA
Vikash Mansinghka	MIT, USA
Radu Marinescu	IBM Research, USA
Benjamin Marlin	University of Massachusetts Amherst, USA
Maria Vanina Martinez	The University of Oxford, UK
Andre Martins	Priberam Labs, Portugal
Jon Mcauliffe	University of California, Berkeley, USA
Andrew McCallum	University of Massachusetts Amherst, USA
Talya Meltzer	Hebrew University of Jerusalem, Israel
Ole Mengshoel	Carnegie Mellon University, USA
Taneli Mielikainen	Nokia Research Center Palo Alto, USA
David Mimno	Princeton University, USA
Thomas Minka	Microsoft Research, UK
Joris Mooij	Radboud University Nijmegen, The Netherlands
Sach Mukherjee	Netherlands Cancer Institute, The Netherlands
Iain Murray	University of Edinburgh, UK
Sriraam Natarajan	Wake Forest University Baptist Medical Center, USA
William S. Noble	University of Washington, USA
Michael Osborne	The University of Oxford, UK
Lars Otten	Google, USA
David Page	Department of Biostatistics, University of Wisconsin, Madison, USA
John Paisley	University of California, Berkeley, USA
Jason Pazis	Duke University, USA
Jose Pena	Linkoping University, Sweden
David M. Pennock	Microsoft, USA
Patrice Perny	LIP6–UPMC, France
Jonas Peters	ETH Zurich, Switzerland
Nathalie Peyrard	INRA, France
Kim-Leng Poh	National University of Singapore, Singapore
Hoifung Poon	Microsoft Research, USA
Leonard Poon	Hong Kong Institute of Education, Hong Kong
Bob Price	PARC, USA
David V. Pynadath	University of Southern California, Institute for Creative Technologies, USA
Yuan (Alan) Qi	Purdue University, USA
Zengchang Qin	Beihang University, China
Erik Quaeghebeur	Ghent University, Belgium
Emmanuel Rachelson	ISAE Supaero, France
Piyush Rai	University of Texas, Austin, USA
Roland Ramsahai	University of Cambridge, UK
Vinayak Rao	Duke University, USA
Silja Renooij	University of Utrecht, The Netherlands
Teemu Roos	Helsinki Institute for Information Technology, Finland
Dan Roy	University of Cambridge, UK
Mike Ruberry	Harvard University, USA
Rafael Rumi	Almeria University, Spain

Brian Ruttenberg	Charles River Analytics, USA
S.V.N. Vishwanathan	Purdue University, USA
Regis Sabbadin	INRA, France
Ruslan Salakhutdinov	University of Toronto, Canada
Antonio Salmeron	Almeria University, Spain
Jose San Pedro	Telefonica, Spain
Scott Sanner	NICTA/Australian National University, Australia
Cristina Savin	Cambridge University, UK
Mark Schmidt	INRIA/ENS, France
Jeff Schneider	Carnegie Mellon University, USA
Matthias Seeger	EPFL, Switzerland
Bart Selman	Cornell University, USA
Ross D. Shachter	Stanford University, USA
Cosma Shalizi	Carnegie Mellon University, USA
Ilya Shpitser	University of Southampton, UK
Gerardo I. Simari	The University of Oxford, UK
Tomas Singliar	Boeing, USA
Jim Smith	The University of Warwick, UK
Le Song	GeorgiaTech, USA
Mark Steyvers	University of California, Irvine, USA
Andreas Stuhlmuller	MIT and Stanford University, USA
L. Enrique Sucar	INAOE, Mexico
Joe Suzuki	Osaka University, Japan
Graham Taylor	University of Guelph, Canada
Joshua Tenenbaum	MIT, USA
Robert E. Tillman	Rotella Capital Management, USA
Volker Tresp	Siemens AG, Germany
Matthias Troffaes	Durham University, UK
Charles Twardy	George Mason University, USA
Raquel Urtasun	Toyota Technological Institute at Chicago, USA
Marco Valtorta	University of South Carolina, USA
Laurens van der Maaten	Delft University of Technology, The Netherlands
Vladimir Vovk	Royal Holloway, University of London, UK
Christopher Wakefield	Duke University, USA
Chong Wang	Princeton University, USA
Hao Wang	University of Southern California, USA
Yi Wang	Institute of High Performance Computing, Singapore
Renata Wassermann	Universidade de Sao Paulo, Brazil
Wim Wiegerinck	Radboud University Nijmegen, The Netherlands
Changhe Yuan	Queens College/City University of New York, USA
Kun Zhang	Max Planck Institute for Intelligent Systems, Germany
Onno Zoeter	Xerox Research Centre Europe, France
Geoff Zweig	Microsoft Research, USA

Additional Reviewers

Additional Acknowledgments

Sponsors

We gratefully acknowledge the generous support provided by our sponsors, including support for student best paper awards. Without our sponsors' support it would not be feasible to organize a conference such as UAI 2013 without charging much higher registration fees.

Tutorials

Thursday July 11th

8:00am–9:50am   Polynomial Methods in Learning and Statistics
Ankur Moitra, Institute for Advanced Study

10:00am–11:50am   Large-Scale Distributed Machine Learning with GraphLab
Carlos Guestrin, University of Washington

1:30pm–3:20pm   Statistical Methods in Genomics
Lior Pachter, University of California, Berkeley

3:30pm–5:20pm   Computational Advertising & Causality
Leon Bottou, Microsoft Research

Tutorial: Polynomial Methods in Learning and Statistics

Ankur Moitra

I will survey some of the emerging applications of the polynomial method in learning and statistics through three in-depth examples. The first part of the talk will focus on the problem of learning the parameters of a mixture of Gaussians, given random samples from the distribution. The theme here is that the method of moments can be shown to provably recover good estimates to the true parameters precisely because systems of polynomial equations are relatively stable. In fact, I will show that noisy estimates of the first six moments of a mixture of two Gaussians suffice to recover good estimates to the true mixture parameters, thus justifying an approach that dates back to Karl Pearson.

Next, I will consider the problem of learning the parameters of a topic model given random samples (documents) from this generative model. There has been considerable recent progress on this problem, and one line of attack exploits the spectral properties of certain tensors to give an algorithm for learning the parameters of a Latent Dirichlet Allocation model provided that the topic vectors are linearly independent. Another approach introduces a natural condition (separability) under which simple algorithms again based on the method of moments can be shown to work for arbitrary topic models, even ones which allow correlations between pairs of topics.

Lastly, I will introduce the population recovery problem where the goal is to find a representative set of templates given highly incomplete samples. This is a basic problem that arises in a number of settings, and I will present the first polynomial time algorithm (improving on an earlier quasi-polynomial time algorithm). Even though this algorithm is not based on the method of moments, again the polynomial method and tools from complex analysis turn out to be useful for analyzing our algorithm.

This survey will cover in part a number of papers by many authors including Anandkumar, Arora, Belkin, Foster, Ge, Hsu, Kalai, Kakade, Liu, Saks, Sinha, Wigderson, Valiant, Yehudayoff and myself.

Tutorial: Large-Scale Distributed Machine Learning with GraphLab

Carlos Guestrin

Today, machine learning (ML) methods play a central role in industry and science. The growth of the Web and improvements in sensor data collection technology have been rapidly increasing the magnitude and complexity of the ML tasks we must solve. This growth is driving the need for scalable, parallel ML algorithms that can handle "Big Data."

Unfortunately, translating these parallel algorithms into distributed systems that can handle this big data remains a significant challenge. Thus, the state-of-the-art algorithms in ML are often seen as "academic exercises," which are implemented in the real-world.

In this tutorial, I will focus on three steps for addressing this problem:

1. Examining common algorithmic patterns in distributed ML methods for tackling big data.

2. Qualifying the challenges of implementing these algorithms in real distributed systems.

3. Describing computational frameworks for implementing these algorithms at scale.

The goal of this tutorial is to help researchers tune their efforts on algorithms for big data, and enable them to easily translate their algorithms into systems that can be used in practical settings in industry and academia.

In the latter part, we will focus mainly on the GraphLab framework, which naturally expresses asynchronous, dynamic graph computations that are key for state-of-the-art ML algorithms. When these algorithms are expressed in our higher-level abstraction, GraphLab will effectively address many of the underlying parallelism challenges, including data distribution, optimized communication, and guaranteeing sequential consistency, a property that is surprisingly important for many ML algorithms. On a variety of large-scale tasks, GraphLab provides 20-100x performance improvements over Hadoop. In recent months, GraphLab has received many tens of thousands of downloads, and is being actively used by a number of startups, companies, research labs and universities.

Tutorial: Statistical Methods in Genomics

Lior Pachter

"Sequence census methods" couple high-throughput DNA sequencing to biochemical assays and are transforming standard functional genomics assays into powerful genome-wide probes of biochemical activity.

In this tutorial, I'll provide an introduction to such methods with a focus on the crucial role of machine learning in the development, execution and interpretation of experiments. I will first focus on the prototype experiment, called RNA-Seq, and provide a walk-through of the chemistry, biology and computation that comprise the experiment, followed by a survey of machine learning techniques that have been applied to RNA-seq data. I will also highlight new questions in machine learning that arise in this context.

I will then examine the broader scope of the field, with a view towards providing a classification and categorization of sequence census methods.

Tutorial: Computational Advertising & Causality

Leon Bottou

Statistical machine learning technologies in the real world are never without a purpose. Using their predictions, humans or machines make decisions whose circuitous consequences often violate the modeling assumptions that

justified the system design in the first place. Such contradictions appear very clearly in computational advertisement systems. The design of the ad placement engine directly influences the occurrence of clicks and the corresponding advertiser payments. It also has important indirect effects : (a) ad placement decisions impact the satisfaction of the users and therefore their willingness to frequent this web site in the future, (b) ad placement decisions impact the return on investment observed by the advertisers and therefore their future bids, and (c) ad placement decisions change the nature of the data collected for training the statistical models in the future.

Popular theoretical approaches, such as auction theory or multi-armed bandits, only address selected aspects of such a system. In contrast, the language and the methods of causal inference provide a full set of tools to answer the vast collection of questions facing the designer of such a system. Is it useful to pass new input signals to the statistical models? Is it worthwhile to collect and label a new training set? What about changing the loss function or the learning algorithm? In order to answer such questions, one needs to unravel how the information produced by the statistical models traverses the web of causes and consequences and produces measurable losses and rewards.

This tutorial provides a real world example demonstrating the value of causal inference for large-scale machine learning. It also illustrates a collection of practical counterfactual analysis techniques applicable to many real-life machine learning systems, including causal inferences techniques applicable to continuously valued variables with meaningful confidence intervals, and quasi-static analysis techniques for estimating how small interventions affect certain causal equilibria. In the context of computational advertisement, this analysis elucidates the connection between auction theory and machine learning.

Conference Program

Friday July 12th

8:30am - Welcome and Opening Remarks

Keynote Talk

8:40am   Never Ending Learning
Tom Mitchell, CMU

Session 1: Classification

9:40am   Modeling Documents with Deep Boltzmann Machines
Nitish Srivastava, Ruslan Salakhutdinov, Geoffrey Hinton

10:05am   Generative Multiple-Instance Learning Models For Quantitative Electromyography
Tameem Adel, Ruth Urner, Benn Smith, Daniel Stashuk, Dan Lizotte

10:30 - Coffee Break

Session 2: Structure Learning

11:00am   Advances in Bayesian Network Learning using Integer Programming
James Cussens, Mark Bartlett

11:25am   Treedy: A Heuristic for Counting and Sampling Subsets
Teppo Niinimäki, Mikko Koivisto

11:50am   Evaluating Anytime Algorithms for Learning Optimal Bayesian Networks
Brandon Malone, Changhe Yuan

12:15 - Lunch Break

Session 3a: Stochastic Processes

2:00pm   Learning Periodic Human Behaviour Models from Sparse Data for Crowdsourcing Aid Delivery in Developing Countries
James McInerney, Alex Rogers, Nicholas Jennings

2:25pm   Hilbert Space Embeddings of Predictive State Representations
Byron Boots, Geoffrey Gordon, Arthur Gretton

Session 3b: Novel Modeling Approaches

2:50pm   On the Complexity of Strong and Epistemic Credal Networks
Denis Maua, Cassio de Campos, Alessio Benavoli, Alessandro Antonucci

3:15pm   Sparse Nested Markov models with Log-linear Parameters
Ilya Shpitser, Robin Evans, Thomas Richardson, James Robins

Poster Spotlights 1

3:40pm - 1 minute poster spotlight presentations from papers:

248   Preference Elicitation For General Random Utility Models
  Hossein Azari Soufiani, David Parkes, Lirong Xia
61   Bennett-type Generalization Bounds: Large-deviation Case and Faster Rate of Convergence
  Chao Zhang, Jieping Ye
80   Inverse Covariance Estimation for High-Dimensional Data in Linear Time and Space:
  Spectral Methods for Riccati and Sparse Models
  Jean Honorio, Tommi Jaakkola
146   High-dimensional Joint Sparsity Random Effects Model for Multi-task Learning
  Krishnakumar Balasubramanian, Kai Yu, Tong Zhang
176   Hinge-loss Markov Random Fields: Convex Inference for Structured Prediction
  Stephen Bach, Bert Huang, Ben London, Lise Getoor
30   SparsityBoost: A New Scoring Function for Learning Bayesian Network Structure
  Eliot Brenner, David Sontag
47   Scoring and Searching over Bayesian Networks with Causal and Associative Priors
  Giorgos Borboudakis, Ioannis Tsamardinos
63   Speedy Model Selection (SMS) for Copula Models
  Yaniv Tenzer, Gal Elidan
66   Learning Max-margin Tree Predictors
  Ofer Meshi, Elad Eban, Gal Elidan, Amir Globerson
76   Multiple Instance Learning by Discriminative Training of Markov Networks
  Hossein Hajimirsadeghi, Jinling Li, Greg Mori, Tarek Sayed, Mohammad Zaki
88   Collective Diffusion Over Networks: Models and Inference
  Akshat Kumar, Dan Sheldon, Biplav Srivastava
225   Probabilistic Conditional Preference Networks
  Damien Bigot, Bruno Zanuttini, Helene Fargier, Jérôme Mengin
37   Reasoning about Probabilities in Dynamic Systems using Goal Regression
  Vaishak Belle, Hector Levesque
151   Structured Convex Optimization under Submodular Constraints
  Kiyohito Nagano, Yoshinobu Kawahara
161   Finite-Time Analysis of Kernelised Contextual Bandits
  Remi Munos, Michal Valko, Nathaniel Korda, Ilias Flaounas, Nello Cristianini
54   Unsupervised Learning of Noisy-OR Bayesian Networks
  Yonatan Halpern, David Sontag

4pm to 6pm - Poster Session 1 and Coffee Break
All oral presentations from Friday and all poster spotlights.

7pm - Conference Banquet

Saturday July 13th

Keynote Talk

8:30am   Bayesian Learning in Online Service: Statistics Meets Systems
Ralf Herbrich, Amazon

Session 4: Ranking and Semi-Supervised Learning

9:40am   The Lovasz Bregman Divergences and connections to rank aggregation, clustering and web ranking
Rishabh Iyer, Jeff Bilmes

10:05am   Active Learning with Expert Advice
Peilin Zhao, Steven Hoi, Jinfeng Zhuang

10:30 - Coffee Break

Session 5: Inference Algorithms

11:00am   Optimization With Parity Constraints: From Binary Codes to Discrete Integration
Stefano Ermon, Carla Gomes, Ashish Sabharwal, Bart Selman

11:25am   Structured Message Passing
Vibhav Gogate, Pedro Domingos

11:50am   Lower Bounds for Exact Model Counting and Applications in Probabilistic Databases
Paul Beame, Jerry Li, Sudeepa Roy, Dan Suciu

12:15 - Lunch Break

Session 6a: Causal Models

2:00pm   Learning Sparse Causal Models is not NP-hard
Tom Claassen, Joris Mooij, Tom Heskes

2:25pm   Cyclic Causal Discovery from Continuous Equilibrium Data
Joris Mooij, Tom Heskes

Session 6b: Learning

2:50pm   Scalable Matrix-valued Kernel Learning for High-dimensional Nonlinear Multivariate Regression and Granger Causality
Vikas Sindhwani, Ha Quang Minh, Aurelie Lozano

3:15pm   Constrained Bayesian Inference for Low Rank Multitask Learning
Oluwasanmi Koyejo, Joydeep Ghosh

Poster Spotlights 2

3:40pm - 1 minute poster spotlight presentations from papers:

4   One-Class Support Measure Machines for Group Anomaly Detection
     Krikamol Muandet, Bernhard Schölkopf
8   Convex Relaxations of Bregman Divergence Clustering
     Hao Cheng, Xinhua Zhang, Dale Schuurmans
132 Stochastic Rank Aggregation
     Shuzi Niu, Yanyan Lan, Jiafeng Guo, Xueqi Cheng
109 Warped Mixtures for Nonparametric Cluster Shapes
     Tomoharu Iwata, David Duvenaud, Zoubin Ghahramani
110 Identifying Finite Mixtures of Nonparametric Product Distributions and Causal Inference
     of Confounders
     Eleni Sgouritsa, Dominik Janzing, Jonas Peters, Bernhard Schölkopf
118 The Supervised IBP: Neighbourhood Preserving Infinite Latent Feature Models
     Novi Quadrianto, Viktoriia Sharmanska, David A. Knowles, Zoubin Ghahramani
192 Normalized Online Learning
     John Langford, Paul Mineiro, Stephane Ross
131 Discovering Cyclic Causal Models with Latent Variables: A General SAT-Based Procedure
     Antti Hyttinen, Patrik Hoyer, Frederick Eberhardt, Matti Järvisalo
197 A Sound and Complete Algorithm for Learning Causal Models from Relational Data
     Marc Maier, Katerina Marazopoulou, David Arbour, David Jensen
103 Tighter Linear Program Relaxations for High Order Graphical Models
     Elad Mezuman, Daniel Tarlow, Amir Globerson, Yair Weiss
199 Solving Limited-Memory Influence Diagrams Using Branch-and-Bound Search
     Arindam Khaled, Changhe Yuan, Eric Hansen
244 Gaussian Processes for Big Data
     James Hensman, Nicolo Fusi, Neil Lawrence
24   From Ordinary Differential Equations to Structural Causal Models: the deterministic case
     Joris Mooij, Dominik Janzing, Bernhard Schölkopf
70   Causal Transportability of Experiments on Controllable Subsets of Variables: z-Transportability
     Sanghack Lee, Vasant Honavar
189 Calculation of Entailed Rank Constraints in Partially Non-Linear and Cyclic Models
     Peter Spirtes

4pm to 6pm - Poster Session 2 and Coffee Break
All oral presentations from Saturday and all poster spotlights.

Sunday July 14th

Keynote Talk

8:30am   Modeling Common-Sense Scene Understanding with Probabilistic Programs
Josh Tenenbaum, MIT

Session 7: Modeling Human Behavior

9:40am   Active Sensing as Bayes-Optimal Sequential Decision Making
Sheeraz Ahmad, Angela Yu

10:05am   Evaluating computational models of explanation using human judgments
Michael Pacer, Joseph Williams, Xi Chen, Tania Lombrozo, Thomas Griffiths

10:30 - Coffee Break

Session 8: Markov Decision Processes

11:00am   Bounded Approximate Symbolic Dynamic Programming for Hybrid MDPs
Luis Gustavo Vianna, Scott Sanner, Leliane de Barros

11:25am   Approximation of Lorenz-Optimal Solutions in Multiobjective Markov Decision Processes
Judy Goldsmith, Josiah Hanna, Patrice Perny, Paul Weng

11:50am   POMDPs under Probabilistic Semantics
Krishnendu Chatterjee, Martin Chmelík

12:15 - Lunch Break

Session 9: Inference Algorithms

2:00pm   Bethe-ADMM for Tree Decomposition based Parallel MAP Inference
Qiang Fu, Huahua Wang, Arindam Banerjee

2:25pm   On MAP Inference by MWSS on Perfect Graphs
Adrian Weller, Tony Jebara

2:50pm   Automorphism Groups of Graphical Models and Lifted Variational Inference
Hung Bui, Tuyen Huynh, Sebastian Riedel

Poster Spotlights 3

3:40pm - 1 minute poster spotlight presentations from papers:

135   Boosting in the presence of label noise
Jakramate Bootkrajang, Ata Kaban
143   Building Bridges: Viewing Active Learning from the Multi-Armed Bandit Lens
Ravi Ganti, Alexander Gray
134   Parallel Gaussian Process Regression with Low-Rank Covariance Matrix Approximations
Jie Chen, Nannan Cao, Kian Hsiang Low, Colin Keng-Yan Tan, Patrick Jaillet

4pm to 6pm - Poster Session 3 and Coffee Break

Invited Speakers

Keynote Lecture: Never ending learning

Tom Mitchell

We will never really understand machine learning until we can build machines that learn many different things, over years, and become better learners over time.

This talk describes our research to build a Never-Ending Language Learner (NELL) that runs 24 hours per day, forever, learning to read the web. Each day NELL extracts (reads) more facts from the web, and integrates these into its growing knowledge base of beliefs. Each day NELL also learns to read better than yesterday, enabling it to go back to the text it read yesterday, and extract more facts, more accurately. NELL has been running 24 hours/day for over two years now. The result so far is a collection of 15 million interconnected beliefs (e.g., servedWith(coffee, applePie), isA(applePie, bakedGood)), that NELL is considering at different levels of confidence, along with hundreds of thousands of learned phrasings, morphoogical features, and web page structures that NELL uses to extract beliefs from the web. Track NELL's progress at http://rtw.ml.cmu.edu.

Keynote Lecture: Bayesian Learning in Online Service: Statistics Meets Systems

Ralf Herbrich

Over the past few years, we have entered the world of big and structured data - a trend largely driven by the exponential growth of Internet-based online services such as Search, eCommerce and Social Networking as well as the ubiquity of smart devices with sensors in everyday life. This poses new challenges for statistical inference and decision-making as some of the basic assumptions are shifting:

- The ability to optimize both the likelihood and loss functions
- The ability to store the parameters of (data) models
- The level of granularity and 'building blocks' in the data modeling phase
- The importance of priors vs. likelihoods
- The interplay of computation, storage, communication and inference and decision-making techniques

In this talk, I will discuss the implications of big and structured data for Statistics and the convergence of statistical model and distributed systems. I will present one of the most versatile modeling techniques that combines systems and statistical properties - factor graphs - and review a series of approximate inference techniques such as distributed message passing and Gibbs sampling. The talk will be concluded with an overview of real-world problems at Amazon.

Keynote Lecture: Modeling common-sense scene understanding with probabilistic programs

Josh Tenenbaum

To see is, famously, to "know what is where by looking". Yet to see is also to know what will happen, what can be done, and what is being done – to detect not only objects and their locations, but the physical dynamics

governing how objects in the scene interact with each other and how agents can act on them, and the psychological dynamics governing how intentional agents in the scene interact with these objects and each other to achieve their goals. I will talk about recent efforts to capture these core aspects of human common-sense scene understanding in computational models, which can also be used for building more human-like machine vision and reasoning systems. These models of intuitive physics and intuitive psychology take the form of *probabilistic programs*: probabilistic generative models defined not over graphs, as in many current machine learning and vision models, but over programs whose execution traces describe the causal processes giving rise to the behavior of physical objects and intentional agents. Common-sense physical and psychological scene understanding can then be characterized as approximate Bayesian inference over these probabilistic programs. We study how this approach can solve a wide range of problems including inferring scene structure from images, predicting physical dynamics and inferring latent physical attributes from static images or short movies, and reasoning about the goals and beliefs of agents from observations of short action traces.

Workshop Program

Monday July 15th

Room: Lake Washington A

Bayes Application Workshop I: Big Data Meets Complex Models

8:30am	Combined Applications Workshop Introductions
9:00–10:15am	Session 1: Extracting information about individuals from complex data
10:15–10:30am	Coffee Break
10:30-11:45am	Session 2: Combining learning and expert opinion
11:45am-1:00pm	Break (lunch on your own)

Bayes Application Workshop II: Spatial, Temporal, and Network Models

1:00–2:30pm	Session 3
2:30-2:45pm	Coffee Break
2:45–3:45pm	Session 4
4:00–5:00pm	Combined Applications Workshop Wrap-up and Planning Discussion

Room: Lake Chelan

Approaches to Causal Structure Learning Workshop

8:45–10:00am	Session 1
10:00–10:30am	Coffee Break
10:30am–12:15pm	Session 2
12:15–2:00pm	Break (lunch on your own)
2:00–3:25pm	Session 3
3:25–3:35pm	Poster Spotlights
3:35–4:50pm	Posters
4:50–6:00pm	Session 4

Room: Lake Coeur d'Alene

New Challenges in E-Commerce Recommendations

8:30–9:30am	*Alex Smola*
9:30–10:00am	Coffee Break
10:00am–12:15pm	Oral Presentation 1
12:15–2:00pm	Break (lunch on your own)
2:00–3:00pm	Recommendations at Netflix, *Carlos Gomez Uribe (Netflix)*
3:00–3:15pm	Coffee Break
3:15–4:45pm	Oral Presentation 2
4:45–5:00pm	Panel Discussion

Bayes Application Workshop I: Big Data Meets Complex Models

`http://pluto.coe.fsu.edu/UAI2013Workshops/BigData/`

Determining causal relationships from observations and experiments is fundamental to human reasoning, decision making and the advancement of science. The aim of this workshop is to bring together researchers interested in the challenges of causal structure learning from observational and experimental data especially when latent or confounding variables may be present.

The focus this year is on the intersection of the complex models studied by the UAI community with the emerging challenges of Big Data. In particular, we invite papers on the following themes:

- Data preparation and cleaning

- Robustness, resistance (to errors in the data) and treatment of outliers

- Transforming data for further processing

- Prepossessing of data and use of the output of one model as the input to another (for example, using the output of natural language tools as the input to a Bayesian network)

- Data fusion: combining data of multiple types

- Treatment of missing data and selection bias

- Combining data and expert opinion

- Dealing with problems in estimating model parameters from incomplete data, or models that are not identifiable from the data

Organizers:

Russell Almond (Chair) - Florida State University

Thomas O'Neill - The American Board of Family Medicine

Marek J Druzdzel - University of Pittsburgh and Bialystok University of Technology, Poland

Julia Flores - Universidad de Castilla-La Mancha, Spain

Linda van der Gaag - Utrecht University, The Netherlands

Lionel Jouffe - Bayesia SAS

Kathryn Laskey - George Mason University

Suzanne Mahoney - Innovative Decisions, Inc.

Bayes Application Workshop II: Models for Spatial, Temporal and Network Data

http://pluto.coe.fsu.edu/UAI2013Workshops/SpatialTemporal/

Over 29 years, the annual Uncertainty in Artificial Intelligence (UAI) conference has explored complex models, many partially or fully Bayesian, which attempt to capture some of the complexities of human reasoning and decision making. As the capacity of modern computing has increased, so has the complexity of the models explored by the UAI community: complexity defined by many variables and many parameters which must be estimated from data or tuned to expert opinion. Implementing these models in practice often requires going beyond the theoretical development of the models: the difficulties arising from the practical application of UAI models has been the constant focus of the Bayesian Applications Workshop.

Due to the past success of the Bayesian Applications Workshop, there will this year for the first time be two workshops with an applications theme. The theme of this applications workshop is spatial, temporal, and network data. One example of such data is the following. We are interested in thinking about the uncertainty in "mobile data," which may come from a GPS-enabled phone or a car. In mobile applications, one important aspect is the uncertainty associated with modeling things in the context of space and time as you are moving. In automotive, for example - you are moving, but with constraints, and these constraints impact your "search space." You are typically only looking for potential destinations in the cone of where you are traveling towards. There is also a time window over which certain goals are relevant and the goodness of a goal changes as you move. The social network of the driver and the passengers could also play a role.

Another example of spatial, temporal, and network data is this: A scientist or an engineer (at ESA, NASA, USGS, or elsewhere) develops a probabilistic model in the form of a Bayesian network or a Markov random field. This is typically a declarative model of the domain - for example of earth fault motion due to earthquakes. A computer scientist or computer engineer is then concerned about how to efficiently compile and execute the model in order to compute posterior distributions or estimate parameters on a multi-core CPU, a GPU, a Hadoop cluster, a supercomputer, or another computer architecture. How well has this model worked in different domains, what are current challenges and opportunities?

The focus this workshop is on models that deal with spatial, temporal, and network data as studied by the UAI community. In particular, we invite papers on the following themes:

- Recommendation, plan recognition, and link prediction under uncertainty

- Role of user interfaces and user interaction, visualization, speech, dialogue management, etc.

- Combining data and expert knowledge in models

- Data fusion: combination of data of different types, including interaction between spatial, temporal, and network data

- Hardware and software platforms for handling complex models and large data sets

- Integration of data collected over time and space while the state of the system may also be changing (for example, a driver's goals may change as the car moves)

- Support for hard or soft real-time response; safety and security

- Integration with techniques from other disciplines, such as aerospace, automotive, biology, computer networking, earth science, ecology, education, electrical engineering, feedback control, medicine, software engineering, etc.

- Handling of missing or incomplete data, including models that are not identifiable

- Affective, emotional, context-aware, and considerate user interfaces and user interactions

- Applications in automotive, aerospace, smart phones, mobility, electrical power networks, smart grid, social networks, ecology, medicine, earth sciences, etc.

- System health management, including diagnosis, prognosis, and detection

Note that the workshop is not restricted to a particular vertical market or discipline. Instead, the workshop seeks to cross-fertilize and inspire across disciplines, with a focus on how issues related to modeling of spatial, temporal, and network data is handled.

Organizers:

Ole J. Mengshoel (Chair) - CMU

Dennis Buede - Innovative Decisions

Asela Gunawardana - Microsoft

Jennifer Healey - Intel

Oscar Kipersztok - Boeing

Branislav Kveton - Technicolor

Helge Langseth - NTNU

Tomas Singliar - Boeing

Enrique Sucar - INAOE

Tom Walsh - MIT

Workshop 3: Approaches to Causal Structure Learning Workshop

`http://www.statslab.cam.ac.uk/~rje42/uai13/main.htm`

Causality is central to how we view and react to the world around us, to our decision making, and to the advancement of science. Causal inference in statistics and machine learning has advanced rapidly in the last 20 years, leading to a plethora of new methods. However, a side-effect of the increased sophistication of these approaches is that they have grown apart, rather than together.

The aim of this workshop is to bring together researchers interested in the challenges of causal structure learning from observational and experimental data especially when latent or confounding variables may be present. Topics related to causal structure learning will be explored through a set of invited talks, presentations and a poster session.

Organizers:

Robin Evans (Chair) - University of Cambridge

Marloes Maathuis - ETH Zurich

Thomas Richardson - University of Washington

Ilya Shpitser - University of Southampton

Jin Tian - Iowa State University

Workshop 4: New Challenges in E-Commerce Recommendations

http://amazonml.com/event/uai2013-workshop/

The "New Challenges in e-Commerce Product Recommendations" Workshop solicits submissions in the area of e-commerce product recommendations. Over the past decade, researchers in recommender systems have focused on algorithms such as matrix factorization and their application to relatively static and long-lived content catalogs such as movies. However, the continued surge of e-Commerce has surfaced a lot of new challenges that directly influence the design of algorithms. Our desire is to foster a discussion on this topic by bringing together industry leaders who have developed these experiences and connecting them with researchers in the field of Machine Learning as well as to broaden the areas of research in this space.

Recommending products in e-Commerce poses some unique scientific challenges that we would like to discuss in this workshop:

- Hybrid Discovery Experiences – Customers use a combination of search, browse and recommendations to find products. Today, these experiences are mostly siloed. Hybrid experiences such as personalized search and personalized browse offer new ways to improve the overall product discovery experience.

- Context Awareness – Understanding the user's intent or their need is very important. Today, recommendations are made in many contexts such as item page, checkout page, search result page, or even within the consumption experience such as at the end of an e-book or movie.

- Breadth of Selection – e-Commerce spans a wide and growing range of products from movies and music to books, games, electronics, apparel, groceries, automotive, industrial supplies, pet supplies, home and garden, sporting goods etc.

- Mobile – The mobile setting allows novel discovery experiences on phones and tablets which are challenged by small surface area.

- Social – In today's world, data from social networks are easily accessible to recommender systems. This provides additional signals to personalize product discovery as well as leverage homophily.

- Evaluation – In e-Commerce, root-mean-squared-error (RMSE) or mean-reciprocal rank are only proxies for evaluating customer satisfaction. Also, offline and online evaluation can vary widely.

- Sparsity/Cold Start – People watch movies and listen to music often, but they don't buy a juicer that often.

- Age Progression – Recommendations for kids, students and parents need to be age aware. Parents who purchased size 2 diapers a few months back cannot be recommended more size 2 diapers after a while. Same with textbooks for students.

- Product Meta Data – Understanding product compatibility is very important for recommendations in categories such as electronics and automotive. Unfortunately such metadata is rarely available or is noisy.

- Periodicity – Some products get consumed regularly (e.g., food) and the value of recommendations is also in discovering renewal patterns.

- Free content – With the surge of free digital content–books, music, movies, apps and games–recommender systems need to find the right balance between free vs. paid content. The long-term value of free content is important to quantify.

Organizers:

Srikanth Thirumalai – Director of Personalization, Amazon.com

Ralf Herbrich – Director of Machine Learning Science, Amazon.com

Paper Abstracts

185: Generative Multiple-Instance Learning Models For Quantitative Electromyography

Tameem Adel, Ruth Urner, Benn Smith, Daniel Stashuk, Dan Lizotte

We present a comprehensive study of the use of generative modeling approaches for Multiple-Instance Learning (MIL) problems. In MIL a learner receives training instances grouped together into bags with labels for the bags only (which might not be correct for the comprised instances). Our work was motivated by the task of facilitating the diagnosis of neuromuscular disorders using sets of motor unit potential trains (MUPTs) detected within a muscle which can be cast as a MIL problem. Our approach leads to a state-of-the-art solution to the problem of muscle classification. By introducing and analyzing generative models for MIL in a general framework and examining a variety of model structures and components, our work also serves as a methodological guide to modelling MIL tasks. We evaluate our proposed methods both on MUPT datasets and on the MUSK1 dataset, one of the most widely used benchmarks for MIL.

85: Active Sensing as Bayes-Optimal Sequential Decision Making

Sheeraz Ahmad, Angela Yu

Sensory inference under conditions of uncertainty is a major problem in both machine learning and computational neuroscience. An important but poorly understood aspect of sensory processing is the role of active sensing. Here, we present a Bayes-optimal inference and control framework for active sensing, C-DAC (Context-Dependent Active Controller). Unlike previously proposed algorithms that optimize abstract statistical objectives such as information maximization (Infomax) [Butko & Movellan, 2010] or one-step look-ahead accuracy [Najemnik & Geisler, 2005], our active sensing model directly minimizes a combination of behavioral costs, such as temporal delay, response error, and effort. We simulate these algorithms on a simple visual search task to illustrate scenarios in which context-sensitivity is particularly beneficial and optimization with respect to generic statistical objectives particularly inadequate. Motivated by the geometric properties of the C-DAC policy, we present both parametric and non-parametric approximations, which retain context-sensitivity while significantly reducing computational complexity. These approximations enable us to investigate the more complex problem involving peripheral vision, and we notice that the difference between C-DAC and statistical policies becomes even more evident in this scenario.

148: The Bregman Variational Dual-Tree Framework

Saeed Amizadeh, Bo Thiesson, Milos Hauskrecht

Graph-based methods provide a powerful tool set for many non-parametric frameworks in Machine Learning. In general, the memory and computational complexity of these methods is quadratic in the number of examples in the data which makes them quickly infeasible for moderate to large scale datasets. A significant effort to find more efficient solutions to the problem has been made in the literature. One of the state-of-the-art methods that has been recently introduced is the Variational Dual-Tree (VDT) framework. Despite some of its unique features, VDT is currently restricted only to Euclidean spaces where the Euclidean distance quantifies the similarity. In this paper, we extend the VDT framework beyond the Euclidean distance to more general Bregman divergences that include the Euclidean distance as a special case. By exploiting the properties of the general Bregman divergence, we show how the new framework can maintain all the pivotal features of the VDT framework and yet

significantly improve its performance in non-Euclidean domains. We apply the proposed framework to different text categorization problems and demonstrate its benefits over the original VDT.

176: Hinge-loss Markov Random Fields: Convex Inference for Structured Prediction

Stephen Bach, Bert Huang, Ben London, Lise Getoor

Graphical models for structured domains are powerful tools, but the computational complexities of combinatorial prediction spaces can force restrictions on models, or require approximate inference in order to be tractable. Instead of working in a combinatorial space, we use hinge-loss Markov random fields (HL-MRFs), an expressive class of graphical models with log-concave density functions over continuous variables, which can represent confidences in discrete predictions. This paper demonstrates that HL-MRFs are general tools for fast and accurate structured prediction. We introduce the first inference algorithm that is both scalable and applicable to the full class of HL-MRFs, and show how to train HL-MRFs with several learning algorithms. Our experiments show that HL-MRFs match or surpass the predictive performance of state-of-the-art methods, including discrete models, in four application domains.

146: High-dimensional Joint Sparsity Random Effects Model for Multi-task Learning

Krishnakumar Balasubramanian, Kai Yu, Tong Zhang

Joint sparsity regularization in multi-task learning has attracted much attention in recent years. The traditional convex formulation employs the group Lasso relaxation to achieve joint sparsity across tasks. Although this approach leads to a simple convex formulation, it suffers from several issues due to the looseness of the relaxation. To remedy this problem, we view jointly sparse multi-task learning as a specialized random effects model, and derive a convex relaxation approach that involves two steps. The first step learns the covariance matrix of the coefficients using a convex formulation which we refer to as sparse covariance coding; the second step solves a ridge regression problem with a sparse quadratic regularizer based on the covariance matrix obtained in the first step. It is shown that this approach produces an asymptotically optimal quadratic regularizer in the multitask learning setting when the number of tasks approaches infinity. Experimental results demonstrate that the convex formulation obtained via the proposed model significantly outperforms group Lasso (and related multi-stage formulations).

165: Lower Bounds for Exact Model Counting and Applications in Probabilistic Databases

Paul Beame, Jerry Li, Sudeepa Roy, Dan Suciu

The best current methods for exactly computing the number of satisfying assignments, or the satisfying probability, of Boolean formulas can be seen, either directly or indirectly, as building 'decision-DNNF' (decision decomposable negation normal form) representations of the input Boolean formulas. Decision-DNNFs are a special case of 'd-DNNF's where 'd' stands for 'deterministic'. We show that any decision-DNNF can be converted into an equivalent 'FBDD' (free binary decision diagram) – also known as a 'read-once branching program' (ROBP or 1-BP) – with only a quasipolynomial increase in representation size in general, and with only a polynomial increase in size in the special case of monotone k-DNF formulas. Leveraging known exponential lower bounds for FBDDs, we then obtain similar exponential lower bounds for decision-DNNFs which provide lower bounds for the recent algorithms. We also separate the power of decision-DNNFs from d-DNNFs and a generalization of decision-DNNFs known as AND-FBDDs. Finally we show how these imply exponential lower bounds for natural problems associated with probabilistic databases.

37: Reasoning about Probabilities in Dynamic Systems using Goal Regression

Vaishak Belle, Hector Levesque

Reasoning about degrees of belief in uncertain dynamic worlds is fundamental to many applications, such as robotics and planning, where actions modify state properties and sensors provide measurements, both of which are prone to noise. With the exception of limited cases such as Gaussian processes over linear phenomena, belief state evolution can be complex and hard to reason with in a general way. This paper proposes a framework with new results that allows the reduction of subjective probabilities after sensing and acting to questions about the initial state only. We build on an expressive probabilistic first-order logical account by Bacchus, Halpern and Levesque, resulting in a methodology that, in principle, can be coupled with a variety of existing inference solutions.

225: Probabilistic Conditional Preference Networks

Damien Bigot, Bruno Zanuttini, Helene Fargier, Jérôme Mengin

In order to represent the preferences of a group of individuals, we introduce Probabilistic CP-nets (PCP-nets). PCP-nets provide a compact language for representing probability distributions over preference orderings. We argue that they are useful for aggregating preferences or modelling noisy preferences. Then we give efficient algorithms for the main reasoning problems, namely for computing the probability that a given outcome is preferred to another one, and the probability that a given outcome is optimal. As a by-product, we obtain an unexpected linear-time algorithm for checking dominance in a standard, tree-structured CP-net.

135: Boosting in the presence of label noise

Jakramate Bootkrajang, Ata Kaban

Boosting is known to be sensitive to label noise. We studied two approaches to improve AdaBoost's robustness against labelling errors. One is to employ a label-noise robust classifier as a base learner, while the other is to modify the AdaBoost algorithm to be more robust. Empirical evaluation shows that a committee of robust classifiers, although converges faster than non label-noise aware AdaBoost, is still susceptible to label noise. However, pairing it with the new robust Boosting algorithm we propose here results in a more resilient algorithm under mislabelling.

127: Hilbert Space Embeddings of Predictive State Representations

Byron Boots, Geoffrey Gordon, Arthur Gretton

Predictive State Representations (PSRs) are an expressive class of models for controlled stochastic processes. PSRs represent state as a set of predictions of future observable events. Because PSRs are defined entirely in terms of observable data, statistically consistent estimates of PSR parameters can be learned efficiently by manipulating moments of observed training data. Most learning algorithms for PSRs have assumed that actions and observations are finite with low cardinality. In this paper, we generalize PSRs to infinite sets of observations and actions, using the recent concept of Hilbert space embeddings of distributions. The essence is to represent the state as a nonparametric conditional embedding operator in a Reproducing Kernel Hilbert Space (RKHS) and leverage recent work in kernel methods to estimate, predict, and update the representation. We show that these Hilbert space embeddings of PSRs are able to gracefully handle continuous actions and observations, and that our learned models outperform competing system identification algorithms on several prediction benchmarks.

47: Scoring and Searching over Bayesian Networks with Causal and Associative Priors

Giorgos Borboudakis, Ioannis Tsamardinos

A significant theoretical advantage of search-and-score methods for learning Bayesian Networks is that they can accept informative prior beliefs for each possible network, thus complementing the data. In this paper, a method is presented for assigning priors based on beliefs on the presence or absence of certain paths in the true network. Such beliefs correspond to knowledge about the possible causal and associative relations between pairs of variables. This type of knowledge naturally arises from prior experimental and observational data, among others. In addition, a novel search-operator is proposed to take advantage of such prior knowledge. Experiments show that, using path beliefs improves the learning of the skeleton, as well as the edge directions in the network.

30: SparsityBoost: A New Scoring Function for Learning Bayesian Network Structure

Eliot Brenner, David Sontag

We give a new consistent scoring function for structure learning of Bayesian networks. In contrast to traditional approaches to scorebased structure learning, such as BDeu or MDL, the complexity penalty that we propose is data-dependent and is given by the probability that a conditional independence test correctly shows that an edge cannot exist. What really distinguishes this new scoring function from earlier work is that it has the property of becoming computationally easier to maximize as the amount of data increases. We prove a polynomial sample complexity result, showing that maximizing this score is guaranteed to correctly learn a structure with no false edges and a distribution close to the generating distribution, whenever there exists a Bayesian network which is a perfect map for the data generating distribution. Although the new score can be used with any search algorithm, we give empirical results showing that it is particularly effective when used together with a linear programming relaxation approach to Bayesian network structure learning.

188: Sample Complexity of Multi-task Reinforcement Learning

Emma Brunskill, Lihong Li

Transferring knowledge across a sequence of reinforcement-learning tasks is challenging, and has a number of important applications. Though there is encouraging empirical evidence that transfer can improve performance in subsequent reinforcement-learning tasks, there has been very little theoretical analysis. In this paper, we introduce a new multi-task algorithm for a sequence of reinforcement-learning tasks when each task is sampled independently from (an unknown) distribution over a finite set of Markov decision processes whose parameters are initially unknown. For this setting, we prove under certain assumptions that the per-task sample complexity of exploration is reduced significantly due to transfer compared to standard single-task algorithms. Our multi-task algorithm also has the desired characteristic that it is guaranteed not to exhibit negative transfer: in the worst case its per-task sample complexity is comparable to the corresponding single-task algorithm.

158: Automorphism Groups of Graphical Models and Lifted Variational Inference

Hung Bui, Tuyen Huynh, Sebasitan Riedel

Using the theory of group action, we first introduce the concept of the automorphism group of an exponential family or a graphical model, thus formalizing the general notion of symmetry of a probabilistic model. This automorphism group provides a precise mathematical framework for lifted inference in the general exponential family. Its group action partitions the set of random variables and feature functions into equivalent classes (called orbits) having identical marginals and expectations. Then the inference problem is effectively reduced to that of computing marginals or expectations for each class, thus avoiding the need to deal with each individual

variable or feature. We demonstrate the usefulness of this general framework in lifting two classes of variational approximation for maximum a posteriori (MAP) inference: local linear programming (LP) relaxation and local LP relaxation with cycle constraints; the latter yields the first lifted variational inference algorithm that operates on a bound tighter than the local constraints.

31: POMDPs under Probabilistic Semantics

Krishnendu Chatterjee, Martin Chmelík

We consider partially observable Markov decision processes (POMDPs) with limit-average payoff, where a reward value in the interval $[0, 1]$ is associated to every transition, and the payoff of an infinite path is the long-run average of the rewards. We consider two types of path constraints: (i) quantitative constraint defines the set of paths where the payoff is at least a given threshold $\lambda_1 \in (0, 1]$; and (ii) qualitative constraint which is a special case of quantitative constraint with $\lambda_1 = 1$. We consider the computation of the almost-sure winning set, where the controller needs to ensure that the path constraint is satisfied with probability 1. Our main results for qualitative path constraint are as follows: (i) the problem of deciding the existence of a finite-memory controller is EXPTIME-complete; and (ii) the problem of deciding the existence of an infinite-memory controller is undecidable. For quantitative path constraint we show that the problem of deciding the existence of a finite-memory controller is undecidable.

134: Parallel Gaussian Process Regression with Low-Rank Covariance Matrix Approximations

Jie Chen, Nannan Cao, Kian Hsiang Low, Ruofei Ouyang, Colin Keng-Yan Tan, Patrick Jaillet

Gaussian processes (GP) are Bayesian non-parametric models that are widely used for probabilistic regression. Unfortunately, it cannot scale well with large data nor perform real-time predictions due to its cubic time cost in the data size. This paper presents two parallel GP regression methods that exploit low-rank covariance matrix approximations for distributing the computational load among parallel machines to achieve time efficiency and scalability. We theoretically guarantee the predictive performances of our proposed parallel GPs to be equivalent to that of some centralized approximate GP regression methods: The computation of their centralized counterparts can be distributed among parallel machines, hence achieving greater time efficiency and scalability. We analytically compare the properties of our parallel GPs such as time, space, and communication complexity. Empirical evaluation on two real-world datasets in a cluster of 20 computing nodes shows that our parallel GPs are significantly more time-efficient and scalable than their centralized counterparts and exact/full GP while achieving predictive performances comparable to full GP.

8: Convex Relaxations of Bregman Divergence Clustering

Hao Cheng, Xinhua Zhang, Dale Schuurmans

Although many convex relaxations of clustering have been proposed in the past decade, current formulations remain restricted to spherical Gaussian or discriminative models and are susceptible to imbalanced clusters. To address these shortcomings, we propose a new class of convex relaxations that can be flexibly applied to more general forms of Bregman divergence clustering. By basing these new formulations on normalized equivalence relations we retain additional control on relaxation quality, which allows improvement in clustering quality. We furthermore develop optimization methods that improve scalability by exploiting recent implicit matrix norm methods. In practice, we find that the new formulations are able to efficiently produce tighter clusterings that improve the accuracy of state of the art methods.

121: Learning Sparse Causal Models is not NP-hard

Tom Claassen, Joris Mooij, Tom Heskes

This paper shows that causal model discovery is not an NP-hard problem, in the sense that for sparse graphs bounded by node degree k the sound and complete causal model can be obtained in worst case order $N^{2(k+2)}$ independence tests, even when latent variables and selection bias may be present. We present a modification of the well-known FCI algorithm that implements the method for an independence oracle, and suggest improvements for sample/real-world data versions. It does not contradict any known hardness results, and does not solve an NP-hard problem: it just proves that sparse causal discovery is perhaps more complicated, but not as hard as learning minimal Bayesian networks.

18: Advances in Bayesian Network Learning using Integer Programming

James Cussens, Mark Bartlett

We consider the problem of learning Bayesian networks (BNs) from complete discrete data. This problem of discrete optimisation is formulated as an integer program (IP). We describe the various steps we have taken to allow efficient solving of this IP. These are (i) efficient search for cutting planes, (ii) a fast greedy algorithm to find high-scoring (perhaps not optimal) BNs and (iii) tightening the linear relaxation of the IP. After relating this BN learning problem to set covering and the multidimensional 0-1 knapsack problem, we present our empirical results. These show improvements, sometimes dramatic, over earlier results.

240: Qualitative Possibilistic Mixed-Observable MDPs

Nicolas Drougard, Florent Teichteil-Königsbuch, Jean-Loup Farges, Didier Dubois

Possibilistic and qualitative POMDPs (π-POMDPs) are counterparts of POMDPs used to model situations where the agent's initial belief or observation probabilities are imprecise due to lack of past experiences or insufficient data collection. However, like probabilistic POMDPs, optimally solving π-POMDPs is intractable: the finite belief state space exponentially grows with the number of system's states. In this paper, a possibilistic version of Mixed-Observable MDPs is presented to get around this issue: the complexity of solving π-POMDPs, some state variables of which are fully observable, can be then dramatically reduced. A value iteration algorithm for this new formulation under infinite horizon is next proposed and the optimality of the returned policy (for a specified criterion) is shown assuming the existence of a "stay" action in some goal states. Experimental work finally shows that this possibilistic model outperforms probabilistic POMDPs commonly used in robotics, for a target recognition problem where the agent's observations are imprecise.

117: Optimization With Parity Constraints: From Binary Codes to Discrete Integration

Stefano Ermon, Carla Gomes, Ashish Sabharwal, Bart Selman

Many probabilistic inference tasks involve summations over exponentially large sets. Recently, it has been shown that these problems can be reduced to solving a polynomial number of MAP inference queries for a model augmented with randomly generated parity constraints. By exploiting a connection with max-likelihood decoding of binary codes, we show that these optimizations are computationally hard. Inspired by iterative message passing decoding algorithms, we propose an Integer Linear Programming (ILP) formulation for the problem, enhanced with new sparsification techniques to improve decoding performance. By solving the ILP through a sequence of LP relaxations, we get both lower and upper bounds on the partition function, which hold with high probability and are much tighter than those obtained with variational methods.

144: Monte-Carlo Planning: Theoretically Fast Convergence Meets Practical Efficiency

Zohar Feldman, Carmel Domshlak

Popular Monte-Carlo tree search (MCTS) algorithms for online planning, such as epsilon-greedy tree search and UCT, aim at rapidly identifying a reasonably good action, but provide rather poor worst-case guarantees on performance improvement over time. In contrast, a recently introduced MCTS algorithm BRUE guarantees exponential-rate improvement over time, yet it is not geared towards identifying reasonably good choices right at the go. We take a stand on the individual strengths of these two classes of algorithms, and show how they can be effectively connected. We then rationalize a principle of "selective tree expansion", and suggest a concrete implementation of this principle within MCTS. The resulting algorithm,s favorably compete with other MCTS algorithms under short planning times, while preserving the attractive convergence properties of BRUE.

211: Bethe-ADMM for Tree Decomposition based Parallel MAP Inference

Qiang Fu, Huahua Wang, Arindam Banerjee

We consider the problem of maximum a posteriori (MAP) inference in discrete graphical models. We present a parallel MAP inference algorithm called Bethe-ADMM based on two ideas: tree-decomposition of the graph and the alternating direction method of multipliers (ADMM). However, unlike the standard ADMM, we use an inexact ADMM augmented with a Bethe-divergence based proximal function, which makes each subproblem in ADMM easy to solve in parallel using the sum-product algorithm. We rigorously prove global convergence of Bethe-ADMM. The proposed algorithm is extensively evaluated on both synthetic and real datasets to illustrate its effectiveness. Further, the parallel Bethe-ADMM is shown to scale almost linearly with increasing number of cores.

143: Building Bridges: Viewing Active Learning from the Multi-Armed Bandit Lens

Ravi Ganti, Alexander Gray

In this paper we propose a multi-armed bandit inspired, pool based active learning algorithm for the problem of binary classification. By carefully constructing an analogy between active learning and multi-armed bandits, we utilize ideas such as lower confidence bounds, and self-concordant regularization from the multi-armed bandit literature to design our proposed algorithm. Our algorithm is a sequential algorithm, which in each round assigns a sampling distribution on the pool, samples one point from this distribution, and queries the oracle for the label of this sampled point. The design of this sampling distribution is also inspired by the analogy between active learning and multi-armed bandits. We show how to derive lower confidence bounds required by our algorithm. Experimental comparisons to previously proposed active learning algorithms show superior performance on some standard UCI datasets.

139: Batch-iFDD for Representation Expansion in Large MDPs

Alborz Geramifard, Tom Walsh, Nicholas Roy, Jonathan How

Matching pursuit (MP) methods are a promising class of feature construction algorithms for value function approximation. Yet existing MP methods require creating a pool of potential features, mandating expert knowledge or enumeration of a large feature pool, both of which hinder scalability. This paper introduces batch incremental feature dependency discovery (Batch-iFDD) as an MP method that inherits a provable convergence property. Additionally, Batch-iFDD does not require a large pool of features, leading to lower computational complexity. Empirical policy evaluation results across three domains with up to one million states highlight the scalability of Batch-iFDD over the previous state of the art MP algorithm.

91: Structured Message Passing

Vibhav Gogate, Pedro Domingos

In this paper, we present structured message passing (SMP), a unifying framework for approximate inference algorithms that take advantage of structured representations such as algebraic decision diagrams and sparse hash tables. These representations can yield significant time and space savings over the conventional tabular representation when the message has several identical values (context-specific independence) or zeros (determinism) or both in its range. Therefore, in order to fully exploit the power of structured representations, we propose to artificially introduce context-specific independence and determinism in the messages. This yields a new class of powerful approximate inference algorithms which includes popular algorithms such as cluster-graph Belief propagation (BP), expectation propagation and particle BP as special cases. We show that our new algorithms introduce several interesting bias-variance trade-offs. We evaluate these trade-offs empirically and demonstrate that our new algorithms are more accurate and scalable than state-of-the-art techniques.

76: Multiple Instance Learning by Discriminative Training of Markov Networks

Hossein Hajimirsadeghi, Jinling Li, Greg Mori, Mohammad Zaki, Tarek Sayed

We introduce a graphical framework for multiple instance learning (MIL) based on Markov networks. This framework can be used to model the traditional MIL definition as well as more general MIL definitions. Different levels of ambiguity – the portion of positive instances in a bag – can be explored in weakly supervised data. To train these models, we propose a discriminative max-margin learning algorithm leveraging efficient inference for cardinality-based cliques. The efficacy of the proposed framework is evaluated on a variety of data sets. Experimental results verify that encoding or learning the degree of ambiguity can improve classification performance.

54: Unsupervised Learning of Noisy-Or Bayesian Networks

Yonatan Halpern, David Sontag

This paper considers the problem of learning the parameters in Bayesian networks of discrete variables with known structure and hidden variables. Previous approaches in these settings typically use expectation maximization; when the network has high treewidth, the required expectations might be approximated using Monte Carlo or variational methods. We show how to avoid inference altogether during learning by giving a polynomial-time algorithm based on the method-of-moments, building upon recent work on learning discrete-valued mixture models. In particular, we show how to learn the parameters for a family of bipartite noisy-or Bayesian networks. In our experimental results, we demonstrate an application of our algorithm to learning QMR-DT, a large Bayesian network used for medical diagnosis. We show that it is possible to fully learn the parameters of QMR-DT even when only the findings are observed in the training data (ground truth diseases unknown).

244: Gaussian Processes for Big Data

James Hensman, Nicolo Fusi, Neil Lawrence

We introduce stochastic variational inference for Gaussian process models. This enables the application of Gaussian process (GP) models to data sets containing millions of data points. We show how GPs can be variationally decomposed to depend on a set of globally relevant inducing variables which factorize the model in the necessary manner to perform variational inference. Our ap- proach is readily extended to models with non-Gaussian likelihoods and latent variable models based around Gaussian processes. We demonstrate the approach on a simple toy problem and two real world data sets.

80: Inverse Covariance Estimation for High-Dimensional Data in Linear Time and Space: Spectral Methods for Riccati and Sparse Models

Jean Honorio, Tommi Jaakkola

We propose maximum likelihood estimation for learning Gaussian graphical models with a Gaussian (ℓ_2^2) prior on the parameters. This is in contrast to the commonly used Laplace (ℓ_1) prior for encouraging sparseness. We show that our optimization problem leads to a Riccati matrix equation, which has a closed form solution. We propose an efficient algorithm that performs a singular value decomposition of the training data. Our algorithm is $O(NT^2)$-time and $O(NT)$-space for N variables and T samples. Our method is tailored to high-dimensional problems ($N \gg T$), in which sparseness promoting methods become intractable. Furthermore, instead of obtaining a single solution for a specific regularization parameter, our algorithm finds the whole solution path. We show that the method has logarithmic sample complexity under the spiked covariance model. We also propose sparsification of the dense solution with provable performance guarantees. We provide techniques for using our learnt models, such as removing unimportant variables, computing likelihoods and conditional distributions. Finally, we show promising results in several gene expressions datasets.

131: Discovering Cyclic Causal Models with Latent Variables: A General SAT-Based Procedure

Antti Hyttinen, Patrik Hoyer, Frederick Eberhardt, Matti Järvisalo

We present a very general approach to learning the structure of causal models based on d-separation constraints, obtained from any given set of overlapping passive observational or experimental data sets. The procedure allows for both directed cycles (feedback loops) and the presence of latent variables. Our approach is based on a logical representation of causal pathways, which permits the integration of quite general background knowledge, and inference is performed using a Boolean satisfiability (SAT) solver. The procedure is complete in that it exhausts the available information on whether any given edge can be determined to be present or absent, and returns "unknown" otherwise. Many existing constraint-based causal discovery algorithms can be seen as special cases, tailored to circumstances in which one or more restricting assumptions apply. Simulations illustrate the effect of these assumptions on discovery and how the present algorithm scales.

109: Warped Mixtures for Nonparametric Cluster Shapes

Tomoharu Iwata, David Duvenaud, Zoubin Ghahramani

A mixture of Gaussians fit to a single curved or heavy-tailed cluster will report that the data contains many clusters. To produce more appropriate clusterings, we introduce a model which warps a latent mixture of Gaussians to produce nonparametric cluster shapes. The possibly low-dimensional latent mixture model allows us to summarize the properties of the high-dimensional clusters (or density manifolds) describing the data. The number of manifolds, as well as the shape and dimension of each manifold is automatically inferred. We derive a simple inference scheme for this model which analytically integrates out both the mixture parameters and the warping function. We show that our model is effective for density estimation, performs better than infinite Gaussian mixture models at recovering the true number of clusters, and produces interpretable summaries of high-dimensional datasets.

160: The Lovasz-Bregman Divergence and connections to rank aggregation, clustering, and web ranking

Rishabh Iyer, Jeff Bilmes

We extend the recently introduced theory of Lovasz-Bregman (LB) divergences (Iyer & Bilmes 2012) in several ways. We show that they represent a distortion between a "score" and an "ordering", thus providing a new view of rank aggregation and order based clustering with interesting connections to web ranking. We show how the LB divergences have a number of properties akin to many permutation based metrics, and in fact have as special cases forms very similar to the Kendall-tau metric. We also show how the LB divergences subsume a number of commonly used ranking measures in information retrieval, like NDCG and AUC. Unlike the traditional permutation based metrics, however, the LB divergence naturally captures a notion of "confidence"

in the orderings, thus providing a new representation to applications involving aggregating scores as opposed to just orderings. We show how a number of recently used web ranking models are forms of Lovasz-Bregman rank aggregation and also observe that a natural form of Mallow's model using the LB divergence has been used as conditional ranking models for the "Learning to Rank" problem.

199: Solving Limited-Memory Influence Diagrams Using Branch-and-Bound Search

Arindam Khaled, Eric Hansen, Changhe Yuan

A limited-memory influence diagram (LIMID) generalizes a traditional influence diagram by relaxing the assumptions of regularity and no-forgetting, allowing a wider range of decision problems to be modeled. Algorithms for solving traditional influence diagrams are not easily generalized to solve LIMIDs, however, and only recently have exact algorithms for solving LIMIDs been developed. In this paper, we introduce an exact algorithm for solving LIMIDs that is based on branch-and-bound search. Our approach is related to the approach of solving an influence diagram by converting it to an equivalent decision tree, with the difference that the LIMID is converted to a much smaller decision graph that can be searched more efficiently.

97: Constrained Bayesian Inference for Low Rank Multitask Learning

Oluwasanmi Koyejo, Joydeep Ghosh

We present a novel approach for constrained Bayesian inference. Unlike current methods, our approach does not require convexity of the constraint set. We reduce the constrained variational inference to a parametric optimization over the feasible set of densities and propose a general recipe for such problems. We apply the proposed constrained Bayesian inference approach to multitask learning subject to rank constraints on the weight matrix. Further, constrained parameter estimation is applied to recover the sparse conditional independence structure encoded by prior precision matrices. Our approach is motivated by reverse inference for high dimensional functional neuroimaging, a domain where the high dimensionality and small number of examples requires the use of constraints to ensure meaningful and effective models. For this application, we propose a model that jointly learns a weight matrix and the prior inverse covariance structure between different tasks. We present experimental validation showing that the proposed approach outperforms strong baseline models in terms of predictive performance and structure recovery.

88: Collective Diffusion Over Networks: Models and Inference

Akshat Kumar, Dan Sheldon, Biplav Srivastava

Diffusion processes in networks are increasingly used to model the spread of information and social influence. In several applications in computational sustainability such as the spread of wildlife, infectious diseases and traffic mobility pattern, the observed data often consists of only *aggregate* information. In this work, we present new models that generalize standard diffusion processes to such collective settings. We also present optimization based techniques that can accurately learn the underlying dynamics of the given contagion process, including the hidden network structure, by only observing the time a node becomes active and the associated aggregate information. Empirically, our technique is highly robust and accurately learns network structure with more than 90

70: Causal Transportability of Experiments on Controllable Subsets of Variables: z-Transportability

Sanghack Lee, Vasant Honavar

We introduce z-transportability, the problem of estimating the causal effect of a set of variables X on another set of variables Y in a target domain from experiments on any subset of controllable variables Z where Z is an

arbitrary subset of observable variables V in a source domain. z-Transportability generalizes z-identifiability, the problem of estimating in a given domain the causal effect of X on Y from surrogate experiments on a set of variables Z such that Z is disjoint from X;. z-Transportability also generalizes transportability which requires that the causal effect of X on Y in the target domain be estimable from experiments on any subset of all observable variables in the source domain. We first generalize z-identifiability to allow cases where Z is not necessarily disjoint from X. Then, we establish a necessary and sufficient condition for z-transportability in terms of generalized z-identifiability and transportability. We provide a correct and complete algorithm that determines whether a causal effect is z-transportable; and if it is, produces a transport formula, that is, a recipe for estimating the causal effect of X on Y in the target domain using information elicited from the results of experimental manipulations of Z in the source domain and observational data from the target domain. Our results also show that do-calculus is complete for z-transportability.

197: A Sound and Complete Algorithm for Learning Causal Models from Relational Data

Marc Maier, Katerina Marazopoulou, David Arbour, David Jensen

The PC algorithm learns maximally oriented causal Bayesian networks. However, there is no equivalent complete algorithm for learning the structure of relational models, a more expressive generalization of Bayesian networks. Recent developments in the theory and representation of relational models support lifted reasoning about conditional independence. This enables a powerful constraint for orienting bivariate dependencies and forms the basis of a new algorithm for learning structure. We present the relational causal discovery (RCD) algorithm that learns causal relational models. We prove that RCD is sound and complete, and we present empirical results that demonstrate effectiveness.

62: Evaluating Anytime Algorithms for Learning Optimal Bayesian Networks

Brandon Malone, Changhe Yuan

Exact algorithms for learning Bayesian networks guarantee to find provably optimal networks. However, they may fail in difficult learning tasks due to limited time or memory. In this research we adapt several anytime heuristic search-based algorithms to learn Bayesian networks. These algorithms find high-quality solutions quickly, and continually improve the incumbent solution or prove its optimality before resources are exhausted. Empirical results show that the anytime window A* algorithm usually finds higher-quality, often optimal, networks more quickly than other approaches. The results also show that, surprisingly, while generating networks with few parents per variable are structurally simpler, they are harder to learn than complex generating networks with more parents per variable.

142: On the Complexity of Strong and Epistemic Credal Networks

Denis Maua, Cassio de Campos, Alessio Benavoli, Alessandro Antonucci

Credal networks are graph-based statistical models whose parameters take values in a set, instead of being sharply specified as in traditional statistical models (e.g., Bayesian networks). The computational complexity of inferences on such models depends on the irrelevance/independence concept adopted. In this paper, we study inferential complexity under the concepts of epistemic irrelevance and strong independence. We show that inferences under strong independence are NP-hard even in trees with ternary variables. We prove that under epistemic irrelevance the polynomial time complexity of inferences in credal trees is not likely to extend to more general models (e.g. singly connected networks). These results clearly distinguish networks that admit efficient inferences and those where inferences are most likely hard, and settle several open questions regarding computational complexity.

154: Learning Periodic Human Behaviour Models from Sparse Data for Crowd-sourcing Aid Delivery in Developing Countries

James McInerney, Alex Rogers, Nicholas Jennings

In many developing countries, half the population lives in rural locations, where access to essentials such as school materials, mosquito nets, and medical supplies is restricted. We propose an alternative method of distribution (to standard road delivery) in which the existing mobility habits of a local population are leveraged to deliver aid, which raises two technical challenges in the areas optimisation and learning. For optimisation, a standard Markov decision process applied to this problem is intractable, so we provide an exact formulation that takes advantage of the periodicities in human location behaviour. To learn such behaviour models from sparse data (i.e., cell tower observations), we develop a Bayesian model of human mobility. Using real cell tower data of the mobility behaviour of 50,000 individuals in Ivory Coast, we find that our model outperforms the state of the art approaches in mobility prediction by at least 25

66: Learning Max-Margin Tree Predictors

Ofer Meshi, Elad Eban, Gal Elidan, Amir Globerson

Structured prediction is a powerful framework for coping with joint prediction of interacting outputs. A central difficulty in using this framework is that often the correct label dependence structure is unknown. At the same time, we would like to avoid an overly complex structure that will lead to intractable prediction. In this work we address the challenge of learning tree structured predictive models that achieve high accuracy while at the same time facilitate efficient (linear time) inference. We start by proving that this task is in general NP-hard, and then suggest an approximate alternative. Briefly, our CRANK approach relies on a novel Circuit-RANK regularizer that penalizes non-tree structures and that can be optimized using a CCCP procedure. We demonstrate the effectiveness of our approach on several domains and show that, despite the relative simplicity of the structure, prediction accuracy is competitive with a fully connected model that is computationally costly at prediction time.

103: Tighter Linear Program Relaxations for High Order Graphical Models

Elad Mezuman, Daniel Tarlow, Amir Globerson, Yair Weiss

Graphical models with High Order Potentials (HOPs) have received considerable interest in recent years. While there are a variety of approaches to inference in these models, nearly all of them amount to solving a linear program (LP) relaxation with unary consistency constraints between the HOP and the individual variables. In many cases, the resulting relaxations are loose, and in these cases the results of inference can be poor. It is thus desirable to look for more accurate ways of performing inference in these models. In this work, we study the LP relaxations that result from enforcing additional consistency constraints between the HOP and the rest of the model. We address theoretical questions about the strength of the resulting relaxations compared to the relaxations that arise in standard approaches, and we develop practical and efficient message passing algorithms for optimizing the LPs. Empirically, we show that the LPs with additional consistency constraints lead to more accurate inference on some challenging problems that include a combination of low order and high order terms.

23: Cyclic Causal Discovery from Continuous Equilibrium Data

Joris Mooij, Tom Heskes

We propose a method for learning cyclic causal models from a combination of observational and interventional equilibrium data. Novel aspects of the proposed method are its ability to work with continuous data (without assuming linearity) and to deal with feedback loops. Within the context of biochemical reactions, we also propose a novel way of modeling interventions that modify the activity of compounds instead of their abundance. For

computational reasons, we approximate the nonlinear causal mechanisms by (coupled) local linearizations, one for each experimental condition. We apply the method to reconstruct a cellular signaling network from the flow cytometry data measured by Sachs et al. (2005). We show that our method finds evidence in the data for feedback loops and that it gives a more accurate quantitative description of the data at comparable model complexity.

24: From Ordinary Differential Equations to Structural Causal Models: the deterministic case

Joris Mooij, Dominik Janzing, Bernhard Schölkopf

We show how, and under which conditions, the equilibrium states of a first-order Ordinary Differential Equation (ODE) system can be described with a deterministic Structural Causal Model (SCM). Our exposition sheds more light on the concept of causality as expressed within the framework of Structural Causal Models, especially for cyclic models.

4: One-Class Support Measure Machines for Group Anomaly Detection

Krikamol Muandet, Bernhard Schölkopf

We propose one-class support measure machines (OCSMMs) for group anomaly detection which aims at recognizing anomalous aggregate behaviors of data points. The OCSMMs generalize well-known one-class support vector machines (OCSVMs) to a space of probability measures. By formulating the problem as quantile estimation on distributions, we can establish an interesting connection to the OCSVMs and variable kernel density estimators (VKDEs) over the input space on which the distributions are defined, bridging the gap between large-margin methods and kernel density estimators. In particular, we show that various types of VKDEs can be considered as solutions to a class of regularization problems studied in this paper. Experiments on Sloan Digital Sky Survey dataset and High Energy Particle Physics dataset demonstrate the benefits of the proposed framework in real-world applications.

151: Structured Convex Optimization under Submodular Constraints

Kiyohito Nagano, Yoshinobu Kawahara

A number of discrete and continuous optimization problems in machine learning are related to convex minimization problems under submodular constraints. In this paper, we deal with a submodular function with a directed graph structure, and we show that a wide range of convex optimization problems under submodular constraints can be solved much more efficiently than general submodular optimization methods by a reduction to a maximum flow problem. Furthermore, we give some applications, including sparse optimization methods, in which the proposed methods are effective. Additionally, we evaluate the performance of the proposed method through computational experiments.

219: Treedy: A Heuristic for Counting and Sampling Subsets

Teppo Niinimki, Mikko Koivisto

Consider a collection of weighted subsets of a ground set N. Given a query subset Q of N, how fast can one (1) find the weighted sum over all subsets of Q, and (2) sample a subset of Q proportionally to the weights? We present a tree-based greedy heuristic, Treedy, that for a given positive tolerance d answers such counting and sampling queries to within a guaranteed relative error d and total variation distance d, respectively. Experimental results on artificial instances and in application to Bayesian structure discovery in Bayesian networks show that approximations yield dramatic savings in running time compared to exact computation, and that Treedy typically outperforms a previously proposed sorting-based heuristic.

132: Stochastic Rank Aggregation

Shuzi Niu, Yanyan Lan, Jiafeng Guo, Xueqi Cheng

This paper addresses the problem of rank aggregation, which aims to find a consensus ranking among multiple ranking inputs. Traditional rank aggregation methods are deterministic, and can be categorized into explicit and implicit methods depending on whether rank information is explicitly or implicitly utilized. Surprisingly, experimental results on real data sets show that explicit rank aggregation methods would not work as well as implicit methods, although rank information is critical for the task. Our analysis indicates that the major reason might be the unreliable rank information from incomplete ranking inputs. To solve this problem, we propose to incorporate uncertainty into rank aggregation and tackle the problem in both unsupervised and supervised scenario. We call this novel framework *stochastic rank aggregation* (St.Agg for short). Specifically, we introduce a prior distribution on ranks, and transform the ranking functions or objectives in traditional explicit methods to their expectations over this distribution. Our experiments on benchmark data sets show that the proposed St.Agg outperforms the baselines in both unsupervised and supervised scenarios.

93: Pay or Play

Sigal Oren, Michael Schapira, Moshe Tennenholtz

We introduce the class of pay or play games, which captures scenarios in which each decision maker is faced with a choice between two actions: one with a fixed payoff and an- other with a payoff dependent on others selected actions. This is, arguably, the simplest setting that models selection among certain and uncertain outcomes in a multi-agent system. We study the properties of equilibria in such games from both a game-theoretic perspective and a computational perspective. Our main positive result establishes the existence of a semi-strong equilibrium in every such game. We show that although simple, pay of play games contain a large variety of well-studied environments, e.g., vaccination games. We discuss the interesting implications of our results for these environments.

247: Evaluating computational models of explanation using human judgments

Michael Pacer, Joseph Williams, Xi Chen, Tania Lombrozo, Thomas Griffiths

We evaluate four computational models of explanation in Bayesian networks by comparing model predictions to human judgments. In two experiments, we present human participants with causal structures for which the models make divergent predictions and either solicit the best explanation for an observed event (Experiment 1) or have participants rate provided explanations for an observed event (Experiment 2). Across two versions of two causal structures and across both experiments, we find that the Causal Explanation Tree and Most Relevant Explanation models provide better fits to human data than either Most Probable Explanation or Explanation Tree models. We identify strengths and shortcomings of these models and what they can reveal about human explanation. We conclude by suggesting the value of pursuing computational and psychological investigations of explanation in parallel.

208: Approximation of Lorenz-Optimal Solutions in Multiobjective Markov Decision Processes

Patrice Perny, Paul Weng, Judy Goldsmith, Josiah Hanna

This paper is devoted to fair optimization in Multiobjective Markov Decision Processes (MOMDPs). A MOMDP is an extension of the MDP model for planning under uncertainty while trying to optimize several reward functions simultaneously. This applies to multiagent problems when rewards define individual utility functions, or in multicriteria problems when rewards refer to different features. In this setting, we study the determination of policies leading to Lorenz-non-dominated tradeoffs. Lorenz dominance is a refinement of Pareto dominance

that was introduced in Social Choice for the measurement of inequalities. In this paper, we introduce methods to efficiently approximate the sets of Lorenz-non-dominated solutions of infinite-horizon, discounted MOMDPs. The approximations are polynomial-sized subsets of those solutions.

167: Solution Methods for Constrained Markov Decision Process with Continuous Probability Modulation

Marek Petrik, Dharmashankar Subramanian, Janusz Marecki

We propose solution methods for previously-unsolved constrained MDPs in which actions can continuously modify the transition probabilities within some acceptable sets. While many methods have been proposed to solve regular MDPs with large state sets, there are few practical approaches for solving constrained MDPs with large action sets. In particular, we show that the continuous action sets can be replaced by their extreme points when the rewards are linear in the modulation. We also develop a tractable optimization formulation for concave reward functions and, surprisingly, also extend it to non- concave reward functions by using their concave envelopes. We evaluate the effectiveness of the approach on the problem of managing delinquencies in a portfolio of loans.

118: The Supervised IBP: Neighbourhood Preserving Infinite Latent Feature Models

Novi Quadrianto, Viktoriia Sharmanska, David Knowles, Zoubin Ghahramani

We propose a probabilistic model to infer supervised latent variables in the Hamming space from observed data. Our model allows simultaneous inference of the number of binary latent variables, and their values. The latent variables preserve neighbourhood structure of the data in a sense that objects in the same semantic concept have similar latent values, and objects in different concepts have dissimilar latent values. We formulate the supervised infinite latent variable problem based on an intuitive principle of pulling objects together if they are of the same type, and pushing them apart if they are not. We then combine this principle with a flexible Indian Buffet Process prior on the latent variables. We show that the inferred supervised latent variables can be directly used to perform a nearest neighbour search for the purpose of retrieval. We introduce a new application of dynamically extending hash codes, and show how to effectively couple the structure of the hash codes with continuously growing structure of the neighbourhood preserving infinite latent feature space.

192: Normalized Online Learning

Stephane Ross, Paul Mineiro, John Langford

We introduce online learning algorithms which are independent of feature scales, proving regret bounds dependent on the ratio of scales existent in the data rather than the absolute scale. This has several useful effects: there is no need to pre-normalize data, the test-time and test-space complexity are reduced, and the algorithms are more robust.

21: Beyond Log-Supermodularity: Lower Bounds and the Bethe Partition Function

Nicholas Ruozzi

A recent result has demonstrated that the Bethe partition function always lower bounds the true partition function of binary, log-supermodular graphical models. We demonstrate that these results can be extended to other interesting classes of graphical models that are not necessarily binary or log-supermodular: the ferromagnetic Potts model with a uniform external field and its generalizations and special classes of weighted graph homomorphism problems.

110: Identifying Finite Mixtures of Nonparametric Product Distributions and Causal Inference of Confounders

Eleni Sgouritsa, Dominik Janzing, Jonas Peters, Bernhard Schölkopf

We propose a kernel method to identify finite mixtures of nonparametric product distributions. It is based on a Hilbert space embedding of the joint distribution. The rank of the constructed tensor is equal to the number of mixture components. We present an algorithm to recover the components by partitioning the data points into clusters such that the variables are jointly conditionally independent given the cluster. This method can be used to identify finite confounders.

200: Determinantal Clustering Processes - A Nonparametric Bayesian Approach to Kernel Based Semi-Supervised Clustering

Amar Shah, Zoubin Ghahramani

Semi-supervised clustering is the task of clustering data points into clusters where only a fraction of the points are labelled. The true number of clusters in the data is often unknown and most models require this parameter as an input. Dirichlet process mixture models are appealing as they can infer the number of clusters from the data. However, these models do not deal with high dimensional data well and can encounter difficulties in inference. We present a novel nonparameteric Bayesian kernel based method to cluster data points without the need to prespecify the number of clusters or to model complicated densities from which data points are assumed to be generated from. The key insight is to use determinants of submatrices of a kernel matrix as a measure of how close together a set of points are. We explore some theoretical properties of the model and derive a natural Gibbs based algorithm with MCMC hyperparameter learning. The model is implemented on a variety of synthetic and real world data sets.

207: Sparse Nested Markov models with Log-linear Parameters

Ilya Shpitser, Robin Evans, Thomas Richardson, James Robins

Hidden variables are ubiquitous in practical data analysis, and therefore modeling marginal densities and doing inference with the resulting models is an important problem in statistics, machine learning, and causal inference. Recently, a new type of graphical model, called the nested Markov model, was developed which captures equality constraints found in marginals of directed acyclic graph (DAG) models. Some of these constraints, such as the so called 'Verma constraint', strictly generalize conditional independence. To make modeling and inference with nested Markov models practical, it is necessary to limit the number of parameters in the model, while still correctly capturing the constraints in the marginal of a DAG model. Placing such limits is similar in spirit to sparsity methods for undirected graphical models, and regression models. In this paper, we give a log-linear parameterization which allows sparse modeling with nested Markov models. We illustrate the advantages of this parameterization with a simulation study.

25: Scalable Matrix-valued Kernel Learning for High-dimensional Nonlinear Multivariate Regression and Granger Causality

Vikas Sindhwani, Ha Quang Minh, Aurelie Lozano

We propose a general matrix-valued multiple kernel learning framework for high-dimensional nonlinear multivariate regression problems. This framework allows a broad class of mixed norm regularizers, including those that induce sparsity, to be imposed on a dictionary of vector-valued Reproducing Kernel Hilbert Spaces. We develop a highly scalable and eigendecomposition-free algorithm that orchestrates two inexact solvers for simultaneously learning both the input and output components of separable matrix-valued kernels. As a key application enabled by our framework, we show how high-dimensional causal inference tasks can be naturally cast as sparse function

estimation problems, leading to novel nonlinear extensions of a class of Graphical Granger Causality techniques. Our algorithmic developments and extensive empirical studies are complemented by theoretical analyses in terms of Rademacher generalization bounds.

248: Preference Elicitation For General Random Utility Models

Hossein Azari Soufiani, David Parkes, Lirong Xia

This paper discusses *General Random Utility Models (GRUMs)*. These are a class of parametric models that generate partial ranks over alternatives given attributes of agents and alternatives. We propose two preference elicitation scheme for GRUMs developed from principles in Bayesian experimental design, one for social choice and the other for personalized choice. We couple this with a general Monte-Carlo-Expectation-Maximization (MC-EM) based algorithm for MAP inference under GRUMs. We also prove uni-modality of the likelihood functions for a class of GRUMs. We examine the performance of various criteria by experimental studies, which show that the proposed elicitation scheme increases the precision of estimation.

189: Calculation of Entailed Rank Constraints in Partially Non-Linear and Cyclic Models

Peter Spirtes

The Trek Separation Theorem (Sullivant et al. 2010) states necessary and sufficient conditions for a linear directed acyclic graphical model to entail for all possible values of its linear coefficients that the rank of various sub-matrices of the covariance matrix is less than or equal to n, for any given n. In this paper, I extend the Trek Separation Theorem in two ways: I prove that the same necessary and sufficient conditions apply even when the generating model is partially non-linear and contains some cycles. This justifies application of constraint-based causal search algorithms such as the BuildPureClusters algorithm (Silva et al. 2006) for discovering the causal structure of latent variable models to data generated by a wider class of causal models that may contain non-linear and cyclic relations among the latent variables.

166: Modeling Documents with Deep Boltzmann Machines

Nitish Srivastava, Ruslan Salakhutdinov, Geoffrey Hinton

We introduce a Deep Boltzmann Machine model suitable for modeling and extracting latent semantic representations from a large unstructured collection of documents. We overcome the apparent difficulty of training a DBM with judicious parameter tying. This parameter tying enables an efficient pretraining algorithm and a state initialization scheme that aids inference. The model can be trained just as efficiently as a standard Restricted Boltzmann Machine. Our experiments show that the model assigns better log probability to unseen data than the Replicated Softmax model. Features extracted from our model outperform LDA, Replicated Softmax, and DocNADE models on document retrieval and document classification tasks.

63: Speedy Model Selection (SMS) for Copula Models

Yaniv Tenzer, Gal Elidan

We tackle the challenge of efficiently learning the structure of expressive multivariate real-valued densities of copula graphical models. We start by theoretically substantiating the conjecture that for many copula families the magnitude of Spearman's rank correlation coefficient is monotone in the expected contribution of an edge in network, namely the negative copula entropy. We then build on this theory and suggest a novel Bayesian approach that makes use of a prior over values of Spearman's rho for learning copula-based models that involve a mix of copula families. We demonstrate the generalization effectiveness of our highly efficient approach on sizable and varied real-life datasets.

106: Probabilistic inverse reinforcement learning in unknown environments

Aristide Tossou, Christos Dimitrakakis

We consider the problem of learning by demonstration from agents acting in unknown stochastic Markov environments or games. Our aim is to estimate agent preferences in order to construct improved policies for the same task that the agents are trying to solve. To do so, we extend previous probabilistic approaches for inverse reinforcement learning in known MDPs to the case of unknown dynamics or opponents. We do this by deriving two simplified probabilistic models of the demonstrator's policy and utility. For tractability, we use maximum a posteriori estimation rather than full Bayesian inference. Under a flat prior, this results in a convex optimisation problem. We find that the resulting algorithms are highly competitive against a variety of other methods for inverse reinforcement learning that do have knowledge of the dynamics.

14: Approximate Kalman Filter Q-Learning for Continuous State-Space MDPs

Charles Tripp, Ross Shachter

We seek to learn an effective policy for a Markov Decision Process (MDP) with continuous states via Q-Learning. Given a set of basis functions over state action pairs we search for a corresponding set of linear weights that minimizes the mean Bellman residual. Our algorithm uses a Kalman filter model to estimate those weights and we have developed a simpler approximate Kalman filter model that outperforms the current state of the art projected TD-Learning methods on several standard benchmark problems.

161: Finite-Time Analysis of Kernelised Contextual Bandits

Michal Valko, Nathaniel Korda, Remi Munos, Ilias Flaounas, Nelo Cristianini

We tackle the problem of online reward maximisation over a large finite set of actions described by their contexts. We focus on the case when the number of actions is too big to sample all of them even once. However we assume that we have access to the similarities between actions' contexts and that the expected reward is an arbitrary linear function of the contexts' images in the related reproducing kernel Hilbert space (RKHS). We propose KernelUCB, a kernelised UCB algorithm, and give a cumulative regret bound through a frequentist analysis. For contextual bandits, the related algorithm GP-UCB turns out to be a special case of our algorithm, and our finite-time analysis improves the regret bound of GP-UCB for the agnostic case, both in the terms of the kernel-dependent quantity and the RKHS norm of the reward function. Moreover, for the linear kernel, our regret bound matches the lower bound for contextual linear bandits.

221: Dynamic Blocking and Collapsing for Gibbs Sampling

Deepak Venugopal, Vibhav Gogate

In this paper, we investigate combining blocking and collapsing – two widely used strategies for improving the accuracy of Gibbs sampling – in the context of probabilistic graphical models (PGMs). We show that combining them is not straight-forward because collapsing (or eliminating variables) introduces new dependencies in the PGM and in computation-limited settings, this may adversely affect blocking. We therefore propose a principled approach for tackling this problem. Specifically, we develop two scoring functions, one each for blocking and collapsing, and formulate the problem of partitioning the variables in the PGM into blocked and collapsed subsets as simultaneously maximizing both scoring functions (i.e., a multi-objective optimization problem). We propose a dynamic, greedy algorithm for approximately solving this intractable optimization problem. Our dynamic algorithm periodically updates the partitioning into blocked and collapsed variables by leveraging correlation statistics gathered from the generated samples and enables rapid mixing by blocking together and collapsing highly correlated variables. We demonstrate experimentally the clear benefit of our dynamic approach: as more samples are drawn, our dynamic approach significantly outperforms static graph-based approaches by an order of magnitude in terms of accuracy.

125: Bounded Approximate Symbolic Dynamic Programming for Hybrid MDPs

Luis Gustavo Vianna, Scott Sanner, Leliane de Barros

Recent advances in symbolic dynamic programming (SDP) combined with the extended algebraic decision diagram (XADD) data structure have provided exact solutions for mixed discrete and continuous (hybrid) MDPs with piecewise linear dynamics and continuous actions. Since XADD-based exact solutions may grow intractably large for many problems, we propose a bounded error compression technique for XADDs that involves the solution of a constrained bilinear saddle point problem. Fortuitously, we show that given the special structure of this problem, it can be expressed as a bilevel linear programming problem and solved to optimality in finite time via constraint generation, despite having an infinite set of constraints. This solution permits the use of efficient linear program solvers for XADD compression and enables a novel class of bounded approximate SDP algorithms for hybrid MDPs that empirically offers order-of-magnitude speedups over the exact solution in exchange for a small approximation error.

27: On MAP Inference by MWSS on Perfect Graphs

Adrian Weller, Tony Jebara

Finding the most likely (MAP) configuration of a Markov random field (MRF) is NP-hard in general. A promising, recent technique is to reduce the problem to finding a maximum weight stable set (MWSS) on a derived weighted graph, which if perfect, allows inference in polynomial time. We derive new results for this approach, including a general decomposition theorem for MRFs of any order and number of labels, extensions of results for binary pairwise models with submodular cost functions to higher order, and an exact characterization of which binary pairwise MRFs can be efficiently solved with this method. This defines the power of the approach on this class of models, improves our toolbox and expands the range of tractable models.

35: Integrating Document Clustering and Topic Modeling

Pengtao Xie, Eric Xing

Document clustering and topic modeling are two closely related tasks which can mutually benefit each other. Topic modeling can project documents into a topic space which facilitates effective document clustering. Cluster labels discovered by document clustering can be incorporated into topic models to extract local topics specific to each cluster and global topics shared by all clusters. In this paper, we propose a multi-grain clustering topic model (MGCTM) which integrates document clustering and topic modeling into a unified framework and jointly performs the two tasks to achieve the overall best performance. Our model tightly couples two components: a mixture component used for discovering latent groups in document collection and a topic model component used for mining multi-grain topics including local topics specific to each cluster and global topics shared across clusters. We employ variational inference to approximate the posterior of hidden variables and learn model parameters. Experiments on two datasets demonstrate the effectiveness of our model.

46: Active Learning with Expert Advice

Peilin Zhao, Steven Hoi, Jinfeng Zhuang

Conventional learning with expert advice methods assumes a learner is always receiving the outcome (e.g., class labels) of every incoming training instance at the end of each trial. In real applications, acquiring the outcome from oracle can be costly or time consuming. In this paper, we address a new problem of active learning with expert advice, where the outcome of an instance is disclosed only when it is requested by the online learner. Our goal is to learn an accurate prediction model by asking the oracle the number of questions as small as possible. To address this challenge, we propose a framework of active forecasters for online active learning with expert advice, which attempts to extend two regular forecasters, i.e., Exponentially Weighted Average Forecaster and Greedy Forecaster, to tackle the task of active learning with expert advice. We prove that the proposed algorithms

satisfy the Hannan consistency under some proper assumptions, and validate the efficacy of our technique by an extensive set of experiments.

61: Bennett-type Generalization Bounds: Large-deviation Case and Faster Rate of Convergence

Chao Zhang

In this paper, we present the Bennett-type generalization bounds of the learning process for i.i.d. samples, and then show that the generalization bounds have a faster rate of convergence than the traditional results. In particular, we first develop two types of Bennett-type deviation inequality for the i.i.d. learning process: one provides the generalization bounds based on the uniform entropy number; the other leads to the bounds based on the Rademacher complexity. We then adopt a new method to obtain the alternative expressions of the Bennett-type generalization bounds, which imply that the bounds have a faster rate $o(N^{-\frac{1}{2}})$ of convergence than the traditional results $O(N^{-\frac{1}{2}})$. Additionally, we find that the rate of the bounds will become faster in the large-deviation case, which refers to a situation where the empirical risk is far away from (at least not close to) the expected risk. Finally, we analyze the asymptotical convergence of the learning process and compare our analysis with the existing results.