

# **32nd Conference on Uncertainty in Artificial Intelligence 2016**

Jersey City, New Jersey, USA  
25-29 June 2016

**Editors:**

**Alexander Ihler**

**Dominik Janzing**

ISBN: 978-1-5108-2780-6

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2016) by Association For Uncertainty in Artificial Intelligence (AUAI)  
All rights reserved.

Printed by Curran Associates, Inc. (2016)

For permission requests, please contact Association For Uncertainty in Artificial Intelligence (AUAI)  
at the address below.

Association For Uncertainty in Artificial Intelligence (AUAI)  
Department of Computer Science  
University of British Columbia  
Vancouver BC Canada V6T 1Z4

Phone: (703) 787-8044

Fax: (703) 478-0829

[www.auai.org](http://www.auai.org)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

# Contents

<b>Preface</b>	<b>v</b>
<b>Organizing Committee</b>	<b>vii</b>
<b>Acknowledgments</b>	<b>ix</b>
<b>Sponsors</b>	<b>xvii</b>
<b>Best Paper Awards</b>	<b>xix</b>
<b>1 Proceedings</b>	<b>1</b>
Gradient Methods for Stackelberg Games.	
<i>Kareem Amin, Michael Wellman, Satinder Singh</i> . . . . .	2
Inferring Causal Direction from Relational Data.	
<i>David Arbour, Katerina Marazopoulou, David Jensen</i> . . . . .	12
Convex Relaxation Regression: Black-Box Optimization of Smooth Functions by Learning Their Convex Envelopes.	
<i>Mohamm Gheshlaghi Azar, Eva Dyer, Konrad Kording</i> . . . . .	22
The Mondrian Kernel.	
<i>Matej Balog, Balaji Lakshminarayanan, Zoubin Ghahramani, Daniel Roy, Yee Whye Teh</i> . . . . .	32
Sequential Nonparametric Testing with the Law of the Iterated Logarithm.	
<i>Akshay Balsubramani, Aaditya Ramdas</i> . . . . .	42
Budget Allocation using Weakly Coupled, Constrained Markov Decision Processes.	
<i>Craig Boutilier, Tyler Lu</i> . . . . .	52
Analysis of Nyström method with sequential ridge leverage scores.	
<i>Daniele Calandriello, Alessandro Lazaric, Michal Valko</i> . . . . .	62
Unsupervised Discovery of El Nino Using Causal Feature Learning on Microlevel Climate Data.	
<i>Krzysztof Chalupka, Tobias Bischoff, Frederick Eberhardt, Pietro Perona</i> . . . . .	72
Individual Planning in Open and Typed Agent Systems.	
<i>Muthukumaran Chandrasekaran, Adam Eck, Prashant Doshi, Leenkiat Soh</i> . . . . .	82
Modeling Transitivity in Complex Networks.	
<i>Morteza Haghiri Chehreghani, Mostafa Haghiri Chehreghani</i> . . . . .	92
Adversarial Inverse Optimal Control for General Imitation Learning Losses and Embodiment Transfer.	
<i>Xiangli Chen, Mathew Monfort, Brian Ziebart, Peter Carr</i> . . . . .	102
A Generative Block-Diagonal Model for Clustering.	
<i>Junxiang Chen, Jennifer Dy</i> . . . . .	112
Optimal Stochastic Strongly Convex Optimization with a Logarithmic Number of Projections.	
<i>Jianhui Chen, Tianbao Yang, Qihang Lin, Lijun Zhang, Yi Chang</i> . . . . .	122
Accelerated Stochastic Block Coordinate Gradient Descent for Sparsity Constrained Nonconvex Optimization.	
<i>Jinghui Chen, Quanquan Gu</i> . . . . .	132
Online Bayesian Multiple Kernel Bipartite Ranking.	
<i>Changying Du, Changde Du, Guoping Long, Qing He, Yucheng Li</i> . . . . .	142

Pruning Rules for Learning Parsimonious Context Trees.	
<i>Ralf Eggeling, Mikko Koivisto</i>	152
Learning Network of Multivariate Hawkes Processes: A Time Series Approach.	
<i>Jalal Etesami, negar kiyavash, Kun Zhang, Kushagra Singhal</i>	162
Elliptical Slice Sampling with Expectation Propagation.	
<i>Francois Fagan, Jalaj Bhandari, John Cunningham</i>	172
Bayesian Learning of Kernel Embeddings.	
<i>Seth Flaxman, Dino Sejdinovic, John Cunningham, Sarah Filippi</i>	182
On the Theory and Practice of Privacy-Preserving Bayesian Data Analysis.	
<i>James Foulds, Joseph Geumlek, Max Welling, Kamalika Chaudhuri</i>	192
Taming the Noise in Reinforcement Learning via Soft Updates.	
<i>Roy Fox, Ari Pakman, Naftali Tishby</i>	202
Quasi-Newton Hamiltonian Monte Carlo.	
<i>Tianfan Fu, Luo Luo, Zhihua Zhang</i>	212
Scalable Joint Modeling of Longitudinal and Point Process Data for Disease Trajectory Prediction and Improving Management of Chronic Kidney Disease.	
<i>Joseph Futoma, Mark Sendak, Blake Cameron, Katherine Heller</i>	222
Degrees of Freedom in Deep Neural Networks.	
<i>Tianxiang Gao, Vladimir Jojic</i>	232
Training Neural Nets to Aggregate Crowdsourced Responses.	
<i>Alex Gaunt, Diana Borsa, Yoram Bachrach</i>	242
Model-Free Reinforcement Learning with Skew-Symmetric Bilinear Utilities.	
<i>Hugo Gilbert, Bruno Zanuttini, Paul Weng, Paolo Viappiani, Esther Nicart</i>	252
Bridging Heterogeneous Domains With Parallel Transport For Vision and Multimedia Applications.	
<i>Raghuraman Gopalan</i>	262
Structured Prediction: From Gaussian Perturbations to Linear-Time Principled Algorithms.	
<i>Jean Honorio, Tommi Jaakkola</i>	271
Efficient Feature Group Sequencing for Anytime Linear Prediction.	
<i>Hanzhang Hu, Alexander Grubb, J. Andrew Bagnell, Martial Hebert</i>	279
Scalable Nonparametric Bayesian Multilevel Clustering.	
<i>Viet Huynh, Dinh Phung, Svetha Venkatesh, Long Nguyen, Matthew Hoffman, Hung Bui</i>	289
Hierarchical learning of grids of microtopics.	
<i>Nebojsa Jojic, Alessandro Perina, Dongwoo Kim</i>	299
Active Uncertainty Calibration in Bayesian ODE Solvers.	
<i>Hans Kersting, Philipp Hennig</i>	309
Faster Stochastic Variational Inference using Proximal-Gradient Methods with General Divergence Functions.	
<i>Mohammad Emtiyaz Khan, Reza Babanezhad Harikandeh, Wu Lin, Mark Schmidt, Masashi Sugiyama</i>	319
Probabilistic Size-constrained Microclustering.	
<i>Arto Klami, Aditya Jitta</i>	329
Online learning with Erdos-Renyi side-observation graphs.	
<i>Tomáš Kocák, gergely Neu, Michal Valko</i>	339
Conjugate Conformal Prediction for Online Binary Classification.	
<i>Mustafa Kocak, Dennis Shasha, Elza Erkip</i>	347
Online Forest Density Estimation.	
<i>Frederic Koriche</i>	357
Towards a Theoretical Understanding of Negative Transfer in Collective Matrix Factorization.	
<i>Chao Lan, Jianxin Wang, Jun Huan</i>	367
Budgeted Semi-supervised Support Vector Machine .	
<i>Trung Le, Phuong Duong, Mi Dinh, Tu Nguyen, Vu Nguyen, Dinh Phung</i>	377
A Characterization of Markov Equivalence Classes of Relational Causal Models under Path Semantics.	
<i>Sanghack Lee, Vasant Honavar</i>	387

Improving Imprecise Compressive Sensing Models. <i>Donggeun Lee, Rafael de Lima, Jaesik Choi</i> . . . . .	397
Bounded Rational Decision-Making in Feedforward Neural Networks. <i>Felix Leibfried, Daniel Braun</i> . . . . .	407
Thompson Sampling is Asymptotically Optimal in General Environments. <i>Jan Leike, Tor Lattimore, Laurent Orseau, Marcus Hutter</i> . . . . .	417
A Formal Solution to the Grain of Truth Problem. <i>Jan Leike, Jessica Taylor, Benya Fallenstein</i> . . . . .	427
Bayesian Hyperparameter Optimization for Ensemble Learning. <i>Julien-Charles Levesque, Christian Gagne, Robert Sabourin</i> . . . . .	437
Political Dimensionality Estimation Using a Probabilistic Graphical Model. <i>Yoad Lewenberg, Yoram Bachrach, Lucas Bordeaux, Pushmeet Kohli</i> . . . . .	447
Correlated Tag Learning in Topic Model. <i>Shuangyin Li, Rong Pan, Yu Zhang, Qiang Yang</i> . . . . .	457
Utilize Old Coordinates: Faster Doubly Stochastic Gradients for Kernel Methods. <i>Chun-Liang Li, Barnabas Poczos</i> . . . . .	467
On Hyper-Parameter Estimation In Empirical Bayes: A Revisit of The MacKay Algorithm. <i>Chune Li, Yongyi Mao, Richong Zhang, Jinpeng Huai</i> . . . . .	477
Dantzig Selector with an Approximately Optimal Denoising Matrix and its Application in Sparse Reinforcement Learning. <i>Bo Liu, Luwan Zhang, Ji Liu</i> . . . . .	487
Importance Weighted Consensus Monte Carlo for Distributed Bayesian Inference. <i>Qiang Liu</i> . . . . .	497
Large-scale Submodular Greedy Exemplar Selection with Structured Similarity Matrices. <i>Dmitry Malioutov, Abhishek Kumar, Ian Yen</i> . . . . .	507
Sparse Gaussian Processes for Bayesian Optimization. <i>Mitchell McIntire, Daniel Ratner, Stefano Ermon</i> . . . . .	517
A General Statistical Framework for Designing Strategy-proof Assignment Mechanisms. <i>Harikrishna Narasimhan, David Parkes</i> . . . . .	527
A Correlated Worker Model for Grouped, Imbalanced and Multitask Data. <i>An Nguyen, Byron Wallace, Matthew Lease</i> . . . . .	537
Convergence Rates for Greedy Kaczmarz Algorithms, and Randomized Kaczmarz Rules Using the Orthogonality Graph. <i>Julie Nutini, Behrooz Sepehry, Issam Laradji, Mark Schmidt, Hoyt Koepke, Alim Virani</i> .	547
Safely Interruptible Agents. <i>Laurent Orseau, Stuart Armstrong</i> . . . . .	557
Super-Sampling with a Reservoir. <i>Brooks Paige, Dino Sejdinovic, Frank Wood</i> . . . . .	567
Alternative Markov and Causal Properties for Acyclic Directed Mixed Graphs. <i>Jose Peña</i> . . . . .	577
Bounded Rationality in Wagering Mechanisms. <i>David Pennock, Vasilis Syrgkanis, Jennifer Wortman Vaughan</i> . . . . .	587
Incremental Preference Elicitation for Decision Making Under Risk with the Rank-Dependent Utility Model. <i>Patrice Perny, Paolo Viappiani, abdellah Boukhatem</i> . . . . .	597
Interpretable Policies for Dynamic Product Recommendations. <i>Marek Petrik, Ronny Luss</i> . . . . .	607
Merging Strategies for Sum-Product Networks: From Trees to Graphs. <i>Tahrima Rahman, Vibhav Gogate</i> . . . . .	617
MDPs with Unawareness in Robotics. <i>Nan Rong, Joseph Halpern, Ashutosh Saxena</i> . . . . .	627
A Kernel Test for Three-Variable Interactions with Random Processes. <i>Paul Rubenstein, Kacper Chwialkowski, Arthur Gretton</i> . . . . .	637
Overdispersed Black-Box Variational Inference. <i>Francisco Ruiz, Michalis Titsias, david Blei</i> . . . . .	647

Stochastic Portfolio Theory: A Machine Learning Approach. <i>Yves-Laurent Kom Samo, Alexander Vervuurt</i> . . . . .	657
Stability of Causal Inference. <i>Leonard Schulman, Piyush Srivastava</i> . . . . .	666
Markov Beta Processes for Time Evolving Dictionary Learning. <i>Amar Shah, Zoubin Ghahramani</i> . . . . .	676
Finite Sample Complexity of Rare Pattern Anomaly Detection. <i>Md Amran Siddiqui, Alan Fern, Thomas Dietterich, Shubhomoy Das</i> . . . . .	686
The Deterministic Information Bottleneck. <i>DJ Strouse, david Schwab</i> . . . . .	696
Learning to Smooth with Bidirectional Predictive State Inference Machines. <i>Wen Sun, Roberto Capobianco, Geoffrey J. Gordon, J. Andrew Bagnell, Byron Boots</i> . . .	706
Context-dependent feature analysis with random forests. <i>Antonio Suter, Gilles Louppe, Vân Anh Huynh-Thu, Louis Wehenkel, Pierre Geurts</i> . .	716
Content-based Modeling of Reciprocal Relationships using Hawkes and Gaussian Processes. <i>Xi Tan, Syed Naqvi, Yuan, Qi, Katherine Heller, vinayak Rao</i> . . . . .	726
Forward Backward Greedy Algorithms for Multi-Task Learning with Faster Rates. <i>Lu Tian, Pan Xu, Quanquan Gu</i> . . . . .	735
Non-parametric Domain Approximation for Scalable Gibbs Sampling in MLNs. <i>Deepak Venugopal, Somdeb Sarkhel, Kyle Cherry</i> . . . . .	745
Efficient Observation Selection in Probabilistic Graphical Models Using Bayesian Lower Bounds. <i>Dilin Wang, John Fisher III, Qiang Liu</i> . . . . .	755
Characterizing Tightness of LP Relaxations by Forbidding Signed Minors. <i>Adrian Weller</i> . . . . .	765
Subspace Clustering with a Twist. <i>David Wipf, Yue Dong, Bo Xin</i> . . . . .	775
Bayesian Estimators As Voting Rules. <i>Lirong Xia</i> . . . . .	785
Lighter-Communication Distributed Machine Learning via Sufficient Factor Broadcasting. <i>Pengtao Xie, Jin Kyu Kim, Yi Zhou, Qirong Ho, Abhimanu Kumar, Yaoliang Yu, Eric Xing</i> . . . . .	795
Efficient Multi-Class Selective Sampling on Graphs. <i>Peng Yang, Peilin Zhao, Zhen Hai, Wei Liu, Steven C.H. Hoi, Xiao-Li Li</i> . . . . .	805
Adaptive Algorithms and Data-Dependent Guarantees for Bandit Convex Optimization. <i>Scott Yang, Mehryar Mohri</i> . . . . .	815
On the Identifiability and Estimation of Functional Causal Models in the Presence of Outcome-Dependent Selection. <i>Kun Zhang, Jiji Zhang, Biwei Huang, Bernhard Schölkopf, Clark Glymour</i> . . . . .	825
Cascading Bandits for Large-Scale Recommendation Problems. <i>Shi Zong, Hao Ni, Kenny Sung, Rosemary Ke, Zheng Wen, Branislav Kveton</i> . . . . .	835