NIPS 2017 Workshop book

Workshop organizers make last-minute changes to their schedule. Download this document again to get the latest changes, or use the NIPS mobile application.

Schedule Highlights

**Dec. 9, 2017**

101-A, Deep Learning at Supercomputer Scale  
Eisen, Hafner, Stone, Saeta

101-B, (Almost) 50 shades of Bayesian Learning: PAC-Bayesian trends and insights  
Guedj, Germain, Bach

102-A-B, Optimal Transport and Machine Learning  
Bousquet, Cuturi, Peyré, Sharma

102-C, Emergent Communication Workshop  
Foerster, Mordatch, Lazaridou, Cho, Kiela, Abbeel

103 A-B, Machine Learning on the Phone and other Consumer Devices  
Aradhye, Quinonero Candela, Prasad

104-A, Teaching Machines, Robots, and Humans  
Cakmak, Rafferty, Singla, Zhu, Zilles

104-B, Cognitively Informed Artificial Intelligence: Insights From Natural Intelligence  
Mozzer, Lake, Yu

104-C, Machine Learning in Computational Biology  
Zou, Kundaje, Quon, Fusi, Mostafavi

201-A, BigNeuro 2017: Analyzing brain data from nano to macroscale  
Dyer, Klar, Gray Ronca, Koering, Vogelstein

201-B, 2017 NIPS Workshop on Machine Learning for Intelligent Transportation Systems  
Li, Dragan, Niebles, Savarese

202, The future of gradient-based machine learning software & techniques  
Wittchko, van Merriënboer, Lamblin

203, NIPS Highlights (MLTrain), Learn how to code a paper with state of the art frameworks  
Dimakis, Vasiloglou, Van den Broeck, Ihler, Araki

204, Machine Learning Challenges as a Research Tool  
Guyon, Viégas, Escalera, Abernethy

205-B, Deep Learning: Bridging Theory and Practice  
Arora, Raghu, Salakhutdinov, Schmidt, Vinyals

205-C, Hierarchical Reinforcement Learning  
Precup, Mannor, Schaul, Fox, Florensa Campo

206, Bayesian Deep Learning  
Gal, Hernández-Lobato, Louizos, Wilson, Kingma, van den Oord, van der Maaten

207, Interpreting, Explaining and Visualizing Deep Learning - Now what?  
Müller, Vedaldi, Hansen, Samek, Montavon

208, Workshop on Prioritising Online Content  
Shaw-Taylor, Pontil, Cesa-Bianchi, Yilmaz, Watkins, Riedel, Groebnlik

209, Collaborate & Communicate: An exploration and practical skills workshop that builds on the experience of AIML experts who are both successful collaborators and great communicators  
Gorman

210, Synergies in Geometric Data Analysis (2nd day)  
Melia, Chazal

211, Medical Imaging meets NIPS  
Glocker, Konukoglu, Lambaert, Bhatia

**Dec. 8, 2017**

Grand Ballroom A, OPT 2017: Optimization for Machine Learning  
Sra, J. Reddi, Agarwal, Recht

Grand Ballroom B, NIPS 2017 Time Series Workshop  
Kuznetsov, Anava, Yang, Khaleghi

Hall C, From 'What If?' To 'What Next?': Causal Inference and Machine Learning for Intelligent Decision Making  
Volfovsky, Swaminathan, Toulis, Kallus, Silva, Shaw-Taylor, Joachims, Li

Hyatt Beach Ballroom D+E+F+H, Extreme Classification: Multi-class & Multi-label Learning in Extremely Large Label Spaces  
Varma, Kloft, Dembczyński

Hyatt Regency Ballroom A+B+C, Conversational AI - today's practice and tomorrow's potential  
Germán, Williams, Heck, Glass, Bordes, Young, Tesfau

Hyatt Seaview Ballroom, Discrete Structures in Machine Learning  
Singer, Bimnes, Krause, Jegelka, Karbasi

Hyatt Shoreline, Nearest Neighbors for Modern Applications with Massive Data: An Age-old Solution with New Challenges  
Chen, Shah, Lee

Room 101-A, Visually grounded interaction and language  
Strub, de Vries, Das, Kotur, Lee, Malinowski, Pietquin, Parikh, Batra, Courville, Mar

Room 101-B, Learning in the Presence of Strategic Behavior  
Haghtalab, Mansour, Roughgarden, Syrgkanis, Wortman Vaughan

Room 102 A+B, Machine Learning for Creativity and Design  
Eck, Ha, Eslami, Dieleman, Fiebrink, Elliott

Room 102-C, Machine Learning for Molecules and Materials  
Chmiela, Hernández-Lobato, Schütt, Aspuru-Guzik, Tkatchenko, Rasmussen, von Lilienfeld, Kusner, Tsuda, Paige, Müller

Room 103 A+B, Advances in Modeling and Learning Interactions from Complex Data  
Dasarathy, Kolar, Baraniuk

Room 103-C, 6th Workshop on Automated Knowledge Base Construction (AKBC)  
Pujara, Chen, Dalvi Mishra, Rocktäschel

Room 104-A, Advances in Approximate Bayesian Inference  
Ruiz, Mandt, Zhang, McInerney, Tran, Broderick, Titsias, Blei, Welling

Room 104-B, Machine Learning for Health (ML4H) - What Parts of Healthcare are Ripe for Disruption by Machine Learning Right Now?  
Beam, Fiterau, Schulam, Fries, Hughes, Wittchko, Snoek, Antropova, Ranganath, Jedaynak, Naumann, Dalca, Dalca, Althoff, ASTHANA, Tandon, Randola, Ratner, Kale, Shalit, Ghassemi, Kohane

Room 104-C, Acting and Interacting in the Real World: Challenges in Robot Learning  
Posner, Hadsell, Riedmiller, Wulfmeier, Paul
Room 201-A, Transparent and interpretable Machine Learning in Safety Critical Environments Tosi, Vellido, Álvarez

Room 201-B, Machine Learning for Audio Signal Processing (ML4Audio) Purwins, Sturm, Plumbley

Room 202, Deep Learning for Physical Sciences Baydin, Prabhat, Cranmer, Wood

Room 203, Machine Deception Goodfellow, Hwang, Goodman, Rodriguez

Room 204, ML Systems Workshop @ NIPS 2017 Lakshmiratan, Bird, Sen, Ré, Li, Gonzalez, Crankshaw

Room S-1, Machine Learning for the Developing World De-Arteaga, Herlands

Room S-3, Workshop on Worm's Neural Information Processing (WNIP) Hasani, Zimmer, Larson, Grosu

Room S-5, Synergies in Geometric Data Analysis Meila, Chazal

Room S-7, Competition track Escalera, Weimer

S-4, Machine Learning and Computer Security Steinhardt, Papernot, Li, Liu, Liang, Song

Seaside Ballroom, Learning on Distributions, Functions, Graphs and Groups d'Alché-Buc, Muandet, Sriperumbudur, Szabó
Deep Learning at Supercomputer Scale

Erich Elsen, Danijar Hafner, Zak Stone, Brennan Saeta

101-A, Sat Dec 09, 08:00 AM

Five years ago, it took more than a month to train a state-of-the-art image recognition model on the ImageNet dataset. Earlier this year, Facebook demonstrated that such a model could be trained in an hour. However, if we could parallelize this training problem across the world’s fastest supercomputers (~100 PFlops), it would be possible to train the same model in under a minute. This workshop is about closing that gap: how can we turn months into minutes and increase the productivity of machine learning researchers everywhere?

This one-day workshop will facilitate active debate and interaction across many different disciplines. The conversation will range from algorithms to infrastructure to silicon, with invited speakers from Cerebras, DeepMind, Facebook, Google, OpenAI, and other organizations. When should synchronous training be preferred over asynchronous training? Are large batch sizes the key to reach supercomputer scale, or is it possible to fully utilize a supercomputer at batch size one? How important is sparsity in enabling us to scale? Should sparsity patterns be structured or unstructured? To what extent do we expect to customize model architectures for particular problem domains, and to what extent can a “single model architecture” deliver state-of-the-art results across many different domains? How can new hardware architectures unlock even higher real-world training performance?

Our goal is bring people who are trying to answer any of these questions together in hopes that cross pollination will accelerate progress towards deep learning at true supercomputer scale.

(Almost) 50 shades of Bayesian Learning: PAC-Bayesian trends and insights

Benjamin Guedj, Pascal Germain, Francis Bach

101-B, Sat Dec 09, 08:00 AM

Industry-wide successes of machine learning at the dawn of the (so-called) big data era has led to an increasing gap between practitioners and theoreticians. The former are using off-the-shelf statistical and machine learning methods, while the latter are designing and studying the mathematical properties of such algorithms. The tradeoff between those two movements is somewhat addressed by Bayesian researchers, where sound mathematical guarantees often meet efficient implementation and provide model selection criteria. In the late 90s, a new paradigm has emerged in the statistical learning community, used to derive probably approximately correct (PAC) bounds on Bayesian-flavored estimators. This PAC-Bayesian theory has been pioneered by Shawe-Taylor and Williamson (1997), and McAllester (1998, 1999). It has been extensively formalized by Catoni (2004, 2007) and has triggered, slowly but surely, increasing research efforts during last decades.

We believe it is time to pinpoint the current PAC-Bayesian trends relatively to other modern approaches in the (statistical) machine learning community. Indeed, we observe that, while the field grows by its own, it took some undesirable distance from some related areas. Firstly, it seems to us that the relation to Bayesian methods has been forsaken in numerous works, despite the potential of PAC-Bayesian theory to bring new insights to the Bayesian community and to go beyond the classical Bayesian/frequentist divide. Secondly, the PAC-Bayesian methods share similarities with other quasi-Bayesian (or pseudo-Bayesian) methods studying Bayesian practices from a frequentist standpoint, such as the Minimum Description Length (MDL) principle (Grünwald, 2007). Last but not least, even if some practical and theory grounded learning algorithm has emerged from PAC-Bayesian works, these are almost unused for real-world problems.

In short, this workshop aims at gathering statisticians and machine learning researchers to discuss current trends and the future of (PAC,quasi)-Bayesian learning. From a broader perspective, we aim to bridge the gap between several communities that can all benefit from sharper statistical guarantees and sound theory-driven learning algorithms.

References

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Speaker</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30 AM</td>
<td>Overture</td>
<td>Guedj</td>
</tr>
<tr>
<td>08:45 AM</td>
<td>François Laviolette</td>
<td>Laviolette</td>
</tr>
<tr>
<td>09:35 AM</td>
<td>Peter Grünwald</td>
<td>Grünwald</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Jean-Michel Marin</td>
<td>Marin</td>
</tr>
<tr>
<td>11:50 AM</td>
<td>Contributed talk 1</td>
<td></td>
</tr>
<tr>
<td>02:00 PM</td>
<td>Daniel Roy</td>
<td>Roy</td>
</tr>
<tr>
<td>02:25 PM</td>
<td>Olivier Catoni</td>
<td>Catoni</td>
</tr>
<tr>
<td>03:30 PM</td>
<td>Yevgeny Seldin</td>
<td>Seldin</td>
</tr>
<tr>
<td>04:20 PM</td>
<td>John Shawe-Taylor</td>
<td>Shawe-Taylor</td>
</tr>
<tr>
<td>05:10 PM</td>
<td>Contributed talk 2</td>
<td></td>
</tr>
</tbody>
</table>
Optimal Transport and Machine Learning

Olivier Bousquet, Marco Cuturi, Gabriel Peyré, Fei Sha, Justin Solomon

102 A+B, Sat Dec 09, 08:00 AM

Optimal transport (OT) is gradually establishing itself as a powerful and essential tool to compare probability measures, which in machine learning take the form of point clouds, histograms, bags-of-features, or more generally datasets to be compared with probability densities and generative models. OT can be traced back to early work by Monge, and later to Kantorovich and Dantzig during the birth of linear programming. The mathematical theory of OT has produced several important developments since the 90’s, crowned by Cédric Villani’s Fields Medal in 2010. OT is now transitioning into more applied spheres, including recent applications to machine learning, because it can tackle challenging learning scenarios including dimensionality reduction, structured prediction problems that involve histograms, and estimation of generative models in highly degenerate, high-dimensional problems. This workshop will follow that organized 3 years ago (NIPS 2014) and will seek to amplify that trend. We will provide the audience with an update on all of the very recent successes brought forward by efficient solvers and innovative applications through a long list of invited talks. We will add to that a few contributed presentations (oral, and, if needed posters) and, finally, a panel for all invited speakers to take questions from the audience and formulate more nuanced opinions on this nascent field.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Speaker</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:00 AM</td>
<td>Structured Optimal Transport (with T. Jaakkola, Alvarez Melis S. Jegelka)</td>
<td>Jacob</td>
</tr>
<tr>
<td>08:20 AM</td>
<td>Approximate Bayesian computation with the Wasserstein distance</td>
<td>Jacob</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>Gradient flow in the Wasserstein metric</td>
<td>Craig</td>
</tr>
<tr>
<td>09:40 AM</td>
<td>Approximate inference with Wasserstein gradient flows (with T. Poggio)</td>
<td>Frogner</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>6 x 3 minutes spotlights</td>
<td>Flamary, Chen, Rujeerapaiboon, Adler, Lee, Roberts</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Wasserstein embeddings (tba)</td>
<td>Andoni</td>
</tr>
<tr>
<td>11:40 AM</td>
<td>Laplacian operator and Brownian motions on the Wasserstein space</td>
<td>Gangbo</td>
</tr>
</tbody>
</table>

Abstracts (8):

Abstract 2: Approximate Bayesian computation with the Wasserstein distance in Optimal Transport and Machine Learning, Jacob 08:20 AM

A growing range of generative statistical models prohibits the numerical evaluation of their likelihood functions. Approximate Bayesian computation has become a popular approach to overcome this issue, simulating synthetic data given parameters and comparing summaries of these simulations with the corresponding observed values. We propose to avoid these summaries and the ensuing loss of information through the use of Wasserstein distances between empirical distributions of observed and synthetic data. We describe how the approach can be used in the setting of dependent data such as time series, and how approximations of the Wasserstein distance allow in practice the method to scale to large datasets. In particular, we propose a new approximation to the optimal assignment problem using the Hilbert space-filling curve. The approach is illustrated on various examples including i.i.d. data and time series.

Abstract 3: Gradient flow in the Wasserstein metric in Optimal Transport and Machine Learning, Craig 09:00 AM

Optimal transport not only provides powerful techniques for comparing probability measures, but also for analyzing their evolution over time. For a range of partial differential equations arising in physics, biology, and engineering, solutions are gradient flows in the Wasserstein metric; each equation has a notion of energy for which solutions dissipate energy as
quickly as possible, with respect to the Wasserstein structure. Steady states of the equation correspond to minimizers of the energy, and stability properties of the equation translate into convexity properties of the energy. In this talk, I will compare Wasserstein gradient flow with more classical gradient flows arising in optimization and machine learning. I'll then introduce a class of particle blob methods for simulating Wasserstein gradient flows numerically.

Abstract 5: 6 x 3 minutes spotlights in Optimal Transport and Machine Learning, Flamary, Chen, Rujeraapaiboon, Adler, Lee, Roberts 10:00 AM

1. Nicolas Courty, Rémi Flamary and Mélanie Ducotte. Learning Wasserstein Embeddings
2. Yongxin Chen, Tryphon Georgiou and Allen Tannenbaum. Optimal transport for Gaussian mixture models
5. John Lee, Adam Charles, Nicholas Bertrand and Christopher Rozell. An Optimal Transport Tracking Regularizer

Abstract 7: Laplacian operator and Brownian motions on the Wasserstein space in Optimal Transport and Machine Learning, Gangbo 11:40 AM

We endow the space of probability measures on $\mathfrak{S}(\mathbb{R}^d)$ with the Wasserstein distance. The smoothing effect of the heat equation is established for a class of functions. Special perturbations of the Laplacian operator, denoted $\Delta_{\mathfrak{S}(\mathbb{R}^d)}$, appearing in Mean Field Games theory, are considered. (Joint work with Y. T. Chow).

Abstract 8: Geometrical Insights for Unsupervised Learning in Optimal Transport and Machine Learning, Bottou 01:40 PM

After arguing that choosing the right probability distance is critical for achieving the elusive goals of unsupervised learning, we compare the geometric properties of the two currently most promising distances: (1) the earth-mover distance, and (2) the energy distance, also known as the maximum mean discrepancy. These insights allow us to give a fresh viewpoint on reported experimental results and to risk a couple of predictions. Joint work with Leon Bottou, Martin Arjovsky, David Lopez-Paz, and Maxime Oquab.

Abstract 11: Domain adaptation with optimal transport : from mapping to learning with joint distribution in Optimal Transport and Machine Learning, Flamary 03:30 PM

This presentation deals with the unsupervised domain adaptation problem, where one wants to estimate a prediction function $f$ in a given target domain without any labeled sample by exploiting the knowledge available from a source domain where labels are known. After a short introduction of recent development in domain adaptation and their relation to optimal transport we will present a method that estimates a barycentric mapping between the feature distributions in order to adapt the training dataset prior to learning.

Next we propose a novel method that model with optimal transport the transformation between the joint feature/labels space distributions of the two domains. We aim at recovering an estimated target distribution $\tilde{f}(X,Y)$ by optimizing simultaneously the optimal coupling and $f$. We discuss the generalization of the proposed method, and provide an efficient algorithmic solution. The versatility of the approach, both in terms of class of hypothesis or loss functions is demonstrated with real world classification, regression problems and large datasets where stochastic approaches become necessary.

Joint work with Nicolas COURTY, Devis TUJA, Amaury HABRARD, and Alain RAKOTOMAMONJY

Abstract 12: Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance in Optimal Transport and Machine Learning, Bach 04:10 PM

The Wasserstein distance between two probability measures on a metric space is a measure of closeness with applications in statistics, probability, and machine learning. In this work, we consider the fundamental question of how quickly the empirical measure obtained fromiindependent samples from $\mu$ approaches $\mu$ in the Wasserstein distance of any order. We prove sharp asymptotic and finite-sample results for this rate of convergence for general measures on general compact metric spaces. Our finite-sample results show the existence of multi-scale behavior, where measures can exhibit radically different rates of convergence as $n$ grows. See more details in: J. Weed, F. Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. Technical Report, Arxiv-1707.00087, 2017.

Abstract 13: 7 x 3 minutes spotlights in Optimal Transport and Machine Learning, Cazelles, Genevay, Mena, Brauer, Fischer, Petzka, Seguy, Rolet, SONODA 04:50 PM

5. Henning Petzka, Asja Fischer and Denis Lukovnikov. On the regularization of Wasserstein GANs
6. Vivien Seguy, Bharath Bhushan Damodaran, Rémi Flamary, Nicolas Courty, Antoine Rolet and Mathieu Blondel. Large Scale Optimal Transport and Mapping Estimation

Emergent Communication Workshop

Jakob Foerster, Igor Mordatch, Angeliki Lazaridou, Kyunghyun Cho, Douwe Kiela, Pieter Abbeel
102-C, Sat Dec 09, 08:00 AM

Communication is one of the most impressive human abilities. The question of how communication arises has been studied for many decades, if not centuries. However, due to the computational and representational limitations, in the past problem-settings had to be restricted to low dimensional, simple observation spaces. With the rise of deep reinforcement learning methods, this question can now be studied in complex multi-agent settings, which has lead to flourishing activity in the area over the last two years. In these settings agents can learn to communicate in grounded multi-modal environments and rich communication protocols emerge.

However, the recent research has been largely disconnected from the study of emergent communication in other fields and even from work done on this topic in previous decades. This workshop will provide a forum for a variety of researchers from different fields (machine learning, game-theory, linguistics, cognitive science, and programming languages) interested in the question of communication and emergent language to exchange ideas.

https://sites.google.com/site/emecom2017/

Machine Learning on the Phone and other Consumer Devices

Hrishikesh Aradhye, Joaquin Quinonero Candela, Rohit Prasad

103 A+B, Sat Dec 09, 08:00 AM

Deep Machine Learning has changed the computing paradigm. Products of today are built with machine intelligence as a central attribute, and consumers are beginning to expect near-human interaction with the appliances they use. However, much of the Deep Learning revolution has been limited to the cloud, enabled by popular toolkits such as Caffe, TensorFlow, and MxNet, and by specialized hardware such as TPUs. In comparison, mobile devices until recently were just not fast enough, there were limited developer tools, and there were limited use cases that required on-device machine learning. That has recently started to change, with the advances in real-time computer vision and spoken language understanding driving real innovation in intelligent mobile applications. Several mobile-optimized neural network libraries were recently announced (CoreML [1], Caffe2 for mobile [2], TensorFlow Lite [3]), which aim to dramatically reduce the barrier to entry for mobile machine learning. Innovation and competition at the silicon layer has enabled new possibilities for hardware acceleration. To make things even better, mobile-optimized versions of several state-of-the-art benchmark models were recently opened sourced [4]. Widespread increase in availability of connected “smart” appliances for consumers and IoT platforms for industrial use cases means that there is an ever-expanding surface area for mobile intelligence and ambient devices in homes. All of these advances in combination imply that we are likely at the cusp of a rapid increase in research interest on the particular, on-device neural computing.

Significant research challenges remain, however. Mobile devices are even more personal than “personal computers” were. Enabling machine learning while simultaneously preserving user trust requires ongoing advances in the research of differential privacy and federated learning techniques. On-device ML has to keep model size and power usage low while simultaneously optimizing for accuracy. There are a few exciting novel approaches recently being developed in mobile optimization of neural networks. Lastly, the newly prevalent use of camera and voice as interaction models has fueled exciting research towards neural techniques for image and speech/language understanding.

With this emerging interest as well as the wealth of challenging research problems in mind, we are proposing the first NIPS workshop dedicated to on-device machine learning for mobile and ambient home consumer devices. We believe that interest in this space is only going to increase, and we hope that the workshop plays the role of an influential catalyst to foster research and collaboration in this nascent community.

The next wave of ML applications will have significant processing on mobile and ambient devices. Some immediate examples of these are single-image depth estimation, object recognition and segmentation running on-device for creative effects, or on-device recommender and ranking systems for privacy-preserving, low-latency experiences. This workshop will bring ML practitioners up to speed on the latest trends for on-device applications of ML, offer an overview of the latest HW and SW framework developments, and champion active research towards hard technical challenges emerging in this nascent area. The target audience for the workshop is both industrial and academic researchers and practitioners of on-device, native machine learning. The workshop will cover both “informational” and “aspirational” aspects of this emerging research area for delivering groundbreaking experiences on real-world products.


Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:05 AM</td>
<td>Qualcomm presentation on ML-optimized mobile hardware</td>
</tr>
<tr>
<td>08:30 AM</td>
<td>fpgaConvNet: A Toolflow for Mapping Diverse Convolutional Neural Networks on Embedded FPGAs</td>
</tr>
<tr>
<td>08:45 AM</td>
<td>High performance ultra-low-precision convolutions on mobile devices</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>Caffe2: Lessons from Running Deep Learning on the World’s Smart Phones</td>
</tr>
<tr>
<td>09:30 AM</td>
<td>CoreML: High-Performance On-Device Inference</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Data center to the edge: a journey with TensorFlow</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>On-Device ML Frameworks</td>
</tr>
</tbody>
</table>
This workshop focuses on “machine teaching”, the inverse problem of machine learning, in which the goal is to find an optimal training set given a machine learning algorithm and a target model. The study of machine teaching began in the early 1990s, primarily coming out of computational learning theory. Recently, there has been a surge of interest in machine teaching as several different communities within machine learning have found connections to this problem; these connections have included the following:

* machine teaching has close connections to newly introduced models of interaction in machine learning community, such as curriculum learning, self-paced learning, and knowledge distillation. [Hinton et al. 2015; Bengio et al. 2009]

In this workshop, we draw attention to machine teaching by emphasizing how the area of machine teaching interacts with emerging research trends and application domains relevant to the NIPS community. The goal of this workshop is to foster these ideas by bringing together researchers with expertise/interest in the inter-related areas of machine teaching, interactive machine learning, robotics, cyber-security problems, generative adversarial networks, educational technologies, and cognitive science.

Topics of interests in the workshop include (but are not limited to):

* Theoretical foundations of machine teaching:
  ** using tools from information theory to develop better mathematical models of teaching;
  ** characterizing the complexity of teaching when a teacher has limited power, or incomplete knowledge of student’s model, or a mismatch in feature representations;
  ** algorithms for adaptive teaching by interactively inferring the learner’s state;
  ** new notions of Teaching-dimension for generic teaching settings.

* Connections to machine learning models:
  ** the information complexity of teaching and query complexity;
  ** machine teaching vs. curriculum learning and other models of interactive machine learning;
  ** teaching reinforcement learning agents.

* Applications of machine teaching to adversarial attacks, including cyber-security problems, generative adversarial networks, attacks on machine learning algorithms, etc.

* Applications of machine teaching to educational technologies:
  ** using the machine teaching formulation to enable more rigorous and generalizable approaches for developing intelligent tutoring systems;
  ** behavioral experiments to identify good cognitive models of human learning processes.

* Novel applications for machine teaching such as program synthesis, human-robot interactions, social robotics, etc.

Schedule

02:00 PM  Machine Education - the way forward for achieving trust-enabled machine agents  Leu
human-like performance. We conjecture that this inductive bias will need for machine-learning systems to attain the sort of strengths and flexibility of the human cognitive architecture. The adaptive machines, and can endow learning architectures with the from cognitive science that can inform the development of intelligent.html

The premise of this workshop is that there are valuable data and models counterparts.

areas such as memory, attention, and development have rich theoretical and experimental histories, yet these concepts, as applied to AI systems so far, only bear a superficial resemblance to their biological counterparts.

The goal of this workshop is to bring together cognitive scientists, neuroscientists, and AI researchers to discuss opportunities for improving machine learning by leveraging our scientific understanding of human perception and cognition. There is a history of making these connections: artificial neural networks were originally motivated by the massively parallel, deep architecture of the brain; considerations of biological plausibility have driven the development of learning procedures; and architectures for computer vision draw parallels to the connectivity and physiology of mammalian visual cortex. However, beyond these celebrated examples, cognitive science and neuroscience has fallen short of its potential to influence the next generation of AI systems. Areas such as memory, attention, and development have rich theoretical and experimental histories, yet these concepts, as applied to AI systems so far, only bear a superficial resemblance to their biological counterparts.

The long-term aims of this workshop are: * to promote work that incorporates insights from human cognition to suggest novel and improved AI architectures; * to facilitate the development of ML methods that can better predict human behavior; and * to support the development of a field of ‘cognitive computing’ that is more than a marketing slogan—a field that improves on both natural and artificial cognition by synergistically advancing each and integrating their objectives.

The workshop will focus on three interconnected topics of particular relevance to ML:

1. **Learning and development**. Cognitive capabilities expressed early in a child’s development are likely to be crucial for bootstrapping adult learning and intelligence. Intuitive physics and intuitive psychology allow the developing organism to build an understanding of the world and of other agents. Additionally, children and adults often demonstrate “learning-to-learn,” where previous concepts and skills form a compositional basis for learning new concepts and skills.

2. **Memory**. Human memory operates on multiple time scales, from memories that literally persist for the blink of an eye to those that persist for a lifetime. These different forms of memory serve different computational purposes. Although forgetting is typically thought of as a disadvantage, the ability to selectively forget/override irrelevant knowledge in nonstationary environments is highly desirable.

3. **Attention and Decision Making**. These refer to relatively high-level cognitive functions that allow task demands to purposefully control an agent’s external environment and sensory data stream, dynamically reconfigure internal representation and architecture, and devise action plans that strategically trade off multiple, oft-conflicting behavioral objectives.

The long-term aims of this workshop are: * to promote work that incorporates insights from human cognition to suggest novel and improved AI architectures; * to facilitate the development of ML methods that can better predict human behavior; and * to support the development of a field of ‘cognitive computing’ that is more than a marketing slogan—a field that improves on both natural and artificial cognition by synergistically advancing each and integrating their strengths in complementary manners.

### Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30 AM</td>
<td>Workshop overview</td>
</tr>
<tr>
<td>08:40 AM</td>
<td>Cognitive AI</td>
</tr>
<tr>
<td>09:05 AM</td>
<td>Computational modeling of human face processing</td>
</tr>
<tr>
<td>09:30 AM</td>
<td>People infer object shape in a 3D, object-centered coordinate system</td>
</tr>
<tr>
<td>09:55 AM</td>
<td>Relational neural expectation maximization</td>
</tr>
<tr>
<td>10:10 AM</td>
<td>Contextual dependence of human preference for complex objects: A Bayesian statistical account</td>
</tr>
<tr>
<td>Time</td>
<td>Session</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>10:15 AM</td>
<td>A biologically-inspired sparse, topographic recurrent neural network model for robust change detection</td>
</tr>
<tr>
<td>10:20 AM</td>
<td>Visual attention guided deep imitation learning</td>
</tr>
<tr>
<td>10:25 AM</td>
<td>Human learning of video games</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>COFFEE BREAK AND POSTER SESSION</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Life history and learning: Extended human childhood as a way to resolve explore/exploit trade-offs and improve hypothesis search</td>
</tr>
<tr>
<td>10:45 AM</td>
<td>Meta-reinforcement learning in brains and machines</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Revealing human inductive biases and metacognitive processes with rational models</td>
</tr>
<tr>
<td>11:15 AM</td>
<td>Learning to select computations</td>
</tr>
<tr>
<td>12:15 PM</td>
<td>Mapping the spatio-temporal dynamics of cognition in the human brain</td>
</tr>
<tr>
<td>12:30 PM</td>
<td>Scale-invariant temporal memory in AI</td>
</tr>
<tr>
<td>12:45 PM</td>
<td>Scale-invariant temporal history (SITH): Optimal slicing of the past in an uncertain world</td>
</tr>
<tr>
<td>1:00 PM</td>
<td>COFFEE BREAK AND POSTER SESSION II</td>
</tr>
<tr>
<td>1:00 PM</td>
<td>From deep learning of disentangled representations to higher-level cognition</td>
</tr>
<tr>
<td>1:15 PM</td>
<td>Access consciousness and actionable representations</td>
</tr>
<tr>
<td>1:30 PM</td>
<td>Evaluating the capacity to reason about beliefs</td>
</tr>
<tr>
<td>1:45 PM</td>
<td>Efficient human-like semantic representations via the information bottleneck principle</td>
</tr>
<tr>
<td>2:00 PM</td>
<td>The mutation sampler: A sampling approach to causal representation</td>
</tr>
<tr>
<td>2:15 PM</td>
<td>Generating more human-like recommendations with a cognitive model of generalization</td>
</tr>
<tr>
<td>2:30 PM</td>
<td>POSTER: Improving transfer using augmented feedback in progressive neural networks</td>
</tr>
<tr>
<td>2:45 PM</td>
<td>POSTER: Sample-efficient reinforcement learning through transfer and architectural priors</td>
</tr>
<tr>
<td>3:00 PM</td>
<td>POSTER: Variational probability flow for biologically plausible training of deep neural networks</td>
</tr>
<tr>
<td>3:15 PM</td>
<td>POSTER: Curiosity-driven reinforcement learning with hoemostatic regulation</td>
</tr>
<tr>
<td>3:30 PM</td>
<td>POSTER: Context-modulation of hippocampal dynamics and deep convolutional networks</td>
</tr>
<tr>
<td>3:45 PM</td>
<td>POSTER: Cognitive modeling and the wisdom of the crowd</td>
</tr>
<tr>
<td>4:00 PM</td>
<td>POSTER: Learning to organize knowledge with N-gram machines</td>
</tr>
<tr>
<td>4:15 PM</td>
<td>POSTER: Pre-training attentional mechanisms</td>
</tr>
<tr>
<td>4:30 PM</td>
<td>POSTER: Concept acquisition through meta-learning</td>
</tr>
<tr>
<td>4:45 PM</td>
<td>POSTER: Using STDP for unsupervised, event-based online learning</td>
</tr>
<tr>
<td>4:50 PM</td>
<td>POSTER: Power-law temporal discounting over a logarithmically compressed timeline for scale invariant reinforcement learning</td>
</tr>
<tr>
<td>5:05 PM</td>
<td>POSTER: Performance-optimized hierarchical models only partially predict neural responses during perceptual decision making</td>
</tr>
<tr>
<td>5:20 PM</td>
<td>Object-oriented intelligence</td>
</tr>
<tr>
<td>5:35 PM</td>
<td>Representational primitives, in minds and machines</td>
</tr>
</tbody>
</table>
Machine Learning in Computational Biology

James Zou, Anshul Kundaje, Gerald Quon, Nicolo Fusi, Sara Mostafavi

104-C, Sat Dec 09, 08:00 AM

The field of computational biology has seen dramatic growth over the past few years. A wide range of high-throughput technologies developed in the last decade now enable us to measure parts of a biological system at various resolutions—at the genome, epigenome, transcriptome, and proteome levels. These technologies are now being used to collect data for an ever-increasingly diverse set of problems, ranging from classical problems such as predicting differentially regulated genes between time points and predicting subcellular localization of RNA and proteins, to models that explore complex mechanistic hypotheses bridging the gap between genetics and disease, population genetics and transcriptional regulation. Fully realizing the scientific and clinical potential of these data requires developing novel supervised and unsupervised learning methods that are scalable, can accommodate heterogeneity, are robust to systematic noise and confounding factors, and provide mechanistic insights.

The goals of this workshop are to i) present emerging problems and innovative machine learning techniques in computational biology, and ii) generate discussion on how to best model the intricacies of biological data and synthesize and interpret results in light of the current work in the field. We will invite several leaders at the intersection of computational biology and machine learning who will present current research problems in computational biology and lead these discussions based on their own research and experiences. We will also have the usual rigorous screening of contributed talks on novel learning approaches in computational biology. We encourage contributions describing either progress on new bioinformatics problems or work on established problems using methods that are substantially different from established alternatives. Deep learning, kernel methods, graphical models, feature selection, non-parametric models and other techniques applied to relevant bioinformatics problems would all be appropriate for the workshop. We will also encourage contributions to address new challenges in analyzing data generated from gene editing, single cell genomics and other novel technologies. The targeted audience are people with interest in machine learning and applications to relevant problems from the life sciences, including NIPS participants without any existing research link to computational biology. Many of the talks will be of interest to the broad machine learning community.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Speaker(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00 AM</td>
<td>Poster session</td>
<td>Wang, Earnshaw, Boss, Gibson, Levin, Bhat, Rampasek, Shim, Eraslan, Ahsan, Rahmani, Shrikumar, Long, Zhong, Heckerman, CHANG, Goldenberg, sinai, Sankararaman, Park, Grinek, Bepler, Chikina, Lai, Killoran, Järvenpää</td>
</tr>
</tbody>
</table>

BigNeuro 2017: Analyzing brain data from nano to macroscale

Eva Dyer, Greg Kiar, Will Gray Roncal, Konrad P Koerding, Joshua T Vogelstein

201-A, Sat Dec 09, 08:00 AM

Datasets in neuroscience are increasing in size at alarming rates relative to our ability to analyze them. This workshop aims at discussing new frameworks for processing and making sense of large neural datasets.

The morning session will focus on approaches for processing large neuroscience datasets. Examples include: distributed + high-performance computing, GPU and other hardware accelerations, spatial databases and other compression schemes used for large neuroimaging datasets, online machine learning approaches for handling large data sizes, randomization and stochastic optimization.

The afternoon session will focus on abstractions for modelling large neuroscience datasets. Examples include graphs, graphical models, manifolds, mixture models, latent variable models, spatial models, and factor learning.

In addition to talks and discussions, we plan to have papers submitted and peer reviewed. Workshop "proceedings" will consist of links to unpublished arXiv or bioarXiv papers that are of exceptional quality and are well aligned with the workshop scope. Some accepted papers will also be invited for an oral presentation; the remaining authors will be invited to present a poster.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:40 AM</td>
<td>Opening Remarks</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>Session 1 Keynote</td>
</tr>
<tr>
<td>09:35 AM</td>
<td>Session 1 Invited Talk</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Session 1 Lightning Talks/Discussion Panel</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Coffee Break</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Session 2 Keynote</td>
</tr>
<tr>
<td>11:35 AM</td>
<td>Session 2 Invited Talk</td>
</tr>
<tr>
<td>12:00 PM</td>
<td>Session 2 Lightning Talks/Discussion Panel</td>
</tr>
<tr>
<td>12:30 PM</td>
<td>Lunch Break</td>
</tr>
<tr>
<td>01:30 PM</td>
<td>Session 3 Keynote</td>
</tr>
<tr>
<td>02:05 PM</td>
<td>Session 3 Invited Talk</td>
</tr>
<tr>
<td>02:30 PM</td>
<td>Session 3 Contributed Talk</td>
</tr>
<tr>
<td>02:40 PM</td>
<td>Session 3 Invited Talk</td>
</tr>
<tr>
<td>03:00 PM</td>
<td>Session 3 Lightning Talks/Discussion Panel</td>
</tr>
<tr>
<td>03:30 PM</td>
<td>Poster Session</td>
</tr>
</tbody>
</table>
Our transportation systems are poised for a transformation as we make progress on autonomous vehicles, vehicle-to-vehicle (V2V) and vehicle-to-everything (V2X) communication infrastructures, and smart road infrastructures such as smart traffic lights. There are many challenges in transforming our current transportation systems to the future vision. For example, how to make perception accurate and robust to accomplish safe autonomous driving? How to learn long term driving strategies (known as driving policies) so that autonomous vehicles can be equipped with adaptive human negotiation skills when merging, overtaking and giving way, etc? how do we achieve near-zero fatality? How do we optimize efficiency through intelligent traffic management and control of fleets? How do we optimize for traffic capacity during rush hours? To meet these requirements in safety, efficiency, control, and capacity, the systems must be automated with intelligent decision making.

Machine learning will be essential to enable intelligent transportation systems. Machine learning has made rapid progress in self-driving, e.g. real-time perception and prediction of traffic scenes, and has started to be applied to ride-sharing platforms such as Uber (e.g. demand forecasting) and crowd-sourced video scene analysis companies such as Nexar (understanding and avoiding accidents). To address the challenges arising in our future transportation system such as traffic management and safety, we need to consider the transportation systems as a whole rather than solving problems in isolation. New machine learning solutions are needed as transportation places specific requirements such as extremely low tolerance on uncertainty and the need to intelligently coordinate self-driving cars through V2V and V2X.

The goal of this workshop is to bring together researchers and practitioners from all areas of intelligent transportations systems to address core challenges with machine learning. These challenges include, but are not limited to accurate and efficient pedestrian detection, pedestrian intent detection, machine learning for object tracking, unsupervised representation learning for autonomous driving, deep reinforcement learning for learning driving policies, cross-modal and simulator to real-world transfer learning, scene classification, real-time perception and prediction of traffic scenes, uncertainty propagation in deep neural networks, efficient inference with deep neural networks predictive modeling of risk and accidents through telematics, modeling, simulation and forecast of demand and mobility patterns in large scale urban transportation systems.

The workshop will include invited speakers, panels, presentations of accepted papers and posters. We invite papers in the form of short, long and position papers to address the core challenges mentioned above. We encourage researchers and practitioners on self-driving cars, transportation systems and ride-sharing platforms to participate. Since this is a topic of broad and current interest, we expect at least 150 participants from leading university researchers, auto-companies and ride-sharing companies.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event Description</th>
<th>Presenter(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:45 AM</td>
<td>Opening Remarks</td>
<td>Li</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>Invited Talk: Raquel Urtasun, Uber ATG and University of Toronto</td>
<td>Urtasun</td>
</tr>
<tr>
<td>09:30 AM</td>
<td>Pieter Abbeel, UC Berkeley and Embodied Intelligence</td>
<td>Abbeel</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Marin Kobilarov, Zoox</td>
<td>Kobilarov</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Coffee</td>
<td></td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Hesham M. Eraqi, Mohamed N. Moustafa, Jens Honer, End-to-End Deep Learning for Steering Autonomous Vehicles Considering Temporal Dependencies</td>
<td></td>
</tr>
<tr>
<td>11:10 AM</td>
<td>Abhinav Jauhri (CMU), Carlee Joe-Wong, John Paul Shen, On the Real-time Vehicle Placement Problem</td>
<td>Jauhri</td>
</tr>
<tr>
<td>11:20 AM</td>
<td>Andrew Best (UNC), Sahil Narang, Lucas Pasqualin, Daniel Barber, Dinesh Manocha, AutonoVi-Sim: Autonomous Vehicle Simulation Platform with Weather, Sensing, and Traffic control</td>
<td></td>
</tr>
<tr>
<td>11:30 AM</td>
<td>Nikita Japuria (MIT), Golnaz Habibi, Jonathan P. How, CASNSC: A context-based approach for accurate pedestrian motion prediction at intersections</td>
<td>Jaipuria</td>
</tr>
<tr>
<td>11:40 AM</td>
<td>Mennatullah Siam (Alberta), Heba Mahgoub, Mohamed Zahran, Senthil Yogamani, Martin Jagersand, Ahmad El-Sallab, Motion and appearance based Multi-Task Learning network for autonomous driving</td>
<td>Siam</td>
</tr>
</tbody>
</table>
Many algorithms in machine learning, computer vision, physical simulation, and other fields require the calculation of gradients and other derivatives. Manual derivation of gradients can be time consuming and error-prone. Automatic differentiation comprises a set of techniques to calculate the derivative of a numerical computation expressed as a computer program. These techniques are commonly used in atmospheric sciences and computational fluid dynamics, and have more recently also been adopted by machine learning researchers. Practitioners across many fields have built a wide set of automatic differentiation tools, using different programming languages, computational primitives and intermediate compiler representations. Each of these choices comes with positive and negative trade-offs, in terms of their usability, flexibility and performance in specific domains. This workshop will bring together researchers in the fields of automatic differentiation and machine learning to discuss ways in which advanced automatic differentiation frameworks and techniques can enable more advanced machine learning models, run large-scale machine learning on accelerators with better performance, and increase the usability of machine learning frameworks for practitioners. Topics for discussion will include: - What abstractions (languages, kernels, interfaces, instruction sets) do we need to develop advanced automatic differentiation frameworks for the machine learning ecosystem? - What different use cases exist in machine learning, from large-scale performance-critical models to small prototypes, and how should our toolsets reflect these needs? - What advanced techniques from the automatic differentiation literature, such as checkpointing, differentiating through iterative processes or chaotic systems, cross-country elimination, etc., could be adopted by the ML community to enable research on new models? - How can we foster greater collaboration between the fields of machine learning and automatic differentiation?

Abstract 18: Micro-Perception Approach to Intelligent Transport, Ramesh Sarukkai (Lyft) in 2017 NIPS Workshop on Machine Learning for Intelligent Transportation Systems, Sarukkai 02:30 PM

In this talk, we will focus on the broader angle of applying machine learning to different aspects of transportation - ranging from traffic congestion, real-time speed estimation, image based localization, and active map making as examples. In particular, as we grow the portfolio of models, we see an unique opportunity in building out a unified framework with a number of micro-perception services for intelligent transport which allows for portability and optimization across multiple transport use cases. We also discuss implications for existing ride-sharing transport as well as potential impact to autonomous.

The future of gradient-based machine learning software & techniques

Alex Wiltschko, Bart van Merriënboer, Pascal Lamblin

202, Sat Dec 09, 08:00 AM

Many algorithms in machine learning, computer vision, physical simulation, and other fields require the calculation of gradients and other derivatives. Manual derivation of gradients can be time consuming and error-prone. Automatic differentiation comprises a set of techniques to calculate the derivative of a numerical computation expressed as a computer program. These techniques are commonly used in atmospheric sciences and computational fluid dynamics, and have more recently also been adopted by machine learning researchers. Practitioners across many fields have built a wide set of automatic differentiation tools, using different programming languages, computational primitives and intermediate compiler representations. Each of these choices comes with positive and negative trade-offs, in terms of their usability, flexibility and performance in specific domains. This workshop will bring together researchers in the fields of automatic differentiation and machine learning to discuss ways in which advanced automatic differentiation frameworks and techniques can enable more advanced machine learning models, run large-scale machine learning on accelerators with better performance, and increase the usability of machine learning frameworks for practitioners. Topics for discussion will include: - What abstractions (languages, kernels, interfaces, instruction sets) do we need to develop advanced automatic differentiation frameworks for the machine learning ecosystem? - What different use cases exist in machine learning, from large-scale performance-critical models to small prototypes, and how should our toolsets reflect these needs? - What advanced techniques from the automatic differentiation literature, such as checkpointing, differentiating through iterative processes or chaotic systems, cross-country elimination, etc., could be adopted by the ML community to enable research on new models? - How can we foster greater collaboration between the fields of machine learning and automatic differentiation?

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00 AM</td>
<td>Introduction and opening remarks</td>
</tr>
</tbody>
</table>

Wiltschko
**NIPS 2017 Workshop book**

**Beyond backprop:**
automatic differentiation in machine learning  
*Baydin*

**09:00 AM**
**Automatic differentiation in PyTorch**  
*Paszke*

**09:10 AM**
**Optimal Smoothing for Pathwise Adjoint**  
*Hüser*

**10:30 AM**
**Morning Coffee Break**

**11:00 AM**
**Poster session**  
*Naumann, Schwartz, Wei, Meissner, Druce, Lin, Pothen*

**11:40 AM**
**30 years of work on automatic differentiation: advanced autodiff techniques**  
*Utke*

**01:40 PM**
**Some highlights on Source-to-Source Adjoint AD**  
*HASCOET*

**02:20 PM**
**Afternoon Coffee Break**

**03:00 PM**
**Divide-and-Conquer Checkpointing for Arbitrary Programs with No User Annotation**  
*Siskind*

**04:00 PM**
**Automatic Differentiation of Parallelised Convolutional Neural Networks - Lessons from Adjoint PDE Solvers**  
*Hueckelheim*

**04:30 PM**
**Panel discussion**  
*Baydin, Paszke, Hüser, Utke, HASCOET, Siskind, Hueckelheim, Griewank*

---

**For more information visit here**  
For more information [https://www.mltrain.cc/](https://www.mltrain.cc/)

**Schedule**

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Speaker(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00 AM</td>
<td>Lessons learned from implementing Edwards</td>
<td>Tran</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Coding Reinforcement Learning Papers</td>
<td>Zhang</td>
</tr>
<tr>
<td>02:00 PM</td>
<td>Simple and Efficient Implementation of Neural Nets with Automatic Operation Batching</td>
<td>Neubig</td>
</tr>
<tr>
<td>04:35 PM</td>
<td>Lessons learned from implementing MLPACK</td>
<td>Curtin</td>
</tr>
</tbody>
</table>

**Abstracts (1):**

**Abstract 3: Simple and Efficient Implementation of Neural Nets with Automatic Operation Batching in NIPS Highlights (MLTrain), Learn How to code a paper with state of the art frameworks, Neubig 02:00 PM**

In this talk I will talk about how to easily and efficiently develop neural network models for complicated problems such as natural language processing using dynamic neural networks. First, I will briefly explain different paradigms in neural networks: static networks (e.g. TensorFlow), dynamic and eager (e.g. PyTorch), and dynamic and lazy (e.g. DyNet). I will discuss about how to efficiently implement models within dynamic neural networks, including minimizing the number of computations and mini-batching. Then I’ll introduce our recently proposed method for automatic batching in dynamic networks, which makes it much easier to implement complicated networks efficiently. Code examples for the implementation will be provided.

---

**NIPS Highlights (MLTrain), Learn How to code a paper with state of the art frameworks**

*Alex Dimakis, Nikolaos Vasiloglou, Guy Van den Broeck, Alexander Ihier, Assaf Araki*

203, Sat Dec 09, 08:00 AM

Every year hundreds of papers are published at NIPS. Although the authors provide sound and scientific description and proof of their ideas, there is no space for explaining all the tricks and details that can make the implementation of the paper work. The goal of this workshop is to help authors evangelize their paper to the industry and expose the participants to all the Machine Learning/Artificial Intelligence know-how that cannot be found in the papers. Also the effect/importance of tuning parameters is rarely discussed, due to lack of space.

**Submissions**

We encourage you to prepare a poster of your favorite paper that explains graphically and at a higher level the concepts and the ideas discussed in it. You should also submit a jupyter notebook that explains in detail how equations in the paper translate to code. You are welcome to use any of the famous platforms like tensorFlow, Keras, MxNet, CNTK, etc.

**Machine Learning Challenges as a Research Tool**

*Isabelle Guyon, Evelyne Viegas, Sergio Escalera, Jacob D Abernethy*

204, Sat Dec 09, 08:00 AM

Challenges in machine learning and data science are competitions running over several weeks or months to resolve problems using provided datasets or simulated environments. The playful nature of challenges naturally attracts students, making challenge a great teaching resource. For this fourth edition of the CiML workshop at NIPS we want to explore the impact of machine learning challenges as a research tool. The workshop will give a large part to discussions around several axes: (1) benefits and limitations of challenges as a research tool; (2) methods to induce and train young researchers; (3) experimental design to foster contributions that will push the state of the art.
CiML is a forum that brings together workshop organizers, platform providers, and participants to discuss best practices in challenge organization and new methods and application opportunities to design high impact challenges. Following the success of last year’s workshop, in which a fruitful exchange led to many innovations, we propose to reconvene and discuss new opportunities for challenges as a research tool, one of the hottest topics identified in last year's discussions. We have invited prominent speakers in this field. We will also reserve time to an open discussion to dig into other topic including open innovation, collaborative challenges (cooperations), platform interoperability, and tool mutualisation. The audience of this workshop is targeted to workshop organizers, participants, and anyone with scientific problem involving machine learning, which may be formulated as a challenge. The emphasis of the workshop is on challenge design. Hence it complements nicely the workshop on the NIPS 2017 competition track and will help paving the way toward next year’s competition program.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Speaker(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:00 AM</td>
<td>Introduction - Isabelle Guyon and Evelyne Viegas</td>
<td>Guyon</td>
</tr>
<tr>
<td>08:10 AM</td>
<td>Kaggle platform - Ben Hamner</td>
<td></td>
</tr>
<tr>
<td>08:40 AM</td>
<td>Automatic evaluation of chatbots - Mikhail Burtsev</td>
<td></td>
</tr>
<tr>
<td>08:40 AM</td>
<td>TrackML - David Rousseau</td>
<td></td>
</tr>
<tr>
<td>09:40 AM</td>
<td>Data Science Bowl - Andre Farris</td>
<td>Farris</td>
</tr>
<tr>
<td>10:10 AM</td>
<td>CrowdAI - Mohanty Sarada</td>
<td></td>
</tr>
<tr>
<td>10:40 AM</td>
<td>Break, poster viewing 1</td>
<td></td>
</tr>
<tr>
<td>11:00 AM</td>
<td>RAMP platform - Balázs Kégl</td>
<td></td>
</tr>
<tr>
<td>11:30 AM</td>
<td>Incentivizing productivity in data science - Jacob Abernethy</td>
<td></td>
</tr>
<tr>
<td>12:20 PM</td>
<td>Lunch Break, poster viewing 2</td>
<td></td>
</tr>
<tr>
<td>12:20 PM</td>
<td>Establishing Uniform Image Segmentation - Maximilian Chen and Michael C. Darling</td>
<td>Chen, Darling</td>
</tr>
<tr>
<td>01:30 PM</td>
<td>Welcome - Isabelle Guyon and Evelyne Viegas</td>
<td></td>
</tr>
<tr>
<td>01:40 PM</td>
<td>Project Malmo, Minecraft - Katja Hofmann</td>
<td></td>
</tr>
<tr>
<td>02:10 PM</td>
<td>Project Alloy - Laura Seaman and Israel Ridgley</td>
<td>Ridgley, Seaman</td>
</tr>
<tr>
<td>02:40 PM</td>
<td>Education and public service - Jonathan C. Stroud</td>
<td></td>
</tr>
<tr>
<td>03:10 PM</td>
<td>Break, poster viewing 3</td>
<td></td>
</tr>
<tr>
<td>03:30 PM</td>
<td>AutoDL (Google challenge) - Olivier Bousquet</td>
<td></td>
</tr>
</tbody>
</table>

Deep Learning: Bridging Theory and Practice

Sanjeev Arora, Maithra Raghu, Russ Salakhutdinov, Ludwig Schmidt, Oriol Vinyals

Grand Ballroom A, Sat Dec 09, 08:00 AM

The past five years have seen a huge increase in the capabilities of deep neural networks. Maintaining this rate of progress however, faces some steep challenges, and awaits fundamental insights. As our models become more complex, and venture into areas such as unsupervised learning or reinforcement learning, designing improvements becomes more laborious, and success can be brittle and hard to transfer to new settings.

This workshop seeks to highlight recent works that use “theory as well as systematic experiments” to isolate the fundamental questions that need to be addressed in deep learning. These have helped flesh out core questions on topics such as generalization, adversarial robustness, large batch training, generative adversarial nets, and optimization, and point towards elements of the theory of deep learning that is expected to emerge in the future.

The workshop aims to enhance this confluence of theory and practice, highlighting influential work with these methods, future open directions, and core fundamental problems. There will be an emphasis on discussion, via panels and round tables, to identify future research directions that are promising and tractable.

Hierarchical Reinforcement Learning

Andrew G Barto, Doina Precup, Shie Mannor, Tom Schaul, Roy Fox, Carlos Florensa Campo

Grand Ballroom B, Sat Dec 09, 08:00 AM

Reinforcement Learning (RL) has become a powerful tool for tackling complex sequential decision-making problems. It has been shown to train agents to reach super-human capabilities in game-playing domains such as Go and Atari. RL can also learn advanced control policies in high-dimensional robotic systems. Nevertheless, current RL agents have considerable difficulties when facing sparse rewards, long planning...
Hierarchical Reinforcement Learning (HRL) is emerging as a key component for finding spatio-temporal abstractions and behavioral patterns that can guide the discovery of useful large-scale control architectures, both for deep-network representations and for analytic and optimal-control methods. HRL has the potential to accelerate planning and exploration by identifying skills that can reliably reach desirable future states. It can abstract away the details of low-level controllers to facilitate long-horizon planning and meta-learning in a high-level feature space. Hierarchical structures are modular and amenable to separation of training efforts, reuse, and transfer. By imitating a core principle of human cognition, hierarchies hold promise for interpretability and explainability.

There is a growing interest in HRL methods for structure discovery, planning, and learning, as well as HRL systems for shared learning and policy deployment. The goal of this workshop is to improve cohesion and planning architectures, both for deep-network representations and for analytic and optimal-control methods. HRL has the potential to accelerate planning and exploration by identifying skills that can reliably reach desirable future states. It can abstract away the details of low-level controllers to facilitate long-horizon planning and meta-learning in a high-level feature space. Hierarchical structures are modular and amenable to separation of training efforts, reuse, and transfer. By imitating a core principle of human cognition, hierarchies hold promise for interpretability and explainability.

Bayesian Deep Learning

Yarin Gal, José Miguel Hernández-Lobato, Christos Louizos, Andrew G Wilson, Diederik P. (Durk) Kingma, Zoubin Ghahramani, Kevin P Murphy, Max Welling

Hall C, Sat Dec 09, 08:00 AM

While deep learning has been revolutionary for machine learning, most modern deep learning models cannot represent their uncertainty nor take advantage of the well studied tools of probability theory. This has started to change following recent developments of tools and techniques combining Bayesian approaches with deep learning. The intersection of the two fields has received great interest from the community over the past few years, with the introduction of new deep learning models that take advantage of Bayesian techniques, as well as Bayesian models that incorporate deep learning elements [1-11]. In fact, the use of Bayesian techniques in deep learning can be traced back to the 1990s, in seminal works by Radford Neal [12], David MacKay [13], and Dayan et al. [14]. These gave us tools to reason about deep models’ confidence, and achieved state-of-the-art performance on many tasks. However earlier tools did not adapt when new needs arose (such as scalability to big data), and were consequently forgotten. Such ideas are now being revisited in light of new advances in the field, yielding many exciting new results.

Extending on last year’s workshop’s success, this workshop will again study the advantages and disadvantages of such ideas, and will be a platform to host the recent flourish of ideas using Bayesian approaches in deep learning and using deep learning tools in Bayesian modelling. The program includes a mix of invited talks, contributed talks, and contributed posters. It will be composed of five main themes: deep generative models, variational inference using neural network recognition models, practical approximate inference techniques in Bayesian neural networks, applications of Bayesian neural networks, and information theory in deep learning. Future directions for the field will be debated in a panel discussion.

Topics:
Probabilistic deep models for classification and regression (such as extensions and application of Bayesian neural networks),
Generative deep models (such as variational autoencoders),
Incorporating explicit prior knowledge in deep learning (such as posterior regularization with logic rules),
Approximate inference for Bayesian deep learning (such as variational Bayes / expectation propagation / etc. in Bayesian neural networks),
Scalable MCMC inference in Bayesian deep models,
Deep recognition models for variational inference (amortized inference),
Model uncertainty in deep learning,
Bayesian deep reinforcement learning,
Deep learning with small data,
Deep learning in Bayesian modelling, Probabilistic semi-supervised learning techniques, Active learning and Bayesian optimization for experimental design, Applying non-parametric methods, one-shot learning, and Bayesian deep learning in general, Implicit inference, Kernel methods in Bayesian deep learning.

References:
[8] - Balan, AK, Rathod, V, Murphy, KP, and Welling, M, ”Bayesian dark knowledge”, 2015.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:05 AM</td>
<td>Deep Probabilistic Programming</td>
<td>Tran</td>
</tr>
<tr>
<td>08:30 AM</td>
<td>TBD 1</td>
<td></td>
</tr>
<tr>
<td>08:45 AM</td>
<td>Automatic Model Selection in BBNs with Horseshoe Priors</td>
<td>Doshi-Velez</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>Deep Bayes for Distributed Learning, Uncertainty Quantification and Compression</td>
<td>Welling</td>
</tr>
<tr>
<td>09:40 AM</td>
<td>Poster spots</td>
<td></td>
</tr>
</tbody>
</table>

Interpreting, Explaining and Visualizing Deep Learning - Now what?

Klaus-Robert Müller, Andrea Vedaldi, Lars K Hansen, Wojciech Samek, Grégoire Montavon

Hyatt Beacon Ballroom D+E+F+H, Sat Dec 09, 08:00 AM

Machine learning has become an indispensable tool for a number of tasks ranging from the detection of objects in images to the understanding of natural languages. While these models reach impressively high predictive accuracy, they are often perceived as black-boxes, and it is not clear what information in the input data is used for predicting. In sensitive applications such as medical diagnosis or self-driving cars, where a single incorrect prediction can be very costly, the reliance of the model on the right features must be guaranteed. This indeed lowers the risk that the model behaves erroneously in presence of novel factors of variation in the test data. Furthermore, interpretability is instrumental when applying machine learning to the sciences, as the detailed understanding of the trained model (e.g., what features it uses to capture the complex relations between physical or biological variables) is a prerequisite for building meaningful new scientific hypotheses. Without such understanding and the possibility of verification that the model has learned something meaningful (e.g. obeying the known physical or biological laws), even the best predictor is of no use for scientific purposes. Finally, also from the perspective of a deep learning engineer,
being able to visualize what the model has (or has not) learned is valuable as it allows to improve current models by e.g. identifying biases in the data or the training procedure, or by comparing the strengths and weaknesses of different architectures.

Not surprisingly, the problem of visualizing and understanding neural networks has recently received a lot of attention in the community. Various techniques for interpreting deep neural networks have been proposed and several workshops have been organized on related topics. However, the theoretical foundations of the interpretability problem are yet to be investigated and the usefulness of the proposed methods in practice still needs to be demonstrated.

Our NIPS 2017 Workshop “Interpreting, Explaining and Visualizing Deep Learning – Now what?” aims to review recent techniques and establish new theoretical foundations for interpreting and understanding deep learning models. However, it will not stop at the methodological level, but also address the “now what?” question. This strong focus on the applications of interpretable methods in deep learning distinguishes this workshop from previous events as we aim to take the next step by exploring and extending the practical usefulness of Interpreting, Explaining and Visualizing in Deep Learning. Also with this workshop we aim to identify new fields of applications for interpretable deep learning. Since the workshop will host invited speakers from various application domains (computer vision, NLP, neuroscience, medicine), it will provide an opportunity for participants to learn from each other and initiate new interdisciplinary collaborations. The workshop will contain invited research talks, short methods and applications talks, a poster and demonstration session and a panel discussion. A selection of accepted papers together with the invited contributions will be published in an edited book by Springer LNCS in order to provide a representative overview of recent activities in this emerging research field.

Schedule

08:15 AM  Opening Remarks  Müller
08:45 AM  Invited Talk 1  Kim
09:15 AM  Invited Talk 2  Batra
09:45 AM  Methods 1  Montavon, Tsang, Ancona
10:30 AM  Coffee Break (morning)
11:00 AM  Methods 2  Kindermans
11:15 AM  Invited Talk 3  Hochreiter
11:45 AM  Posters 1  Lewis, Khalifa Bashier Babiker, Qi, Rieger, Xie, Dabek, NAGASUBRAMANIAN, Zhou, Hupkes, CHANG, Douglas, Ceolini, Doran, Liu, Li
12:15 PM  Lunch
01:15 PM  Posters 2
01:45 PM  Invited Talk 4  Nguyen
02:15 PM  Invited Talk 5  Lee
02:45 PM  Applications 1  Samek

Learning with Limited Labeled Data: Weak Supervision and Beyond

Isabelle Augenstein, Stephen Bach, Eugene Belilovsky, Matthew Blaschko, Christoph Lampert, Edouard Oyallon, Emmanouil Antonios Platanios, Alexander Ratner, Chris Ré

Hyatt Regency Ballroom A+B+C, Sat Dec 09, 08:00 AM

Modern representation learning techniques like deep neural networks have had a major impact both within and beyond the field of machine learning, achieving new state-of-the-art performances with little or no feature engineering on a vast array of tasks. However, these gains are often difficult to translate into real-world settings as they require massive hand-labeled training sets. And in the vast majority of real-world settings, collecting such training sets by hand is infeasible due to the cost of labeling data or the paucity of data in a given domain (e.g. rare diseases in medical applications). In this workshop we focus on techniques for few sample learning and using weaker supervision when large unlabeled datasets are available, as well as theory associated with both.

One increasingly popular approach is to use weaker forms of supervision—i.e. supervision that is potentially noisier, biased, and/or less precise. An overarching goal of such approaches is to use domain knowledge and resources from subject matter experts, but to solicit it in higher-level, lower-fidelity, or more opportunistic ways. Examples include higher-level abstractions such as heuristic labeling rules, feature annotations, constraints, expected distributions, and generalized expectation criteria; noisier or biased labels from distant supervision, annotations, constraints, expected distributions, and generalized expectation criteria; noisier or biased labels from distant supervision, crowd workers, and weak classifiers; data augmentation strategies to express class invariances; and potentially mismatched training data such as in multitask and transfer learning settings.

Along with practical methods and techniques for dealing with limited labeled data settings, this workshop will also focus on the theory of learning in this general setting. Although several classic techniques in the statistical learning theory exist which handle the case of few samples and high dimensions, extending these results for example to the recent success of deep learning is still a challenge. How can the theory or the techniques that have gained success in deep learning be adapted to the case of limited labeled data? How can systems designed (and potentially deployed) for large scale learning be adapted to small data settings? What are efficient and practical ways to incorporate prior knowledge?

This workshop will focus on highlighting both practical and theoretical aspects of learning with limited labeled data, including but not limited to topics such as:
- Learning from noisy labels
- “Distant” or heuristic supervision
- Non-standard labels such as feature annotations, distributions, and
Learning Disentangled Features: from Perception to Control

Emily Denton, Siddharth Narayanaswamy, Tejas Kulkarni, Honglak Lee, Diane Bouchacourt, Josh Tenenbaum, David Pfau

Hyatt Seaview Ballroom, Sat Dec 09, 08:00 AM

An important facet of human experience is our ability to break down what we observe and interact with, along characteristic lines. Visual scenes consist of separate objects, which may have different poses and identities within their category. In natural language, the syntax and semantics of a sentence can often be separated from one another. In planning and cognition plans can be broken down into immediate and long term goals. Inspired by this much research in deep representation learning has gone into finding disentangled factors of variation. However, this research often lacks a clear definition of what disentangling is or much relation to work in other branches of machine learning, neuroscience or cognitive science. In this workshop we intend to bring a wide swathe of scientists studying disentangled representations under one roof to try to come to a unified view of the problem of disentangling.

The workshop will address these issues through 3 focuses:
What is disentangling: Are disentangled representations just the same as statistically independent representations, or is there something more?
How does disentangling relate to interpretability? Can we define what it means to separate style and content, or is human judgement the final arbiter? Are disentangled representations the same as equivariant representations?
How can disentangled representations be discovered: What is the current state of the art in learning disentangled representations? What are the cognitive and neural underpinnings of disentangled representations in animals and humans? Most work in disentangling has focused on perception, but we will encourage dialogue with researchers in natural language processing and reinforcement learning as well as neuroscientists and cognitive scientists.
Why do we care about disentangling: What are the downstream tasks that can benefit from using disentangled representations? Does the downstream task define the relevance of the disentanglement to learn? What does disentangling get us in terms of improved prediction or behavior in intelligent agents?

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Session/Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:00 AM</td>
<td>Poster session</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Invited Talk: Sebastian Riedel</td>
</tr>
<tr>
<td>11:30 AM</td>
<td>Invited Talk: McCullum</td>
</tr>
<tr>
<td>12:00 PM</td>
<td>Lunch</td>
</tr>
<tr>
<td>02:00 PM</td>
<td>Panel: Integrating Domain Expertise as Weak Supervision</td>
</tr>
<tr>
<td>02:30 PM</td>
<td>1-minute Poster Spotlights (Session #2)</td>
</tr>
<tr>
<td>02:50 PM</td>
<td>Poster Session / Coffee Break</td>
</tr>
<tr>
<td>03:00 PM</td>
<td>Invited Talk: Balcan</td>
</tr>
<tr>
<td>04:00 PM</td>
<td>Contributed Talk 2</td>
</tr>
<tr>
<td>04:15 PM</td>
<td>Invited Talk: Goodfellow</td>
</tr>
<tr>
<td>04:45 PM</td>
<td>Invited Talk: Urtasun</td>
</tr>
<tr>
<td>05:15 PM</td>
<td>Contributed Talk 3</td>
</tr>
<tr>
<td>05:30 PM</td>
<td>Contributed Talk 4</td>
</tr>
<tr>
<td>05:45 PM</td>
<td>Invited Talk: Ritter</td>
</tr>
<tr>
<td>06:15 PM</td>
<td>Closing Remarks &amp; Awards</td>
</tr>
</tbody>
</table>

Hyatt Shoreline, Sat Dec 09, 08:00 AM

In order to be helpful to users and to society at large, an autonomous agent needs to be aligned with the objectives of its stakeholders. Misligned incentives are a common and crucial problem with human agents --- we should expect similar challenges to arise from misaligned incentives with artificial agents. For example, it is not uncommon to see reinforcement learning agents ‘hack’ their specified reward function. How do we build learning systems that will reliably achieve a user’s _intended_ objective? How can we ensure that autonomous agents behave reliably in unforeseen situations? How do we design systems...
Bayesian optimization for science and engineering

Ruben Martinez-Cantin, Jose Miguel Hernández-Lobato, Javier Gonzalez

S-1, Sat Dec 09, 08:00 AM

Bayesian optimization (BO) is a recent subfield of machine learning comprising a collection of methodologies for the efficient optimization of expensive black-box functions. BO techniques work by fitting a model to black-box function data and then using the model's predictions to decide where to collect data next, so that the optimization problem can be solved using only a small number of function evaluations. The resulting methods are characterized by their high sample-efficiency when compared to alternative black-box optimization algorithms, enabling the solution of new challenging problems. For example, in recent years, BO has become a popular tool in the machine learning community for the excellent performance attained in the problem of hyperparameter tuning, with important results both in academia and industry. This success has made BO a crucial player in the current trend of "automatic machine learning". As new BO methods have been developed, the area of applicability has been continuously expanding. While the problem of hyperparameter tuning permeates all disciplines, the field has moved towards more specific problems in science and engineering requiring of new advanced methodology. Today, Bayesian optimization is the most promising approach for accelerating and automating science and engineering. Therefore, we have chosen this year's theme for the workshop to be "Bayesian optimization for science and engineering". We enumerate below a few of the recent directions in which BO methodology is being pushed forward to address specific problems in science and engineering: -Beyond Gaussian processes. While GPs are the default choice in BO, specific problems in science and engineering require to collect larger amounts of complex data. In these settings, it is necessary to use alternative models to capture complex patterns with higher accuracy. For example, Bayesian neural networks or deep Gaussian processes. How can we do efficient BO with these type of models? -Optimization in structured domains. Many problems in science and engineering require to perform optimization in complex spaces which are different from the typical box-constrained subset of the real coordinate space. For example, how can we efficiently optimize over graphs, discrete sequences, trees, computer programmes, etc.? -Safe optimization. Critical applications in science and engineering may include configurations in the search space that may be unsafe or harmful and may lead to system failure. How can we perform efficient BO while avoiding these unsafe settings? -Incorporating domain specific knowledge. In many optimization settings there is available a lot of domain specific information that can be used to build bespoke BO methods with significantly better performance than its off-the-shelf counterparts. How can we easily encode and transfer available knowledge into BO methods in an easy and fast manner? -Optimization with structured output response. In specific cases, each black-box may produce additional output values besides an estimate of the objective function. For example, when tuning a simulator, besides the final output value, we may also obtain structured data related to the execution trace of the simulator. How can design BO methods that automatically exploit this extra structured output? -Scalability and fast evaluations. Recent problems require to collect a massive amount of data in parallel and at cost that is usually lower than in typical BO problems. How can we design efficient BO methods that collect very large batches of data in parallel? How can we automatically adjust the cost of BO methods so that they are efficient even when the data collection is not highly expensive? The target audience for this workshop consists of both industrial and academic practitioners of Bayesian optimization as well as researchers working on theoretical and practical advances in model based optimization across different engineering areas. We expect that this pairing of theoretical and applied knowledge will lead to an interesting exchange of ideas and stimulate an open discussion about the long term goals and challenges of the Bayesian optimization community. The main goal of the workshop is to serve as a forum of discussion and to encourage collaboration between the diverse set of scientist that develop and use Bayesian optimization and related techniques. Researchers and practitioners in Academia are welcome, as well people form the wider optimization, engineering and probabilistic modeling communities.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00 AM</td>
<td>Bayesian Optimization Under Uncertainty</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>The Kalai-Smorodinski solution for many-objective</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>Memory Bandits: a Bayesian Optimization</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>Fast Information-theoretic Bayesian Optimisation</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>Context-Dependent Bayesian Optimization in</td>
</tr>
<tr>
<td>09:10 AM</td>
<td>Invited talk: Andreas Krause</td>
</tr>
<tr>
<td>09:40 AM</td>
<td>Invited talk: Yulant Chen</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Invited talk: David Ginsbourger</td>
</tr>
<tr>
<td>02:00 PM</td>
<td>Invited talk: Romy Lorenz</td>
</tr>
<tr>
<td>04:00 PM</td>
<td>Invited talk: Peter Frazier</td>
</tr>
</tbody>
</table>
Workshop on Prioritising Online Content

John S Shawe-Taylor, Massimiliano Pontil, Nicolò Cesa-Bianchi, Emine Yilmaz, Chris Watkins, Sebastian Riedel, Marko Grobelnik

S-3, Sat Dec 09, 08:00 AM

Social Media and other online media sources play a critical role in distributing news and informing public opinion. Initially it seemed that democratising the dissemination of information and news with online media might be wholly good – but during the last year we have witnessed other perhaps less positive effects. The algorithms that prioritise content for users aim to provide information that will be ‘liked’ by each user in order to retain their attention and interest. These algorithms are now well-tuned and are indeed able to match content to different users’ preferences. This has meant that users increasingly see content that aligns with their world view, confirms their beliefs, supports their opinions, in short that maintains their ‘information bubble’, creating the so-called echo-chambers. As a result, views have often become more polarised rather than less, with people expressing genuine disbelief that fellow citizens could possibly countenance alternative opinions, be they pro- or anti-brexit, pro- or anti-Trump. Perhaps the most extreme example is that of fake news in which news is created in order to satisfy and reinforce certain beliefs. This polarisation of views cannot be beneficial for society. As the success of Computer Science and more specifically Machine Learning have led to this undesirable situation, it is natural that we should now ask how Online Content might be prioritised in such a way that users are still satisfied with an outlet but at the same time are not led to more extreme and polarised opinions.

What is the effect of content prioritisation – and more generally, the effect of the affordances of the social network – on the nature of discussion and debate? Social networks could potentially enable society-wide debate and collective intelligence. On the other hand, they could also encourage communal reinforcement by enforcing conformity within friendship groups, in that it is a daring person who posts an opinion at odds with the majority of their friends. Each design of content prioritisation may nudge users towards particular styles of both content-viewing and of content-posting and discussion. What is the nature of the interaction between content-presentation and users’ viewing and debate?

Content may be prioritised either ‘transparently’ according to users’ explicit choices of what they want to see, combined with transparent community voting, and moderators whose decisions can be questioned (e.g. Reddit). At the other extreme, content may be prioritised by proprietary algorithms that model each user’s preferences and then predict what they want to see. What is the range of possible designs and what are their effects? Could one design intelligent power-tools for moderators?

The online portal Reddit is a rare exception to the general rule in that it has proven a popular site despite employing a more nuanced algorithm for the prioritisation of content. The approach was, however, apparently designed to manage traffic flows rather than create a better balance of opinions. It would, therefore, appear that even for this algorithm its effect on prioritisation is only partially understood or intended.

If we view social networks as implementing a large scale message-passing algorithm attempting to perform inference about the state of the world and possible interventions and/or improvements, the current prioritisation algorithms create many (typically short) cycles. It is well known that inference based on message passing fails to converge to an optimal solution if the underlying graph contains cycles because information then becomes incorrectly weighted. Perhaps a similar situation is occurring with the use of social media? Is it possible to model this phenomenon as an approximate inference task?

The workshop will provide a forum for the presentation and discussion of analyses of online prioritisation with emphasis on the biases that such prioritisations introduce and reinforce. Particular interest will be placed on presentations that consider alternative ways of prioritising content where it can be argued that they will reduce the negative side-effects of current methods while maintaining user loyalty.

Call for contributions - see conference web page via link above.

We will issue a call for contributions highlighting but not restricted to the following themes:

- (*) predicting future global events from media
- (*) detecting and predicting new major trends in the scientific literature
- (*) enhancing content with information from fact checkers
- (*) detection of fake news
- (*) detecting and mitigating tribalism among online personas
- (*) adapted and improved mechanisms of information spreading
- (*) algorithmic fairness in machine learning

Schedule

09:00 AM Automating textual claim verification Viachos

Collaborate & Communicate: An exploration and practical skills workshop that builds on the experience of AIML experts who are both successful collaborators and great communicators.

Katherine Gorman

S-4, Sat Dec 09, 08:00 AM

For many in the sciences, collaboration is a given, or at least a given assumption. The field of AIML is no different, and collaboration across fields and disciplines has long been a source of data and funding. But for many, effective collaboration can be confusing and daunting. Good collaboration requires good communication, but more fundamentally, clear communication is a core skillset for anyone. It takes practice, and in highly specialized fields, it is often subject to an all-too-common malady: the curse of knowledge. The curse of knowledge happens when experts in a field, communicating within their field, begin to make assumptions about the knowledge and understanding of their audience and begin to overlook the fundamentals of clear communication. They do this because for an audience of their peers, they seem to become less necessary, while short cuts like jargon seem to make communication faster and more efficient. But today, clear communication around issues and techniques in machine intelligence...
work is crucial not only within the community, but also to foster collaboration across disciplines, and between the community and the lay public. In this workshop we will explore stories of success in both of these topics through a series of short talks by pairs of collaborators followed by a panel discussion. We will then apply some of the principles from these areas by exploring practical models and skills to aid in both. We intend to work with partners to organize attendees at the beginning of the workshop to create real opportunities for good collaborations among them and to inspire them to consider how they might work together. Attendees will leave with resources to help them more effectively collaborate and to help them organize and communicate about their own work.

**Schedule**

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00 AM</td>
<td>Learning From Experts</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>The Story Algorithm</td>
</tr>
<tr>
<td>11:30 AM</td>
<td>Communications Exercise</td>
</tr>
<tr>
<td>12:00 PM</td>
<td>Lunch</td>
</tr>
<tr>
<td>02:00 PM</td>
<td>Collaboration Exercise</td>
</tr>
<tr>
<td>03:00 PM</td>
<td>Coffee Break</td>
</tr>
<tr>
<td>03:30 PM</td>
<td>Honing Personal Skills</td>
</tr>
</tbody>
</table>

**Synergies in Geometric Data Analysis**

*Marina Meila, Frederic Chazal*

**Medical Imaging meets NIPS**

*Ben Glocker, Ender Konukoglu, Hervé Lombaert, Kanwal Bhatia*

S-7, Sat Dec 09, 08:00 AM

**Scope**

'Medical Imaging meets NIPS' is a satellite workshop at NIPS 2017. The workshop aims to bring researchers together from the medical image computing and machine learning community. The objective is to discuss the major challenges in the field and opportunities for joining forces. The event will feature a series of high-profile invited speakers from industry, academia, engineering and medical sciences who aim to give an overview of recent advances, challenges, latest technology and efforts for sharing clinical data.

**Motivation**

Medical imaging is facing a major crisis with an ever increasing complexity and volume of data and immense economic pressure. The interpretation of medical images pushes human abilities to the limit with the risk that critical patterns of disease go undetected. Machine learning has emerged as a key technology for developing novel tools in computer aided diagnosis, therapy and intervention. Still, progress is slow compared to other fields of visual recognition which is mainly due to the domain complexity and constraints in clinical applications which require most robust, accurate, and reliable solutions.

**Call for Abstracts**

We invite submissions of extended abstracts for poster presentation during the workshop. Submitting an abstract is an ideal way of engaging with the workshop and to showcase research in the area of machine learning for medical imaging. Submitted work does not have to be original and can be already published elsewhere and/or can be of preliminary nature. There will be no workshop proceedings, and the poster session may be conditional on receiving sufficiently many submissions. Accepted abstracts together with author information will be made available on this website.

**Dates**

Submissions: Sunday, October 29th, midnight PST
Notifications: Sunday, November 5th
Workshop: Saturday, December 9th, 8:45 AM - 6 PM

**Schedule (tentative)**

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:45 AM</td>
<td>Opening</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>Industry talk - Siemens</td>
</tr>
<tr>
<td>09:30 AM</td>
<td>Industry talk - IBM</td>
</tr>
</tbody>
</table>
Abstracts (1):

Abstract 1: Opening in Medical Imaging meets NIPS, Glocker, Konukoglu, Lombaert, Bhatia 08:45 AM

Opening remarks by the organizers

Workshop on Meta-Learning

Roberto Calandra, Frank Hutter, Hugo Larochelle, Sergey Levine

Seaside Ballroom, Sat Dec 09, 08:00 AM

Recent years have seen rapid progress in meta-learning methods, which learn (and optimize) the performance of learning methods based on data, generate new learning methods from scratch, and learn to transfer knowledge across tasks and domains. Meta-learning can be seen as the logical conclusion of the arc that machine learning has undergone in the last decade, from learning classifiers, to learning representations, and finally to learning algorithms that themselves acquire representations and classifiers. The ability to improve one’s own learning capabilities through experience can also be viewed as a hallmark of intelligent beings, and there are strong connections with work on human learning in neuroscience.

Meta-learning methods are also of substantial practical interest, since they have, e.g., been shown to yield new state-of-the-art automated machine learning methods, novel deep learning architectures, and substantially improved one-shot learning systems.

Some of the fundamental questions that this workshop aims to address are:
- What are the fundamental differences in the learning "task" compared to traditional "non-meta" learners?
- Is there a practical limit to the number of meta-learning layers (e.g., would a meta-meta-meta-learning algorithm be of practical use)?
- How can we design more sample-efficient meta-learning methods?
- How can we exploit our domain knowledge to effectively guide the meta-learning process?
- What are the meta-learning processes in nature (e.g., in humans), and how can we take inspiration from them?
- Which ML approaches are best suited for meta-learning, in which circumstances, and why?
- What principles can we learn from meta-learning to help us design the next generation of learning systems?

The goal of this workshop is to bring together researchers from all the different communities and topics that fall under the umbrella of meta-learning. We expect that the presence of these different communities will result in a fruitful exchange of ideas and stimulate an open discussion about the current challenges in meta-learning, as well as possible solutions.

In terms of prospective participants, our main targets are machine learning researchers interested in the processes related to understanding and improvingcurrent meta-learning algorithms. Specific target communities within machine learning include, but are not limited to: meta-learning, optimization, deep learning, reinforcement learning, evolutionary computation, Bayesian optimization and AutoML. Our invited speakers also include researchers who study human learning, to provide a broad perspective to the attendees.
Dear NIPS Workshop Chairs, We propose to organize the workshop:

OPT 2017: Optimization for Machine Learning. This year marks a major milestone in the history of OPT, as it will be the 10th anniversary edition of this long running NIPS workshop. The previous OPT workshops enjoyed packed to overpacked attendance. This huge interest is no surprise: optimization is the 2nd largest topic at NIPS and is indeed foundational for the wider ML community. Looking back over the past decade, a strong trend is apparent: The intersection of OPT and ML has grown monotonically to the point that now several cutting-edge advances in optimization arise from the ML community. The distinctive feature of optimization within ML is its departure from textbook approaches, in particular, by having a different set of goals driven by "big-data," where both models and practical implementation are crucial. This intimate relation between OPT and ML is the core theme of our workshop. OPT workshops have previously covered a variety of topics, such as frameworks for convex programs (D. Bertsekas), the intersection of ML and optimization, especially SVM training (S. Wright), large-scale learning via stochastic gradient methods and its tradeoffs (L. Bottou, N. Srebro), exploitation of structured sparsity (Vandenberghhe), randomized methods for extremely large-scale convex optimization (A. Nemirovski), complexity theoretic foundations of convex optimization (Y. Nesterov), distributed large-scale optimization (S. Boyd), asynchronous and sparsity based stochastic gradient (B. Recht), algebraic techniques in machine learning (P. Parrilo), insights into nonconvex optimization (A. Lewis), sums-of-squares techniques (J. Lasserre), optimization in the context of deep learning (Y. Bengio), stochastic convex optimization (G. Lan), new views on interior point (E. Hazan), among others. Several ideas propounded in these talks have become important research topics in ML and optimization --- especially in the field of randomized algorithms, stochastic gradient and variance reduced stochastic gradient methods. An edited book "Optimization for Machine Learning" (S. Sra, S. Nowozin, and S. Wright; MIT Press, 2011) grew out of the first three OPT workshops, and contains high-quality contributions from many of the speakers and attendees, and there have been sustained requests for the next edition of such a volume. We wish to use OPT2017 as a platform to foster discussion, discovery, and dissemination of the state-of-the-art in optimization as relevant to machine learning. And even beyond that, as a platform to identify new directions and challenges that will drive future research. Continuing its trend, the workshop will bring experts in optimization to share their perspectives while leveraging crossover experts in ML to share their views and recent advances. Our tentative invited speakers for this year are: Yurii Nesterov (already agreed) Dimitri Bertsekas (already agreed) Francis Bach (already agreed) Distinction from other optimization workshops at NIPS: Compared to the other optimization focused workshops that happen (or have happened) at NIPS, key distinguishing features of OPT are: (a) it provides a unique bridge between the ML community and the wider optimization community, and is the longest running NIPS workshop on optimization (since NIPS 2008); (b) it encourages theoretical work on an equal footing with practical efficiency; and (c) it caters to a wide body of NIPS attendees, experts and beginners alike; (d) it covers optimization in a broad-spectrum, with a focus on bringing in new optimization ideas from different communities into ML while identifying key future directions for the broader OPTML community. Organization ----------- The main features of the proposed workshop are: 1. One day long with morning and afternoon sessions 2. Four invited talks by leading experts from optimization and ML 3. Contributed talks from the broader OPT and ML community 4. A panel discussion exploring key future research directions for OPTML.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30 AM</td>
<td>Invited Talk 1</td>
<td>Lau, Maly, Loizou, Kroer, Yao, Park, Kovacs, Yin, Zhukov, Lim, Barmherzig, Metaxas, Shi, Udawani, Brendel, Zhou, braverman, Liu, Golikov</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Coffee Break 1</td>
<td></td>
</tr>
<tr>
<td>03:00 PM</td>
<td>Coffee Break 2</td>
<td></td>
</tr>
</tbody>
</table>

NIPS 2017 Time Series Workshop

Vitaly Kuznetsov, Oren Anava, Scott Yang, Azadeh Khaleghi

Grand Ballroom B, Fri Dec 08, 08:00 AM

Data, in the form of time-dependent sequential observations emerge in many key real-world problems, ranging from biological data, financial markets, weather forecasting to audio/video processing. However, despite the ubiquity of such data, most mainstream machine learning algorithms have been primarily developed for settings in which sample points are drawn i.i.d. from some (usually unknown) fixed distribution. While there exist algorithms designed to handle non-i.i.d. data, these typically assume specific parametric form for the data-generating distribution. Such assumptions may undermine the complex nature of modern data which can possess long-range dependency patterns, and for which we now have the computing power to discern. On the other extreme lie on-line learning algorithms that consider a more general framework without any distributional assumptions. However, by being purely-agnostic, common on-line algorithms may not fully exploit the stochastic aspect of time-series data. This is the third instalment of time series workshop at NIPS and will build on the success of the previous events: NIPS 2015 Time Series Workshop and NIPS 2016 Time Series Workshop. The goal of this workshop is to bring together theoretical and applied researchers interested in the analysis of time series and development of new algorithms to process sequential data. This includes algorithms for time series prediction, classification, clustering, anomaly and change point detection, correlation discovery, dimensionality reduction as well as a general theory for learning and comparing stochastic processes. We invite researchers from the related areas of batch and online learning, reinforcement learning, data analysis and statistics, econometrics, and many others to contribute to this workshop.
### Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Speaker(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:45 AM</td>
<td>Elastic Motif Segmentation and Alignment of Time Series for Encoding and Classification</td>
<td>Lee</td>
</tr>
<tr>
<td>11:45 AM</td>
<td>Bayesian Time Series Forecasting with Change Point and Anomaly Detection</td>
<td>Lu</td>
</tr>
<tr>
<td>11:45 AM</td>
<td>Towards Desynchronization Detection in Biosignals</td>
<td>Supratak</td>
</tr>
<tr>
<td>11:45 AM</td>
<td>An Efficient ADMM Algorithm for Structural Break Detection in Multivariate Time Series</td>
<td>Tank</td>
</tr>
<tr>
<td>11:45 AM</td>
<td>An Interpretable and Sparse Neural Network Model for Nonlinear Granger Causality Discovery</td>
<td>Covert</td>
</tr>
<tr>
<td>11:45 AM</td>
<td>Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting</td>
<td>Li</td>
</tr>
<tr>
<td>11:45 AM</td>
<td>Variational inference for latent nonlinear dynamics</td>
<td>Hernandez</td>
</tr>
<tr>
<td>11:45 AM</td>
<td>DP Mixture of Warped Correlated GPs for Individualized Time Series Prediction</td>
<td>Yeo</td>
</tr>
<tr>
<td>11:45 AM</td>
<td>Multi-Scale Change Point Detection in Multivariate Time Series</td>
<td>Ebrahimzadeh</td>
</tr>
<tr>
<td>11:45 AM</td>
<td>Vector-Valued Spectral Analysis of Space-Time Data</td>
<td>Slawinska</td>
</tr>
<tr>
<td>11:45 AM</td>
<td>Time Series Forecasting = Matrix Completion</td>
<td>Shen</td>
</tr>
<tr>
<td>11:45 AM</td>
<td>Improving Factor-Based Quantitative Investing by Forecasting Company Fundamentals</td>
<td>Alberg</td>
</tr>
<tr>
<td>11:45 AM</td>
<td>MiDGaP: Mixture Density Gaussian Processes</td>
<td>Zand</td>
</tr>
<tr>
<td>11:45 AM</td>
<td>Temporal Clustering in time-varying Networks with Time Series Analysis</td>
<td>Tu</td>
</tr>
</tbody>
</table>

---

**From 'What If?' To 'What Next?': Causal Inference and Machine Learning for Intelligent Decision Making**

**Alexander Volfovsky, Adith Swaminathan, Panos Toulis, Nathan Kallus, Ricardo Silva, John S Shawe-Taylor, Thorsten Joachims, Lihong Li**

**Hall C, Fri Dec 08, 08:30 AM**

In recent years machine learning and causal inference have both seen important advances, especially through a dramatic expansion of their theoretical and practical domains. Machine learning has focused on ultra high-dimensional models and scalable stochastic algorithms, whereas causal inference has been guiding policy in complex domains involving economics, social and health sciences, and business. Through such advances a powerful cross-pollination has emerged as a new set of methodologies promising to deliver robust data analysis than each field could individually -- some examples include concepts such as doubly-robust methods, targeted learning, double machine learning, causal trees, all of which have recently been introduced.

This workshop is aimed at facilitating more interactions between researchers in machine learning and causal inference. In particular, it is an opportunity to bring together highly technical individuals who are strongly motivated by the practical importance and real-world impact of their work. Cultivating such interactions will lead to the development of theory, methodology, and - most importantly - practical tools, that better target causal questions across different domains.

In particular, we will highlight theory, algorithms and applications on automatic decision making systems, such as recommendation engines, medical decision systems and self-driving cars, as both producers and users of data. The challenge here is the feedback between learning from data and then taking actions that may affect what data will be made available for future learning. Learning algorithms have to reason about how changes to the system will affect future data, giving rise to challenging counterfactual and causal reasoning issues that the learning algorithm has to account for. Modern and scalable policy learning algorithms also require operating with non-experimental data, such as logged user interaction data where users click ads suggested by recommender systems trained on historical user clicks.

To further bring the community together around the use of such interaction data, this workshop will host a Kaggle challenge problem based on the first real-world dataset of logged contextual bandit feedback with non-uniform action-selection propensities. The dataset consists of several gigabytes of data from an ad placement system, which we have processed into multiple well-defined learning problems of increasing complexity, feedback signal, and context. Participants in the challenge problem will be able to discuss their results at the workshop.

**Schedule**

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Speaker(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30 AM</td>
<td>Introductions</td>
<td>Toulis, Volfovsky</td>
</tr>
<tr>
<td>08:45 AM</td>
<td>Looking for a Missing Signal</td>
<td>Bottou</td>
</tr>
<tr>
<td>09:20 AM</td>
<td>Invited Talk (Brunskill)</td>
<td></td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Contributed Talk 1</td>
<td>Heckerman</td>
</tr>
<tr>
<td>10:15 AM</td>
<td>Contributed Talk 2</td>
<td>Mitrovic</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Invited Talk 3</td>
<td>Hahn</td>
</tr>
<tr>
<td>11:35 AM</td>
<td>Invited Talk 4</td>
<td>Sontag</td>
</tr>
</tbody>
</table>
Extreme Classification: Multi-class & Multi-label Learning in Extremely Large Label Spaces

Manik Varma, Marius Kloft, Krzysztof Dembczynski

Hyatt Beacon Ballroom D+E+F+H, Fri Dec 08, 08:00 AM

Extreme classification is a rapidly growing research area focusing on multi-class and multi-label problems involving an extremely large number of labels. Many applications have been found in diverse areas ranging from language modelling to document tagging in NLP, face recognition to learning universal feature representations in computer vision, gene function prediction in bioinformatics, etc. Extreme classification has also opened up a new paradigm for ranking and recommendation by reformulating them as multi-label learning tasks where each item to be ranked or recommended is treated as a separate label. Such reformulations have led to significant gains over traditional collaborative filtering and content based recommendation techniques. Consequently, extreme classifiers have been deployed in many real-world applications in industry.

Extreme classification raises a number of interesting research questions including those related to:

- Large scale learning and distributed and parallel training
- Log-time and log-space prediction and prediction on a test-time budget
- Label embedding and tree based approaches
- Crowd sourcing, preference elicitation and other data gathering techniques
- Bandits, semi-supervised learning and other approaches for dealing with training set biases and label noise
- Bandits with an extremely large number of arms
- Fine-grained classification
- Zero shot learning and extensible output spaces
- Tackling label polysemy, synonymy and correlations
- Structured output prediction and multi-task learning
- Learning from highly imbalanced data
- Dealing with tail labels and learning from very few data points per label
- PU learning and learning from missing and incorrect labels
- Feature extraction, feature sharing, lazy feature evaluation, etc.
- Performance evaluation
- Statistical analysis and generalization bounds
- Applications to new domains

The workshop aims to bring together researchers interested in these areas to encourage discussion and improve upon the state-of-the-art in extreme classification. In particular, we aim to bring together researchers from the natural language processing, computer vision and core machine learning communities to foster interaction and collaboration. Several leading researchers will present invited talks detailing the latest advances in the area. We also seek extended abstracts presenting work in progress which will be reviewed for acceptance as a spotlight + poster or a talk. The workshop should be of interest to researchers in core supervised learning as well as application domains such as recommender systems, computer vision, computational advertising, information retrieval and natural language processing. We expect a healthy participation from both industry and academia.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Session/Activity</th>
<th>Speaker/Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00 AM</td>
<td>Introduction by Manik Varma</td>
<td>Varma</td>
</tr>
<tr>
<td>09:05 AM</td>
<td>John Langford (MSR) on Dreaming Contextual Memory</td>
<td>Langford</td>
</tr>
<tr>
<td>09:35 AM</td>
<td>Ed Chi (Google) on Learned Deep Retrieval for Recommenders</td>
<td>Chi</td>
</tr>
<tr>
<td>10:05 AM</td>
<td>David Sontag (MIT) on Extreme Classification in Healthcare</td>
<td></td>
</tr>
<tr>
<td>10:35 AM</td>
<td>Short Coffee Break</td>
<td></td>
</tr>
</tbody>
</table>
Conversational AI - today's practice and tomorrow's potential
Alborz Geramifard, Jason Williams, Larry Heck, Jim Glass, Antoine Bordes, Steve Young, Gerald Tesauro

Hyatt Regency Ballroom A+B+C, Fri Dec 08, 08:00 AM

In the span of only a few years, conversational systems have become commonplace. Every day, millions of people use natural-language interfaces such as Siri, Google Now, Cortana, Alexa, Facebook M and others via in-home devices, phones, or messaging channels such as Messenger, Slack, Skype, among others. At the same time, interest among the research community in conversational systems has blossomed: for supervised and reinforcement learning, conversational systems often serve as both a benchmark task and an inspiration for new ML methods at conferences which don't focus on speech and language per se, such as NIPS, ICML, IJCAI, and others. Research community challenge tasks are proliferating, including the sixth Dialog Systems Technology Challenge (DSTC6), the Amazon Alexa prize, and the Conversational Intelligence Challenge live competition at NIPS 2017.

Now more than ever, it is crucial to promote cross-pollination of ideas between academic research centers and industry. The goal of this workshop is to bring together researchers and practitioners in this area, to clarify impactful research problems, share findings from large-scale real-world deployments, and generate new ideas for future lines of research.

This workshop will include invited talks from academia and industry, contributed work, and open discussion. In these talks, senior technical leaders from many of the most popular conversational services will give insights into real usage and challenges at scale. An open call for papers will be issued, and we will prioritize forward-looking papers that propose interesting and impactful contributions. We will end the day with an open discussion, including a panel consisting of academic and industrial researchers.

Schedule

08:20 AM  Opening (Alborz)  Geramifard
08:25 AM  Title by Satindar Singh  Singh
09:10 AM  Accepted Paper Oral
09:30 AM  Facebook/McGill Talk (Joelle)  Pineau
10:15 AM  Accepted Paper Oral - 2
10:35 AM  Coffee Break (1) / Poster Session
11:00 AM  Google Talk (Matt and Dilek) Hakkani-Tur
as well as applications in computer vision, speech, NLP, biology and network analysis are all active areas of research, each with an increasingly large body of foundational knowledge. The workshop aims to ask questions that enable communication across these fields. In particular, this year we aim to address the investigation of combinatorial structures allows to capture complex, high-order dependencies in discrete learning problems prevalent in deep learning, social networks, etc. An emphasis will be given on uncertainty and structure that results from problem instances being estimated from data.

Nearest Neighbors for Modern Applications with Massive Data: An Age-old Solution with New Challenges

George H Chen, Devavrat Shah, Christina Lee

Hyatt Shoreline, Fri Dec 08, 08:00 AM

Many modern methods for prediction leverage nearest neighbor (NN) search to find past training examples most similar to a test example, an idea that dates back in text to at least the 11th century in the “Book of Optics” by Alhazen. Today, NN methods remain popular, often as a cog in a bigger prediction machine, used for instance in recommendation systems, forecasting baseball player performance and election outcomes, survival analysis in healthcare, image in-painting, crowdsourcing, graphon estimation, and more. The popularity of NN methods is due in no small part to the proliferation of high-quality fast approximate NN search methods that scale to high-dimensional massive datasets typical of contemporary applications. Moreover, NN prediction readily pairs with methods that learn similarities, such as metric learning methods or Siamese networks. In fact, some well-known pairings that result in nearest neighbor predictors that learn similarities include random forests and many boosting methods.

Despite the popularity, success, and age of nearest neighbor methods, our theoretical understanding of them is still surprisingly incomplete (perhaps much to the chagrin of the initial efforts of analysis by Fix, Hodges, Cover, and Hart!) and can also be disconnected from what practitioners actually want or care about. Many successful approximate nearest neighbor methods in practice do not have known theoretical guarantees, and many of the guarantees for exact nearest neighbor methods do not readily handle approximation. Meanwhile, many applications use variations on NN methods, for which existing theory may not extend to, or for which existing theory is not easily usable by a practitioner. Suffice it to say, a lot is lost in translation between different communities working with NN methods.

In short, NN methods is an exciting field at the intersection of classical statistics, machine learning, data structures and domain specific expertise. The aim of this work is to bring together theoreticians and practitioners alike from these various different backgrounds with a diverse range of perspectives to bring everyone up to speed on:
- Best known statistical/computational guarantees (especially recent non-asymptotic results)
- Latest methods/systems that have been developed especially for fast approximate NN search that scale to massive datasets
- Various applications in which NN methods are heavily used as a critical component in prediction or inference

By gathering a diverse crowd, we hope attendees share their
perspectives, identify ways to bridge theory and practice, and discuss avenues of future research.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Title</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00 AM</td>
<td>Intro/Survey of Nearest Neighbor Methods</td>
<td>Chen</td>
</tr>
<tr>
<td>09:20 AM</td>
<td>Title TBD (Kamalika)</td>
<td>Chaudhuri</td>
</tr>
<tr>
<td>10:05 AM</td>
<td>Title TBD (Sewoong)</td>
<td>Oh</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Title TBD (Piotr)</td>
<td>Indyk</td>
</tr>
<tr>
<td>11:45 AM</td>
<td>Fast k-Nearest Neighbor Search via Prioritized DCI</td>
<td>Li</td>
</tr>
<tr>
<td>01:30 PM</td>
<td>Title TBD (Alyosha)</td>
<td>Efros</td>
</tr>
<tr>
<td>02:15 PM</td>
<td>Title TBD (Celect)</td>
<td>Doshi</td>
</tr>
<tr>
<td>02:35 PM</td>
<td>Poster Session (encompasses coffee break)</td>
<td>Chen, Balle, Lee, frosio, Malik, Kautz, Li, Sugiyama, Carreira-Perpinan, Raziperchikolaei, Tulabandhula, Noh</td>
</tr>
<tr>
<td>04:00 PM</td>
<td>The Boundary Forest Algorithm</td>
<td>Yedidia</td>
</tr>
<tr>
<td>04:45 PM</td>
<td>Iterative Collaborative Filtering for Sparse Matrix Estimation</td>
<td>Lee</td>
</tr>
</tbody>
</table>

Visually grounded interaction and language

Florian Strub, Harm de Vries, Abhishek Das, Satwik Kottur, Stefan Lee, Mateusz Malinowski, Olivier Pietquin, Devi Parikh, Dhruv Batra, Aaron C Courville, Jeremie Mary

Room 101-A, Fri Dec 08, 08:00 AM

Everyday interactions require a common understanding of language, i.e. for people to communicate effectively, words (for example ‘cat’) should invoke similar beliefs over physical concepts (what cats look like, the sounds they make, how they behave, what their skin feels like etc.). However, how this ‘common understanding’ emerges is still unclear.

One appealing hypothesis is that language is tied to how we interact with the environment. As a result, meaning emerges by ‘grounding’ language in modalities in our environment (images, sounds, actions, etc.).

Recent concurrent works in machine learning have focused on bridging visual and natural language understanding through visually-grounded language learning tasks, e.g. through natural images (Visual Question Answering, Visual Dialog), or through interactions with virtual physical environments. In cognitive science, progress in fMRI enables creating a semantic atlas of the cerebral cortex, or to decode semantic information from visual input. And in psychology, recent studies show that a baby’s most likely first words are based on their visual experience, laying the foundation for a new theory of infant language acquisition and learning.

As the grounding problem requires an interdisciplinary attitude, this workshop aims to gather researchers with broad expertise in various fields — machine learning, computer vision, natural language, neuroscience, and psychology — to discuss their cutting edge work as well as perspectives on future directions in this exciting space of grounding and interactions.

We will accept papers related to:
— language acquisition or learning through interactions
— visual captioning, dialog, and question-answering
— reasoning in language and vision
— visual synthesis from language
— transfer learning in language and vision tasks
— navigation in virtual worlds with natural-language instructions
— machine translation with visual cues
— novel tasks that combine language, vision and actions
— understanding and modeling the relationship between language and vision in humans
— semantic systems and modeling of natural language and visual stimuli representations in the human brain

Important dates
---------------------
Submission deadline: 3rd November 2017
Extended Submission deadline: 17th November 2017
Acceptance notification (First deadline): 10th November 2017
Acceptance notification (Second deadline): 24th November 2017

Workshop: 8th December 2017

Paper details
-------------
— Contributed papers may include novel research, preliminary results, extended abstract, positional papers or surveys
— Papers are limited to 4 pages, excluding references, in the latest camera-ready NIPS format: https://nips.cc/Conferences/2017/PaperInformation/StyleFiles
— Papers published at the main conference can be submitted without reformatting
— Please submit via email: nips2017vigil@gmail.com

Accepted papers
-----------------------
— All accepted papers will be presented during 2 poster sessions
— Up to 5 accepted papers will be invited to deliver short talks
— Accepted papers will be made publicly available as non-archival reports, allowing future submissions to archival conferences and journals

Invited Speakers
-----------------------
Raymond J. Mooney - University of Texas
Sanja Fidler - University of Toronto
Olivier Pietquin - DeepMind
Jack Gallant - University of Berkeley
Devi Parikh - Georgia Tech / FAIR
Felix Hill - DeepMind
Jack Gallant - University of Berkeley
Chen Yu - University of Indiana
Learning in the Presence of Strategic Behavior

Nika Haghtalab, Yishay Mansour, Tim Roughgarden, Vasilis Syrgkanis, Jenn Wortman Vaughan

Room 101-B, Fri Dec 08, 08:00 AM

Machine learning is primarily concerned with the design and analysis of algorithms that learn about an entity. Increasingly more, machine learning is being used to design policies that affect the entity. This can cause the entity to react and present a different behavior. Ignoring such interactions could lead to solutions that are ultimately ineffective in practice. For example, to design an effective ad display one has to take into account how a viewer would react to the displayed advertisements, for example by choosing to scroll through or click on them. Additionally, in many environments, multiple learners learn concurrently about one or more related entities. This can bring about a range of interactions between individual learners. For example, multiple firms may compete or collaborate on performing market research. How do the learners and entities interact? How do these interactions change the task at hand? What are some desirable interactions in a learning environment? And what are the mechanisms for bringing about such desirable interactions? These are some of the questions we would like to explore more in this workshop.

Traditionally, learning theory has adopted two extreme views in this respect: First, when learning occurs in isolation from strategic behavior, such as in the classical PAC setting where the data is drawn from a fixed distribution; second, when the learner faces an adversary whose goal is to inhibit the learning process, such as the minimax setting where the data is generated by an adaptive worst-case adversary. While these extreme perspectives have lead to elegant results and concepts, such as VC dimension, Littlestone dimension, regret bounds, and more, many types of problems that we would like to solve involve strategic behaviors that do not fall into these two extremes. Examples of these problems include but are not limited to

1. Learning from data that is produced by agents who have vested interest in the outcome or the learning process. For example, learning a measure of quality of universities by surveying members of the academia who stand to gain or lose from the outcome, or when a GPS routing app has to learn patterns of traffic delay by routing individuals who have no interest in taking slower routes.

2. Learning a model for the strategic behavior of one or more agents by observing their interactions, for example, learning economical demands of buyers by observing their bidding patterns when competing with other buyers.

3. Learning as a model of interactions between agents. Examples of this include applications to swarm robotics, where individual agents have to learn to interact in a multi-agent setting in order to achieve individual or collective goals.

4. Interactions between multiple learners. In many settings, two or more learners learn about the same or multiple related concepts. How do these learners interact? What are the scenarios under which they would share knowledge, information, or data. What are the desirable interactions between learners? As an example, consider multiple competing pharmaceutical firms that are learning about the effectiveness of a certain treatment. In this case, while competing firms would prefer not to share their findings, it is beneficial to the society when such findings are shared. How can we incentivize these learners to perform such desirable interactions?

The main goal of this workshop is to address current challenges and opportunities that arise from the presence of strategic behavior in learning theory. This workshop aims at bringing together members of different communities, including machine learning, economics, theoretical computer science, and social computing, to share recent results, discuss important directions for future research, and foster collaborations.
Learning in Repeated Auctions with Budgets: Regret Minimization and Equilibrium

Gur

10:00 AM

Spotlights

Podimata, Kim

10:15 AM

Invited Talk: Eva Tardos

Tardos

11:00 AM

Invited Talk: Mehryar Mohri

Mohri

11:45 AM

Lunch Break

12:30 PM

Invited Talk: Percy Liang

Liang

01:00 PM

Poster Session & Coffee break

01:30 PM

Invited Talk: Alex Peysakhovich

Peysakhovich

02:35 PM

Statistical Tests of Incentive Compatibility in Display Ad Auctions

03:00 PM

Optimal Economic Design through Deep Learning

Parkes

04:00 PM

Learning Against Non-Stationary Agents with Opponent Modeling & Deep Reinforcement Learning

04:45 PM

Abstract 2: Strategic Classification from Revealed Preferences in Learning in the Presence of Strategic Behavior, Dong 09:45 AM

We study an online linear classification problem, in which the data is generated by strategic agents who manipulate their features in an effort to change the classification outcome. In rounds, the learner deploys a classifier, and an adversarially chosen agent arrives, possibly manipulating her features to optimally respond to the learner. The learner has no knowledge of the agents’ utility functions or “real” features, which may vary widely across agents. Instead, the learner is only able to observe their “revealed preferences” --- i.e. the actual manipulated feature vectors they provide. For a broad family of agent cost functions, we give a computationally efficient learning algorithm that is able to obtain diminishing ”Stackelberg regret” --- a form of policy regret that guarantees that the learner is obtaining loss nearly as small as that of the best classifier in hindsight, even allowing for the fact that agents will best-respond differently to the optimal classifier.

Jinshuo Dong, Aaron Roth, Zachary Schutzman, Bo Waggoner and Zhiwei Steven Wu

Abstract 3: Learning in Repeated Auctions with Budgets: Regret Minimization and Equilibrium in Learning in the Presence of Strategic Behavior, Gur 10:00 AM

In online advertising markets, advertisers often purchase ad placements through bidding in repeated auctions based on realized viewer information. We study how budget-constrained advertisers may compete in such sequential auctions in the presence of uncertainty about future bidding opportunities and competition. We formulate this problem as a sequential game of incomplete information, where bidders know neither their own valuation distribution, nor the budgets and valuation distributions of their competitors. We introduce a family of practical bidding strategies we refer to as adaptive pacing strategies, in which advertisers adjust their bids according to the sample path of expenditures they exhibit. Under arbitrary competitors’ bids, we establish through matching lower and upper bounds the asymptotic optimality of this class of strategies as the number of auctions grows large. When all the bidders adopt these strategies, we establish the convergence of the induced dynamics and characterize a regime (well motivated in the context of display advertising markets) under which these strategies constitute an approximate Nash equilibrium in dynamic strategies: The benefit of unilaterally deviating to other strategies, including ones with access to complete information, becomes negligible as the number of auctions and competitors grows large. This establishes a connection between regret minimization and market stability, by which advertisers can essentially follow equilibrium bidding strategies that also ensure the best performance that can be guaranteed off-equilibrium.

Yonatan Gur and Santiago Balseiro.

Abstract 4: Spotlights in Learning in the Presence of Strategic Behavior, Podimata, Kim 10:15 AM

1. Strategyproof Linear Regression. Yiling Chen, Chara Podimata and Nisarg Shah

2. Incentive-Aware Learning for Large Markets. Mohammad Mahdian, Vahab Mirrokni and Song Zuo.


4. Robust commitments and partial reputation. Vidya Muthukumar and Anant Sahai

Abstract 9: Spotlights in Learning in the Presence of Strategic Behavior, Schmit, Kangasrääsiö, Cai, Feng 02:35 PM

1. Learning to Bid Without Knowing Your Value. Zhe Feng, Chara Podimata and Vasilis Syrgkanis.

2. Incentive-Aware Learning for Large Markets. Mohammad Mahdian, Vahab Mirrokni and Song Zuo.

3. Learning Multi-item Auctions with (or without) Samples. Yang Cai and Constantinos Daskalakis.


5. Robust commitments and partial reputation. Vidya Muthukumar and Anant Sahai

Abstract 12: Statistical Tests of Incentive Compatibility in Display Ad Auctions in Learning in the Presence of Strategic Behavior, 04:15 PM

Consider a buyer participating in a repeated auction in an ad exchange. How does a buyer figure out whether her bids will be used against her in the form of reserve prices? There are many potential A/B testing setups that one can use. However, we will show many natural experimental designs have serious flaws.

For instance, one can use additive or multiplicative perturbation to the bids. We show that additive perturbations to bids can lead to paradoxical results, as reserve prices are not guaranteed to be monotone for non-MHR distributions, and thus higher bids may lead to lower reserve prices!

Similarly, one may be tempted to measure bid influence in reserves by randomly perturbing one's bids. However, unless the perturbations are aligned with the partitions used by the seller to compute optimal reserve prices, the results are guaranteed to be inconclusive.

Finally, in practice additional market considerations play a large role---if the optimal reserve price is further constrained by the seller to satisfy additional business logic, the power of the buyer to detect the effect to which his bids are being used against him is limited.

In this work we develop tests that a buyer can use to measure the impact of current bids on future reserve prices. In addition, we analyze the cost of running such experiments, exposing trade-offs between test accuracy, cost, and underlying market dynamics. We validate our results with experiments on real world data and show that a buyer can detect reserve price optimization done by the seller at a reasonable cost.

Andres Munoz Medina, Sebastien Lahaie, Sergei Vassilvitskii and Balasubramanian Sivan

Abstract 13: Optimal Economic Design through Deep Learning in Learning in the Presence of Strategic Behavior, Parkes 04:30 PM

Designing an auction that maximizes expected revenue is an intricate task. Despite major efforts, only the single-item case is fully understood. We explore the use of tools from deep learning on this topic. The design objective is revenue optimal, dominant-strategy incentive compatible auctions. For a baseline, we show that multi-layer neural networks can learn almost-optimal auctions for a variety of settings for which there are analytical solutions, and even without encoding characterization results into the design of the network. Looking ahead, deep learning has promise for deriving auctions with high revenue for poorly understood problems.

Paul Duetting, Zhe Feng, Harikrishna Narasimhan, and David Parkes

Abstract 14: Learning Against Non-Stationary Agents with Opponent Modeling & Deep Reinforcement Learning in Learning in the Presence of Strategic Behavior, 04:45 PM

Humans, like all animals, both cooperate and compete with each other. Through these interactions we learn to observe, act, and manipulate to maximize our utility function, and continue doing so as others learn with us. This is a decentralized non-stationary learning problem, where to survive and flourish an agent must adapt to the gradual changes of other agents as they learn, as well as capitalize on sudden shifts in their behavior. To date, a majority of the work in deep multi-agent reinforcement learning has focused on only one of these types of adaptations. In this paper, we introduce the Switching Agent Model (SAM) as a way of dealing with both types of non-stationarity through the combination of opponent modeling and deep multi-agent reinforcement learning.

Richard Everett

Machine Learning for Creativity and Design

Douglas Eck, David Ha, Ali Eslami, Sander Dieleman, Rebecca Fiebrink, Luba Elliott

Room 102 A+B, Fri Dec 08, 08:00 AM

In the last year, generative machine learning and machine creativity have gotten a lot of attention in the non-research world. At the same time there have been significant advances in generative models for media creation and for design. This one-day workshop explores several issues in the domain of generative models for creativity and design. First, we will look at algorithms for generation and creation of new media and new designs, engaging researchers building the next generation of generative models (GANs, RL, etc) and also from a more information-theoretic view of creativity (compression, entropy, etc). Second, we will investigate the social and cultural impact of these new models, engaging researchers from HCI/UX communities. Finally, we’ll hear from some of the artists and musicians who are adopting machine learning approaches like deep learning and reinforcement learning as part of their artistic process. We’ll leave ample time for discussing both the important technical challenges of generative models for creativity and design, as well as the philosophical and cultural issues that surround this area of research.

Background

In 2016, DeepMind’s AlphaGo made two moves against Lee Sedol that were described by the Go community as “brilliant,” “surprising,” “beautiful,” and so forth. Moreover, there was little discussion surrounding the fact that these very creative moves were actually made by a machine (Wired); it was enough that they were great examples of go playing. At the same time, the general public showed more concern for other applications of generative models. Algorithms that allow for convincing voice style transfer (Lyrebird) or puppet-like video face control (Face2Face) have raised concerns that generative ML will be used to
make convincing forms of fake news (FastCompany).

Balancing this, the arts and music worlds have positively embraced generative models. Starting with DeepDream and expanding with image and video generation advances (e.g. GANs) we’ve seen lots of new and interesting art and music [citations] technologies provided by the machine learning community. We’ve seen research projects like Google Brain’s Magenta, Sony CSL’s FlowMachines and IBM’s Watson undertake collaborations and attempt to build tools and ML models for use by these communities.

Research
Recent advances in generative models enable new possibilities in art and music production. Language models can be used to write science fiction film scripts (Sunspring) and even replicate the style of individual authors (DeepTingle). Generative models for image and video allow us to create visions of people, places and things that resemble the distribution of actual images (GANs etc). Sequence modelling techniques have opened up the possibility of generating realistic musical scores (MIDI generation etc) and even raw audio that resembles human speech and physical instruments (DeepMind’s WaveNet, MILA’s Char2Wav and Google’s NSynth). In addition, sequence modelling allows us to model vector images to construct stroke-based drawings of common objects according to human doodles (sketch-rnn).

In addition to field-specific research, a number of papers have come out that are directly applicable to the challenges of generation and evaluation such as learning from human preferences (Christiano et al., 2017) and CycleGAN. The application of Novelty Search (Stanley), evolutionary complexification (Stanley - CPPN, NEAT, Nguyen et al - Plug&Play GANs, Innovation Engine) and intrinsic motivation (Oudeyer et al 2007, Schmidhuber on Fun and Creativity) techniques, where objective functions are constantly evolving, is still not common practice in art and music generation using machine learning.

Another focus of the workshop is how to better enable human influence over generative models. This could include learning from human preferences, exposing model parameters in ways that are understandable and relevant to users in a given application domain (e.g., similar to Morris et al. 2008), enabling users to manipulate models through changes to training data (Fiebrink et al. 2011), allowing users to dynamically mix between multiple generative models (Akten & Grierson 2016), or other techniques. Although questions of how to make learning algorithms controllable and understandable to users are relatively nascent in the modern context of deep learning and reinforcement learning, such questions have been a growing focus of work within the human-computer interaction community (e.g., examined in a CHI 2016 workshop on Human-Centred Machine Learning), and the AI Safety community (e.g. Christiano et al. 2017, using human preferences to train deep reinforcement learning systems). Such considerations also underpin the new Google “People + AI Research” (PAIR) initiative.

Artists and Musicians
All the above techniques improve our capabilities of producing text, sound and images. Art and music that stands the test of time however requires more than that. Recent research includes a focus on novelty in creative adversarial networks (Elgammal et al., 2017) and considers how generative algorithms can integrate into human creative processes, supporting exploration of new ideas as well as human influence over generated content (Atken & Grierson 2016a, 2016b). Artists including Mario Klingemann, Gene Kogan, Mike Tyka, and Memo Akten have further contributed to this space of work by creating artwork that compellingly demonstrates capabilities of generative algorithms, and by publicly reflecting on the artistic affordances of these new tools.

The goal of this workshop is to bring together researchers interested in advancing art and music generation to present new work, foster collaborations and build networks.

In this workshop, we are particularly interested in how the following can be used in art and music generation: reinforcement learning, generative adversarial networks, novelty search and evaluation as well as learning from user preferences. We welcome submissions of short papers, demos and extended abstracts related to the above.

There will also be an open call for a display of artworks incorporating machine learning techniques.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Session Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30 AM</td>
<td>Welcome and Introduction</td>
</tr>
<tr>
<td>08:45 AM</td>
<td>Jürgen Schmidhuber</td>
</tr>
<tr>
<td>09:15 AM</td>
<td>Emily Denton</td>
</tr>
<tr>
<td>09:45 AM</td>
<td>Rebecca Fiebrink</td>
</tr>
<tr>
<td>10:15 AM</td>
<td>GANosaic - Mosaic Creation with Generative Texture Manifolds</td>
</tr>
<tr>
<td>10:20 AM</td>
<td>TopoSketch: Drawing in Latent Space</td>
</tr>
<tr>
<td>10:25 AM</td>
<td>Input parameterization for DeepDream</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Ian Goodfellow</td>
</tr>
<tr>
<td>11:30 AM</td>
<td>Improvised Comedy as a Turing Test</td>
</tr>
<tr>
<td>12:00 PM</td>
<td>Lunch</td>
</tr>
<tr>
<td>01:00 PM</td>
<td>Ahmed Elgammal</td>
</tr>
<tr>
<td>01:30 PM</td>
<td>Hierarchical Variational Autoencoders for Music</td>
</tr>
<tr>
<td>02:00 PM</td>
<td>Lexical preferences in an automated story writing system</td>
</tr>
<tr>
<td>02:30 PM</td>
<td>ObamaNet: Photo-realistic lip-sync from text</td>
</tr>
<tr>
<td>03:00 PM</td>
<td>Art / Coffee Break</td>
</tr>
<tr>
<td>03:30 PM</td>
<td>Towards the High-quality Anime Characters</td>
</tr>
<tr>
<td></td>
<td>Generation with Generative Adversarial Networks</td>
</tr>
<tr>
<td>03:35 PM</td>
<td>Crowd Sourcing Clothes</td>
</tr>
<tr>
<td></td>
<td>Design Directed by</td>
</tr>
<tr>
<td></td>
<td>Adversarial Neural Networks</td>
</tr>
</tbody>
</table>
The success of machine learning has been demonstrated time and time again in classification, generative modelling, and reinforcement learning. In particular, we have recently seen interesting developments where ML has been applied to the natural sciences (chemistry, physics, materials science, neuroscience and biology). Here, often the data is not abundant and very costly. This workshop will focus on the unique challenges of applying machine learning to molecules and materials.

Accurate prediction of chemical and physical properties is a crucial ingredient toward rational compound design in chemical and pharmaceutical industries. Many discoveries in chemistry can be guided by screening large databases of computational molecular structures and properties, but high level quantum-chemical calculations can take up to several days per molecule or material at the required accuracy, placing the ultimate achievement of in silico design out of reach for the foreseeable future. In large part the current state of the art for such problems is the expertise of individual researchers or at best highly-specific rule-based heuristic systems. Efficient methods in machine learning, applied to property and structure prediction, can therefore have pivotal impact in enabling chemical discovery and foster fundamental insights.

Because of this, in the past few years there has been a flurry of recent work towards designing machine learning techniques for molecule [1, 2, 4-11, 13-18, 20, 21, 23-32, 34-38] and material data [1-3, 5, 6, 12, 19, 24, 33]. These works have drawn inspiration from and made significant contributions to areas of machine learning as diverse as learning on graphs to models in natural language processing. Recent advances enabled the acceleration of molecular dynamics simulations, contributed to a better understanding of interactions within quantum many-body systems and increased the efficiency of density functional theory based quantum mechanical modeling methods. This young field offers unique opportunities for machine learning researchers and practitioners, as it presents a wide spectrum of challenges and open questions, including but not limited to representations of physical systems, physically
constrained models, manifold learning, interpretability, model bias, and causality.

The goal of this workshop is to bring together researchers and industrial practitioners in the fields of computer science, chemistry, physics, materials science, and biology all working to innovate and apply machine learning to tackle the challenges involving molecules and materials. In a highly interactive format, we will outline the current frontiers and present emerging research directions. We aim to use this workshop as an opportunity to establish a common language between all communities, to actively discuss new research problems, and also to collect datasets by which novel machine learning models can be benchmarked. The program is a collection of invited talks, alongside contributed posters. A panel discussion will provide different perspectives and experiences of influential researchers from both fields and also engage open participant conversation. An expected outcome of this workshop is the interdisciplinary exchange of ideas and initiation of collaboration.

References
quantum chemistry, and chemical space. Reviews in Computational Chemistry, 225-256.


Schedule

08:00 AM TBA Müller
08:20 AM TBA2 Aspuru-Guzik
08:45 AM TBA3 Tkatchenko
09:10 AM TBA4
09:30 AM TBA5

10:10 AM Poster session 1

10:40 AM TBA6 von Lilienfeld
11:00 AM TBA7 Jensen
11:20 AM TBA8 burke
11:45 AM TBA9 Carleo
12:00 PM TBA10 Carleo
01:55 PM TBA11 Smola
02:20 PM TBA12 Kondor
02:40 PM TBA13 Tsuda
02:55 PM TBA14 Ceriotti
03:15 PM TBA15 Chmiela
04:05 PM TBA16
04:25 PM TBA17 Duvenaud
04:40 PM TBA18
05:00 PM TBA19 Schütt
05:50 PM TBA20 Hernández-Lobato

06:05 PM Poster session 3 Vogt-Maranto

---

**Advances in Modeling and Learning Interactions from Complex Data**

**Gautam Dasarathy, Mladen Kolar, Richard Baraniuk**

**Room 103 A+B, Fri Dec 08, 08:00 AM**

Whether it is biological networks of proteins and genes or technological ones like sensor networks and the Internet, we are surrounded today by complex systems composed of entities interacting with and affecting each other. An urgent need has therefore emerged for developing novel techniques for modeling, learning, and conducting inference in such networked systems. Consequently, we have seen progress from a variety of disciplines in both fundamental methodology and in applications of such methods to practical problems. However, much work remains to be done, and a unifying and principled framework for dealing with these problems remains elusive. This workshop aims to bring together theoreticians and practitioners in order to both chart out recent advances and to discuss new directions in understanding interactions in large and complex systems. NIPS, with its attendance by a broad and cross-disciplinary set of researchers offers the ideal venue for this exchange of ideas.

The workshop will feature a mix of contributed talks, contributed posters, and invited talks by leading researchers from diverse backgrounds working in these areas. We will also have a specific segment of the schedule reserved for the presentation of open problems, and will have plenty of time for discussions where we will explicitly look to spark off collaborations amongst the attendees.

We encourage submissions in a variety of topics including, but not limited to:

* Computationally and statistically efficient techniques for learning graphical models from data including convex, greedy, and active approaches.
* New probabilistic models of interacting systems including nonparametric and exponential family graphical models.
* Community detection algorithms including semi-supervised and adaptive approaches.
* Techniques for modeling and learning causal relationships from data.
* Bayesian techniques for modeling complex data and causal relationships.
* Kernel methods for directed and undirected graphical models.
* Applications of these methods in various areas like sensor networks, computer networks, social networks, and biological networks like phylogenetic trees and graphs.

Successful submissions will emphasize the role of statistical and computational learning to the problem at hand. The author(s) of these submissions will be invited to present their work as either a poster or as a contributed talk. Alongside these, we also solicit submissions of open problems that go with the theme of the workshop. The author(s) of the selected open problems will be able to present the problem to the attendees and solicit feedback/collaborations.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:00 AM</td>
<td>Opening Remarks</td>
<td>Dasarathy</td>
</tr>
<tr>
<td>08:05 AM</td>
<td>Invited Talk 1</td>
<td></td>
</tr>
<tr>
<td>08:40 AM</td>
<td>Contributed Talk 1</td>
<td></td>
</tr>
</tbody>
</table>
6th Workshop on Automated Knowledge Base Construction (AKBC)

Jay Pujara, Danqi Chen, Bhavana Dalvi Mishra, Tim Rocktäschel

Room 103-C, Fri Dec 08, 08:00 AM

Extracting knowledge from text, images, audio, and video and translating these extractions into a coherent, structured knowledge base (KB) is a task that spans the areas of machine learning, natural language processing, computer vision, databases, search, data mining and artificial intelligence. Over the past two decades, machine learning techniques used for information extraction, graph construction, and automated knowledge base construction have evolved from simple rule learning to end-to-end neural architectures with papers on the topic consistently appearing at NIPS. Hence, we believe this workshop will appeal to NIPS attendees and be a valuable contribution. Furthermore, there has been significant interest and investment in knowledge base construction in both academia and industry in recent years. Most major internet companies and many startups have developed knowledge bases that power digital assistants (e.g. Siri, Alexa, Google Now) or provide the internet companies and many startups have developed knowledge bases that power digital assistants (e.g. Siri, Alexa, Google Now) or provide the

controversial questions, and suggest debate topics for speaker panels. In addition, for the first time, AKBC will address a longstanding issue in the AKBC, that of equitable comparison and evaluation across methods, by including a shared evaluation platform, Stanford’s KBP Online (https://kbpo.stanford.edu/), which will allow crowdsourced labels for KBs without strong assumptions about the data or methods used. Together, this slate of high-profile research talks, outstanding contributed papers, an interactive research environment, and a novel evaluation service will ensure AKBC is a popular addition to the NIPS program.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:00 AM</td>
<td>Talk by Luna Dong</td>
<td>Rawat, Joulin, Jansen, Lee, Chen, Xu, Verga, Juba, Dumitrache, Jat, Logan, Sindh, Yang, Das, Pezeshkpour, Monath</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Poster Session</td>
<td></td>
</tr>
<tr>
<td>10:45 AM</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Invited Talk 2</td>
<td>Francisco Ruiz, Stephan Mandt, Cheng Zhang, James McInerney, Dustin Tran, Tamara Broderick, Michalis Titsias, David Blei, Max Welling</td>
</tr>
<tr>
<td>11:30 AM</td>
<td>Contributed Talk 3</td>
<td></td>
</tr>
<tr>
<td>12:00 PM</td>
<td>Lunch</td>
<td></td>
</tr>
<tr>
<td>12:30 PM</td>
<td>Invited Talk 3</td>
<td></td>
</tr>
<tr>
<td>01:00 PM</td>
<td>Contributed Talk 4</td>
<td></td>
</tr>
<tr>
<td>01:30 PM</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>01:45 PM</td>
<td>Invited Talk 5</td>
<td></td>
</tr>
<tr>
<td>02:15 PM</td>
<td>Contributed Talk 5</td>
<td></td>
</tr>
</tbody>
</table>

Advances in Approximate Bayesian Inference

Francisco Ruiz, Stephan Mandt, Cheng Zhang, James McInerney, Dustin Tran, Tamara Broderick, Michalis Titsias, David Blei, Max Welling

Room 104-A, Fri Dec 08, 08:00 AM

Approximate inference is key to modern probabilistic modeling. Thanks to the availability of big data, significant computational power, and sophisticated models, machine learning has achieved many breakthroughs in multiple application domains. At the same time, approximate inference becomes critical since exact inference is intractable for most models of interest. Within the field of approximate Bayesian inference, variational and Monte Carlo methods are currently the mainstay techniques. For both methods, there has been considerable progress both on the efficiency and performance.

In this workshop, we encourage submissions advancing approximate inference methods. We are open to a broad scope of methods within the field of Bayesian inference. In addition, we also encourage applications of approximate inference in many domains, such as computational biology, recommender systems, differential privacy, and industry applications.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30 AM</td>
<td>Workshop Introduction</td>
<td>Zhang, Ruiz, Tran, McInerney, Mandt</td>
</tr>
<tr>
<td>08:35 AM</td>
<td>Invited Talk 1: Ian Murray</td>
<td>Murray</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>Contributed Talk 1</td>
<td></td>
</tr>
<tr>
<td>09:15 AM</td>
<td>Invited Talk 2: Yingzhen Li</td>
<td>Li</td>
</tr>
<tr>
<td>09:40 AM</td>
<td>Industry Talk 1: Dawn Liang</td>
<td>Liang</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Industry Talk 2: Cedric Archambeau</td>
<td>Archambeau</td>
</tr>
</tbody>
</table>
Machine Learning for Health (ML4H) - What Parts of Healthcare are Ripe for Disruption by Machine Learning Right Now?

Andrew Beam, Madalina Filateru, Peter Schulam, Jason Fries, Michael Hughes, Alex Wiltschko, Jasper Snoek, Natalia Antropova, Rajesh Ranganath, Bruno Jedynak, Tristan Naumann, Adrian Dalca, Adrian Dalca, Tim Althoff, Shubhi ASTHANA, Prateek Tandon, Jaz Kandola, Alexander Ratner, David Kale, Uri Shalit, Marzyeh Ghassemi, Isaac S Kohane

Room 104-B, Fri Dec 08, 08:00 AM

The goal of the NIPS 2017 Machine Learning for Health Workshop (ML4H) is to foster collaborations that meaningfully impact medicine by bringing together clinicians, health data experts, and machine learning researchers. We aim to build on the success of the last two NIPS ML4H workshops which were widely attended and helped form the foundations of a new research community.

This year’s program emphasizes identifying previously unidentified problems in healthcare that the machine learning community hasn’t addressed, or seeing old challenges through a new lens. While healthcare and medicine are often touted as prime examples for disruption by AI and machine learning, there has been vanishingly little evidence of this disruption to date. To interested parties who are outside of the medical establishment (e.g. machine learning researchers), the healthcare system can appear byzantine and impenetrable, which results in a high barrier to entry. In this workshop, we hope to reduce this activation energy by bringing together leaders at the forefront of both machine learning and healthcare for a dialog on areas of medicine that have immediate opportunities for machine learning. Attendees at this workshop will quickly gain an understanding of the key problems that are unique to healthcare and how machine learning can be applied to addressed these challenges.

The workshop will feature invited talks from leading voices in both medicine and machine learning. A key part of our workshop is the clinician pitch; a short presentation of open clinical problems where data-driven solutions can make an immediate difference. This year’s program will also include spotlight presentations and two poster sessions highlighting novel research contributions at the intersection of machine learning and healthcare. The workshop will conclude with an interactive a panel discussion where all speakers respond to questions provided by the audience.

**Schedule**

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Speaker(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:20 AM</td>
<td>Keynote: Zak Kohane, Harvard DBMI</td>
<td>Kohane</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>Jennifer Chayes, Microsoft Research New England</td>
<td>Chayes</td>
</tr>
<tr>
<td>09:30 AM</td>
<td>Keynote II: Fei-Fei Li, Stanford</td>
<td>Fei-Fei</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Coffee break and Poster Session I</td>
<td></td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Invited clinical panel</td>
<td>Velazquez, Priest, strigo</td>
</tr>
<tr>
<td>12:00 PM</td>
<td>Keynote II: Fei-Fei Li, Stanford</td>
<td></td>
</tr>
</tbody>
</table>
Abstract 6: Invited clinical panel in Machine Learning for Health (ML4H) - What Parts of Healthcare are Ripe for Disruption by Machine Learning Right Now?, Velazquez, Priest, Strigo 11:00 AM

Susann Beier, U. Auckland
James Priest, Stanford
Irina Strigo, UCSF
Enrique Velazquez, Rady Children’s Hospital

Abstract 8: Interactive panel in Machine Learning for Health (ML4H) - What Parts of Healthcare are Ripe for Disruption by Machine Learning Right Now?, 01:40 PM

Interactive panel moderated by Zak Kohane:
- Atul Butte
- Jennifer Chayes
- Fei-Fei Li
- Jill Mesirov
- Susan Murphy
- Mustafa Suleyman

Abstract 13: Award session + A word from our affiliates in Machine Learning for Health (ML4H) - What Parts of Healthcare are Ripe for Disruption by Machine Learning Right Now?, 04:30 PM

Award session and a word from affiliates
- Eunho Yang, KAIST, Korea
- Sung-ju Hwang, UNIST, Korea
The use of machine learning has become pervasive in our society, from specialized scientific data analysis to industry intelligence and practical applications with a direct impact in the public domain. This impact involves different social issues including privacy, ethics, liability and accountability. This workshop aims to discuss the use of machine learning in safety critical environments, with special emphasis on three main application domains:

- Healthcare
- Autonomous systems
- Complainants and liability in data driven industries

We aim to answer some of these questions: How do we make our models more comprehensible and transparent? Shall we always trust our decision making process? How do we involve field experts in the process of making machine learning pipelines more practically interpretable from the viewpoint of the application domain?

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Presenter(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:50 AM</td>
<td>Opening remarks</td>
<td>Tosi, Vellido, Álvarez</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>Invited talk 1</td>
<td>Saria</td>
</tr>
<tr>
<td>09:45 AM</td>
<td>Invited talk 2</td>
<td>Doshi-Velez</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Beyond Sparsity: Tree-based Regularization of Deep Models for Interpretability</td>
<td>Wu, Parbhoo, Doshi-Velez</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Coffee break 1</td>
<td></td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Invited talk 3</td>
<td>Weller</td>
</tr>
<tr>
<td>11:30 AM</td>
<td>Safe Policy Search with Gaussian Process Models</td>
<td>Polymenakos, Roberts</td>
</tr>
<tr>
<td>11:45 AM</td>
<td>Poster spotlights</td>
<td>Kuwajima, Tanaka, Liang, Komorowski, Que, Drumond, Raghu, Celi, Göpfert, Ross, Tan, Caruana, Lou, Kumar, Taylor, Pourabdizi-Sangdeh, Wortman Vaughan, Wallach</td>
</tr>
<tr>
<td>12:00 PM</td>
<td>Poster session part I</td>
<td></td>
</tr>
<tr>
<td>12:30 PM</td>
<td>Lunch break</td>
<td></td>
</tr>
<tr>
<td>02:00 PM</td>
<td>When the classifier doesn’t know: optimum reject options for classification</td>
<td>Hammer</td>
</tr>
<tr>
<td>02:30 PM</td>
<td>Predict Responsibly: Increasing Fairness by Learning To Defer Abstract</td>
<td>Madras, Zemel, Pitassi</td>
</tr>
<tr>
<td>02:45 PM</td>
<td>Deep Motif Dashboard: Visualizing and Understanding Genomic Sequences Using Deep Neural Networks</td>
<td>Lanchantin, Singh, Wang</td>
</tr>
<tr>
<td>03:00 PM</td>
<td>Coffee break and Poster session part II</td>
<td></td>
</tr>
<tr>
<td>03:30 PM</td>
<td>Invited talk 5</td>
<td>Dragan</td>
</tr>
<tr>
<td>04:05 PM</td>
<td>Invited talk 6</td>
<td>Amodei</td>
</tr>
</tbody>
</table>

Abstracts (6):

Abstract 1: Opening remarks in Transparent and interpretable Machine Learning in Safety Critical Environments, Tosi, Vellido, Álvarez 08:50 AM

Opening remarks and introduction to the Workshop on Transparent and Interpretable Machine Learning in Safety Critical Environments.


Contributed talk 1: "Beyond Sparsity: Tree-based Regularization of Deep Models for Interpretability".

The lack of interpretability remains a key barrier to the adoption of deep models in many healthcare applications. In this work, we explicitly regularize deep models so human users might step through the process behind their predictions in little time. Specifically, we train deep time-series models so their class-probability predictions have high accuracy while being closely modeled by decision trees with few nodes. On two clinical decision-making tasks, we demonstrate that this new tree-based regularization is distinct from simpler L2 or L1 penalties, resulting in more interpretable models without sacrificing predictive power.

Contributed talk 2: "Safe Policy Search with Gaussian Process Models". We propose a method to optimise the parameters of a policy which will be used to safely perform a given task in a data-efficient manner. We train a Gaussian process model to capture the system dynamics, based on the PILCO framework. Our model has useful analytic properties, which allow closed form computation of error gradients and estimating the probability of violating given state space constraints. During training, as well as operation, only policies that are deemed safe are implemented on the real system, minimising the risk of failure.


Contributed talk 2: "Safe Policy Search with Gaussian Process Models". We propose a method to optimise the parameters of a policy which will be used to safely perform a given task in a data-efficient manner. We train a Gaussian process model to capture the system dynamics, based on the PILCO framework. Our model has useful analytic properties, which allow closed form computation of error gradients and estimating the probability of violating given state space constraints. During training, as well as operation, only policies that are deemed safe are implemented on the real system, minimising the risk of failure.


"Network Analysis for Explanation"
"Using prototypes to improve convolutional networks interpretability"
"Accelerated Primal-Dual Policy Optimization for Safe Reinforcement Learning"
"Deep Reinforcement Learning for Sepsis Treatment"
"Analyzing Feature Relevance for Linear Reject Option SVM using Relevance Intervals"
"The Neural LASSO: Local Linear Sparsity for Interpretable
Explanations"
"Detecting Bias in Black-Box Models Using Transparent Model
Distillation"
"Data masking for privacy-sensitive learning"
"CLEAR-DR: Interpretable Computer Aided Diagnosis of Diabetic
Retinopathy"
"Manipulating and Measuring Model Interpretability"

Abstract 12: Predict Responsibly: Increasing Fairness by Learning
to Defer Abstract in Transparent and interpretable Machine
Learning in Safety Critical Environments, Madras, Zemel, Pitassi
02:45 PM

Contributed talk 3: "Predict Responsibly: Increasing Fairness by Learning
to Defer Abstract".

Machine learning systems, which are often used for high-stakes
decisions, suffer from two mutually reinforcing problems: unfairness and
opaqueness. Many popular models, though generally accurate, cannot
express uncertainty about their predictions. Even in regimes where a
model is inaccurate, users may trust the model’s predictions too fully,
and allow its biases to reinforce the user’s own.

In this work, we explore models that learn to defer. In our scheme, a
model learns to classify accurately and fairly, but also to
defer if necessary, passing judgment to a downstream decision-maker
such as a human user. We further propose a learning algorithm which
accounts for potential biases held by decision-makers later in a pipeline.
Experiments on real-world datasets demonstrate that learning to defer
can make a model not only more accurate but also less biased. Even
when operated by biased users, we show that deferring models can still
greatly improve the fairness of the entire pipeline.

Abstract 13: Deep Motif Dashboard: Visualizing and Understanding
Genomic Sequences Using Deep Neural Networks in Transparent
and interpretable Machine Learning in Safety Critical Environments,
Lanchantin, Singh, Wang 02:45 PM

Contributed talk 4: "Deep Motif Dashboard: Visualizing and
Understanding Genomic Sequences Using Deep Neural Networks".

Deep neural network (DNN) models have recently obtained
state-of-the-art prediction accuracy for the transcription factor binding
(TFBS) site classification task. However, it remains unclear how these
approaches identify meaningful DNA sequence signals and give insights
as to why TFs bind to certain locations. In this paper, we propose a
toolkit called the Deep Motif Dashboard (DeMo Dashboard) which
provides a suite of visualization strategies to extract motifs, or sequence
patterns from deep neural network models for TFBS classification. We
demonstrate how to visualize and understand three important DNN
models using three visualization methods: saliency maps, temporal
output scores, and class optimization. In addition to providing insights as
to how each model makes its prediction, the visualization techniques
indicate that CNN-RNN makes predictions by modeling both motifs as
well as dependencies among them.
instruments, speech processing, musical HCI, musicology, music technology, music entertainment, and composition.

Invited Speakers:
Karen Livescu (Toyota Technological Institute at Chicago)
Sander Dieleman (Google DeepMind)
Douglas Eck (Google Magenta)
Marco Marchini (Spotify)
Matthew Prockup (Pandora)

Panel Discussion:
Sepp Hochreiter (Johannes Kepler University Linz)
Karen Livescu (Toyota Technological Institute at Chicago)
Oriol Nieto (Pandora)
Malcolm Slaney (Google)
Hendrik Purwins (Aalborg University Copenhagen)

ML4Audio Organisation Committee:
- Hendrik Purwins, Aalborg University Copenhagen, Denmark
  (hpu@create.aau.dk)
- Bob L. Sturm, Queen Mary University of London, UK
  (b.sturm@qmul.ac.uk)
- Mark Plumbley, University of Surrey, UK (m.plumbley@surrey.ac.uk)

PROGRAM COMMITTEE:
Abeer Alwan (University of California, Los Angeles)
Jon Barker (University of Sheffield)
Sebastian Böck (Johannes Kepler University Linz)
Mads Græsbøll Christensen (Aalborg University)
Maximo Cobos (Universitat de Valencia)
Sander Dieleman (Google DeepMind)
Monika Dörfler (University of Vienna)
Shiromi Dubnov (UC San Diego)
Philippe Esling (IRCAM)
Cédric Févotte (IRIT)
Emilia Gómez (Universitat Pompeu Fabra)
Emanuël Habets (International Audio Labs Erlangen)
Jan Larsen (Danish Technical University)
Marco Marchini (Spotify)
Rafael Ramirez (Universitat Pompeu Fabra)
Gaël Richard (TELECOM ParisTech)
Fatemeh Saki (UT Dallas)
Sanjeev Satheesh (Baidu SAIL)
Jan Schlüter (Austrian Research Institute for Artificial Intelligence)
Joan Serrà (Telefonica)
Malcolm Slaney (Google)
Emmanuel Vincent (INRIA Nancy)
Gerhard Widmer (Austrian Research Institute for Artificial Intelligence)
Tao Zhang (Starkey Hearing Technologies)

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Title</th>
<th>Speaker 1</th>
<th>Speaker 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:00 AM</td>
<td>Overture</td>
<td>Purwins</td>
<td>Purwins</td>
</tr>
<tr>
<td>08:15 AM</td>
<td>Acoustic word embeddings for speech search</td>
<td>Livescu</td>
<td>Livescu</td>
</tr>
<tr>
<td>08:45 AM</td>
<td>Learning Word Embeddings from Speech</td>
<td>Glass, Chung</td>
<td>Glass, Chung</td>
</tr>
<tr>
<td>09:05 AM</td>
<td>Multi-Speaker Localization</td>
<td>Using Convolutional Neural Network Trained with Noise</td>
<td>Chakrabarty, Habets</td>
</tr>
<tr>
<td>09:25 AM</td>
<td>Adaptive Front-ends for End-to-end Source Separation</td>
<td>Venkataramani, Smaragdis</td>
<td>Venkataramani, Smaragdis</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Learning and transforming sound for interactive musical applications</td>
<td>Marchini</td>
<td>Marchini</td>
</tr>
<tr>
<td>11:30 AM</td>
<td>Compact Recurrent Neural Network based on Tensor Train for Polyphonic Music Modeling</td>
<td>Sakti</td>
<td>Sakti</td>
</tr>
<tr>
<td>11:50 AM</td>
<td>Singing Voice Separation using Generative Adversarial Networks</td>
<td>Choi, Lee</td>
<td>Choi, Lee</td>
</tr>
<tr>
<td>12:10 PM</td>
<td>Audio Cover Song Identification using Convolutional Neural Network</td>
<td>Chang</td>
<td>Chang</td>
</tr>
<tr>
<td>12:30 PM</td>
<td>Lunch Break</td>
<td></td>
<td></td>
</tr>
<tr>
<td>01:30 PM</td>
<td>Polyphonic piano transcription using deep neural networks</td>
<td>Eck</td>
<td>Eck</td>
</tr>
<tr>
<td>02:00 PM</td>
<td>Deep learning for music recommendation and generation</td>
<td>Dieleman</td>
<td>Dieleman</td>
</tr>
<tr>
<td>02:30 PM</td>
<td>Exploring Ad Effectiveness using Acoustic Features</td>
<td>Prockup, Vahabi</td>
<td>Prockup, Vahabi</td>
</tr>
<tr>
<td>03:00 PM</td>
<td>Music and environmental sounds</td>
<td>Nieto, Pons, Raj, Tax, Elizalde, Nam, Kumar</td>
<td>Nieto, Pons, Raj, Tax, Elizalde, Nam, Kumar</td>
</tr>
<tr>
<td>04:00 PM</td>
<td>TBD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>04:30 PM</td>
<td>k-shot Learning of Acoustic Context</td>
<td>de Vries</td>
<td>de Vries</td>
</tr>
<tr>
<td>04:50 PM</td>
<td>Towards Learning Semantic Audio Representations from Unlabeled Data</td>
<td>Jansen</td>
<td>Jansen</td>
</tr>
<tr>
<td>05:10 PM</td>
<td>Cost-sensitive detection with variational autoencoders for environmental acoustic sensing</td>
<td>Li, Roberts</td>
<td>Li, Roberts</td>
</tr>
<tr>
<td>05:30 PM</td>
<td>Machine learning and audio signal processing: State of the art and future perspectives</td>
<td>Purwins, Hochreiter, Livescu, Nieto, Slaney</td>
<td>Purwins, Hochreiter, Livescu, Nieto, Slaney</td>
</tr>
</tbody>
</table>
Abstracts (13):

Abstract 2: Acoustic word embeddings for speech search in Machine Learning for Audio Signal Processing (ML4Audio), Livescu 08:15 AM

For a number of speech tasks, it can be useful to represent speech segments of arbitrary length by fixed-dimensional vectors, or embeddings. In particular, vectors representing word segments -- acoustic word embeddings -- can be used in query-by-example search, example-based speech recognition, or spoken term discovery. *Textual* word embeddings have been common in natural language processing for a number of years now; the acoustic analogue is only recently starting to be explored. This talk will present our work on acoustic word embeddings and their application to query-by-example search. I will speculate on applications across a wider variety of audio tasks.

Abstract 3: Learning Word Embeddings from Speech in Machine Learning for Audio Signal Processing (ML4Audio), Glass, Chung 08:45 AM

Yu-An Chung and James R. Glass: Learning Word Embeddings from Speech

In this paper, we propose a novel deep neural network architecture, Sequence-to-Sequence Audio2Vec, for unsupervised learning of fixed-length vector representations of audio segments excised from a speech corpus, where the vectors contain semantic information pertaining to the segments, and are close to other vectors in the embedding space if their corresponding segments are semantically similar. The design of the proposed model is based on the RNN Encoder-Decoder framework, and borrows the methodology of continuous skip-grams for training. The learned vector representations are evaluated on 13 widely used word similarity benchmarks, and achieved competitive results to that of GloVe. The biggest advantage of the proposed model is its capability of extracting semantic information of audio segments taken directly from raw speech, without relying on any other modalities such as text or images, which are challenging and expensive to collect and annotate.

Abstract 4: Multi-Speaker Localization Using Convolutional Neural Network Trained with Noise in Machine Learning for Audio Signal Processing (ML4Audio), Chakrabarty, Habets 09:05 AM

Soumitro Chakrabarty, Emanuël A. P. Habets: Multi-Speaker Localization Using Convolutional Neural Network Trained with Noise

We formulate the multi-speaker localization problem as a multi-class multi-label classification problem, which is solved using a convolutional neural network (CNN) based source localization method. Utilizing the common assumption of disjoint speaker activities, we propose a novel method to train the CNN using synthesized noise signals. The proposed localization method is evaluated for two speakers and compared to a well-known steered response power method.

Abstract 5: Adaptive Front-ends for End-to-end Source Separation in Machine Learning for Audio Signal Processing (ML4Audio), Venkataramani, Smaragdis 09:25 AM

Shrikant Venkataramani, Jonah Casebeer, Paris Smaragdis: Adaptive Front-ends for End-to-end Source Separation

Source separation and other audio applications have traditionally relied on the use of short-time Fourier transforms as a front-end frequency domain representation step. We present an auto-encoder neural network that can act as an equivalent to short-time front-end transforms. We demonstrate the ability of the network to learn optimal, real-valued basis functions directly from the raw waveform of a signal and further show how it can be used as an adaptive front-end for end-to-end supervised source separation.


SPEECH SOURCE SEPARATION
*Minje Kim and Paris Smaragdis. Bitwise Neural Networks for Efficient SingleChannel Source Separation
*Mohit Dubey, Garrett Kenyon, Nils Carlson and Austin Thresher. Does Phase Matter For Monaural Source Separation?
SPEECH ENHANCEMENT
*Rasool Fakoor, Xiaodong He, Ivan Tashev and Shuayb Zarar. Reinforcement Learning To Adapt Speech Enhancement to Instantaneous Input Signal Quality
*Jong Hwan Ko, Josh Fromm, Matthai Phillipose, Ivan Tashev and Shuayb Zarar. Precision Scaling of Neural Networks for Efficient Audio Processing ASR
*Marius Paraschiv, Lasse Borgholt, Tycho Tax, Marco Singh and Lars Maaløe. Exploiting Nontrivial Connectivity for Automatic Speech Recognition
*Brian Mcmahan and Delip Rao. Listening to the World Improves Speech Command Recognition
*Andros Tjandra, Sakriani Sakti and Satoshi Nakamura. End-to-End Speech Recognition with Local Monotonic Attention
*Sri Harsha Dumpala, Rupayan Chakraborty and Sunil Kumar Koppapurapu. A Novel Approach for Effective Learning in Low Resourced Scenarios
SPEECH SYNTHESIS
*Yuxuan Wang, RJ SkerryRyan, Ying Xiao, Daisy Stanton, Joel Shor, Eric Battenberg, Rob Clark and Rif A. Saurous. Uncovering Latent Style Factors for Expressive Speech Synthesis
*Younggun Lee, Azam Rabiiee and Soo-Young Lee. Emotional End-to-End Neural Speech Synthesizer

Abstract 8: Compact Recurrent Neural Network based on Tensor Train for Polyphonic Music Modeling in Machine Learning for Audio Signal Processing (ML4Audio), Sakti 11:30 AM

SPEECH SOURCE SEPARATION
*Minje Kim and Paris Smaragdis: Adaptive Front-ends for End-to-end Source Separation

Source separation and other audio applications have traditionally relied on the use of short-time Fourier transforms as a front-end frequency domain representation step. We present an auto-encoder neural network that can act as an equivalent to short-time front-end transforms. We demonstrate the ability of the network to learn optimal, real-valued basis functions directly from the raw waveform of a signal and further show how it can be used as an adaptive front-end for end-to-end supervised source separation.
This paper introduces a novel compression method for recurrent neural network (RNN) based on Tensor Train (TT) format. The objective in this work is to reduce the number of parameters in RNN and maintain their expressive power.

The key of our approach is by representing the dense matrices weight parameter in the simple RNN and Gated Recurrent Unit (GRU) RNN architectures as the n-dimensional tensor in TT-format. To evaluate our proposed models, we compare it with uncompressed RNN on polyphonic sequence prediction tasks. Our proposed TT-format RNN are able to preserve the performance while reducing the number of RNN parameters significantly up to 80 times smaller.


Hyeong-seok Choi, Kyogu Lee, Ju-heon Lee: Singing Voice Separation using Generative Adversarial Networks

In this paper, we propose a novel approach extending Wasserstein generative adversarial networks (GANs) [3] to separate singing voice from the mixture signal.

We used the mixture signal as a condition to generate singing voices and applied the U-net style network for the stable training of the model. Experiments with the DSD100 dataset show the promising results with the potential of using the GANs for music source separation.

Abstract 10: Audio Cover Song Identification using Convolutional Neural Network in Machine Learning for Audio Signal Processing (ML4Audio), Chang 12:10 PM

Sungkyun Chang, Juheon Lee, Sankeun Choe and Kyogu Lee: Kyogu Lee Audio Cover Song Identification using Convolutional Neural Network

In this paper, we propose a new approach to cover song identification using CNN (convolutional neural network). Most previous studies extract the feature vectors that characterize the cover song relation from a pair of songs and used it to compute the (dis)similarity between the two songs. Based on the observation that there is a meaningful pattern between cover songs and that this can be learned, we have reformulated the cover song identification problem in a machine learning framework. To do this, we first build the CNN using as an input a cross-similarity matrix generated from a pair of songs. We then construct the data set composed of cover song pairs and non-cover song pairs, which are used as positive and negative training samples, respectively. The trained CNN outputs the probability of being in the cover song relation given a cross-similarity matrix generated from any two pieces of music and identifies the cover song by ranking on the probability.

Experimental results show that the proposed algorithm achieves performance better than or comparable to the state-of-the-arts.

Abstract 11: Exploring Ad Effectiveness using Acoustic Features in Machine Learning for Audio Signal Processing (ML4Audio), Prockup, Vahabi 02:30 PM

Online audio advertising is a form of advertising used abundantly in online music streaming services. In these platforms, providing high quality ads ensures a better user experience and results in longer user engagement. In this paper we describe a way to predict ad quality using hand-crafted, interpretable acoustic features that capture timbre, rhythm, and harmonic organization of the audio signal. We then discuss how the characteristics of the sound can be connected to concepts such as the clarity of the ad and its message.

Abstract 12: Music and environmental sounds in Machine Learning for Audio Signal Processing (ML4Audio), Nieto, Pons, Raj, Tax, Elizalde, Nam, Kumar 03:00 PM

MUSIC:

* Jordi Pons, Oriol Nieto, Matthew Prockup, Erik M. Schmidt, Andreas F. Ehmann and Xavier Serra. End-to-end learning for music audio tagging at scale

* Jongpil Lee, Taejun Kim, Jiyoung Park and Juhan Nam. Raw Waveform based Audio Classification Using Samplelevel CNN Architectures

* Alfonso Perez-Carrillo Estimation of violin bowing features from Audio recordings with Convolutional Networks

ENVIRONMENTAL SOUNDS:

* Bhiksha Raj, Benjamin Elizalde, Rohan Badlani, Ankit Shah and Anurag Kumar. NELS NeverEnding Learner of Sounds

* Tycho Tax, Jose Antich, Hendrik Purwins and Lars Maalee Utilizing Domain Knowledge in End-to-End Audio Processing

* Anurag Kumar and Bhiksha Raj. Deep CNN Framework for Audio Event Recognition using Weakly Labeled Web Data

Abstract 13: k-shot Learning of Acoustic Context in Machine Learning for Audio Signal Processing (ML4Audio), de Vries 04:30 PM

Ivan Bocharov, Bert de Vries, and Tjalling Tjalkens: k-shot Learning of Acoustic Context

In order to personalize the behavior of hearing aid devices in different acoustic scenes, we need personalized acoustic scene classifiers. Since we cannot afford to burden an individual hearing aid user with the task to collect a large acoustic database, we will want to train an acoustic scene classifier on one in-situ recorded waveform (of a few seconds duration) per class. In this paper we develop a method that achieves high levels of
classification accuracy from a single recording of an acoustic scene.

Abstract 18: Towards Learning Semantic Audio Representations from Unlabeled Data in Machine Learning for Audio Signal Processing (ML4Audio), Jansen 04:50 PM

Aren Jansen, Manoj Plakal, Ratheet Pandya, Daniel P. W. Ellis, Shawn Hershey, Jiayang Liu, R. Channing Moore, Rif A. Saurous: Towards Learning Semantic Audio Representations from Unlabeled Data

Our goal is to learn semantically structured audio representations without relying on categorically labeled data. We consider several class-agnostic semantic constraints that are inherent to non-speech audio: (i) sound categories are invariant to additive noise and translations in time, (ii) mixtures of two sound events inherit the categories of the constituents, and (iii) the categories of events in close temporal proximity in a single recording are likely to be the same or related. We apply these invariants to samples training data for triplet-loss embedding models using a large unlabeled dataset of YouTube soundtracks. The resulting low-dimensional representations provide both greatly improved query-by-example retrieval performance and reduced labeled data and model complexity requirements for supervised sound classification.

Abstract 19: Cost-sensitive detection with variational autoencoders for environmental acoustic sensing in Machine Learning for Audio Signal Processing (ML4Audio), Li, Roberts 05:10 PM

Yunpeng Li, Ivan Kiskin, Marianne Sinka, Henry Chan, Stephen Roberts: Cost-sensitive detection with variational autoencoders for environmental acoustic sensing

Environmental acoustic sensing involves the retrieval and processing of audio signals to better understand our surroundings. While large-scale acoustic data make manual analysis infeasible, they provide a suitable playground for machine learning approaches. Most existing machine learning techniques developed for environmental acoustic sensing do not provide flexible control of the trade-off between the false positive rate and the false negative rate. This paper presents a cost-sensitive classification paradigm, in which the structures of variational autoencoders and the hyper-parameters of classifiers are selected in a principled Neyman-Pearson framework. We examine the performance of the proposed approach using a dataset from the HumBug project which aims to detect the presence of mosquitoes using sound collected by simple embedded devices.

Deep Learning for Physical Sciences

Atilim Gunes Baydin, Mr. Prabhat, Kyle Cranmer, Frank Wood

Room 202, Fri Dec 08, 08:00 AM

Physical sciences span problems and challenges at all scales in the universe: from finding exoplanets and asteroids in trillions of sky-survey pixels, to automatic tracking of extreme weather phenomena in climate datasets, to detecting anomalies in event streams from the Large Hadron Collider at CERN. Tackling a number of associated data-intensive tasks, including, but not limited to, regression, classification, clustering, dimensionality reduction, likelihood-free inference, generative models, and experimental design are critical for furthering scientific discovery. The Deep Learning for Physical Sciences (DLPS) workshop invites researchers to contribute papers that demonstrate progress in the application of machine and deep learning techniques to real-world problems in physical sciences (including the fields and subfields of astronomy, chemistry, Earth science, and physics).

We will discuss research questions, practical implementation challenges, performance / scaling, and unique aspects of processing and analyzing scientific datasets. The target audience comprises members of the machine learning community who are interested in scientific applications and researchers in the physical sciences. By bringing together these two communities, we expect to strengthen dialogue, introduce exciting new open problems to the wider NIPS community, and stimulate production of new approaches to solving science problems. Invited talks from leading individuals from both communities will cover the state-of-the-art techniques and set the stage for this workshop.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:50 AM</td>
<td>Introduction and opening remarks</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>Invited talk 1: Deep recurrent inverse modeling for radio astronomy and fast MRI imaging Welling</td>
</tr>
<tr>
<td>09:40 AM</td>
<td>Contributed talk 1: Neural Message Passing for Jet Physics Henrion</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Contributed talk 2: A Foray into Using Neural Network Control Policies For Rapid Switching Between Beam Parameters in a Free Electron Laser Edelen</td>
</tr>
</tbody>
</table>
Machine Deception

**Ian Goodfellow, Tim Hwang, Bryce Goodman, Mikel Rodriguez**

Poster session 1 and coffee break

<table>
<thead>
<tr>
<th>Time</th>
<th>Session/Panel</th>
<th>Speakers/Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:20 AM</td>
<td>Invited talk 2: TBA</td>
<td></td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Contributed talk 3: Implicit Causal Models for Genome-wide Association Studies</td>
<td>Tran</td>
</tr>
<tr>
<td>12:00 PM</td>
<td>Contributed talk 4: Graphite: Iterative Generative Modeling of Graphs</td>
<td>Grover</td>
</tr>
<tr>
<td>12:20 PM</td>
<td>Lunch break</td>
<td></td>
</tr>
<tr>
<td>02:00 PM</td>
<td>Invited talk 3: TBA</td>
<td>Murray</td>
</tr>
<tr>
<td>02:40 PM</td>
<td>Contributed talk 5: Deep Learning for Real-time Gravitational Wave Detection and Parameter Estimation with Real LIGO Data</td>
<td>George</td>
</tr>
<tr>
<td>03:00 PM</td>
<td>Poster session 2 and coffee break</td>
<td></td>
</tr>
<tr>
<td>04:00 PM</td>
<td>Invited talk 4: A machine learning perspective on the many-body problem in classical and quantum physics</td>
<td>Carrasquilla</td>
</tr>
<tr>
<td>04:40 PM</td>
<td>Invited talk 5: TBA</td>
<td>von Lilienfeld</td>
</tr>
<tr>
<td>05:20 PM</td>
<td>Contributed talk 6: Physics-guided Learning of Neural Networks: An Application in Lake Temperature Modeling</td>
<td>Karpatne</td>
</tr>
<tr>
<td>05:40 PM</td>
<td>Panel session: TBA</td>
<td></td>
</tr>
<tr>
<td>06:40 PM</td>
<td>Closing remarks</td>
<td></td>
</tr>
</tbody>
</table>

**ML Systems Workshop @ NIPS 2017**

**Aparna Lakshmiratan, Sarah Bird, Siddhartha Sen, Chris Ré, Li Erran Li, Joseph Gonzalez, Dan Crankshaw**

Room-204, Fri Dec 08, 08:00 AM

A new area is emerging at the intersection of artificial intelligence, machine learning, and systems design. This birth is driven by the explosive growth of diverse applications of ML in production, the continued growth in data volume, and the complexity of large-scale learning systems. The goal of this workshop is to bring together experts working at the crossroads of machine learning, system design and software engineering to explore the challenges faced when building practical large-scale ML systems. In particular, we aim to elicit new connections among these diverse fields, and identify tools, best practices and design principles. We also want to think about how to do research in...
this area and properly evaluate it. The workshop will cover ML and AI platforms and algorithm toolkits, as well as dive into machine learning-focused developments in distributed learning platforms, programming languages, data structures, GPU processing, and other topics.

This workshop will follow the successful model we have previously run at ICML, NIPS and SOSP 2017.

Our plan is to run this workshop annually at one ML venue and one Systems venue, and eventually merge these communities into a full conference venue. We believe this dual approach will help to create a low barrier to participation for both communities.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:45 AM</td>
<td>Opening Remarks</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>Invited Talk: Ray: A distributed execution engine for emerging AI applications, Ion Stoica, UC Berkeley</td>
</tr>
<tr>
<td>09:20 AM</td>
<td>Contributed Talk 1</td>
</tr>
<tr>
<td>09:40 AM</td>
<td>Invited Talk: Federated Multi-Task Learning, Virginia Smith, Stanford University</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Poster Previews: 2 min lightning talks</td>
</tr>
<tr>
<td>11:30 AM</td>
<td>Invited Talk: Accelerating Persistent Neural Networks at Datacenter Scale, Daniel Lo, Microsoft Research</td>
</tr>
<tr>
<td>11:50 AM</td>
<td>Contributed Talk 2</td>
</tr>
<tr>
<td>12:10 PM</td>
<td>Lunch</td>
</tr>
<tr>
<td>01:30 PM</td>
<td>Updates from Current ML Systems (TensorFlow, PyTorch, Caffe2, CNTK, MxNet, TVM, Clipper, DAWN)</td>
</tr>
<tr>
<td>02:50 PM</td>
<td>Invited Talk: Machine Learning for Systems and Systems for Machine Learning, Jeff Dean, Google Brain</td>
</tr>
<tr>
<td>03:20 PM</td>
<td>Invited Talk: Creating an Open and Flexible ecosystem for AI models with ONNX, Sarah Bird, Facebook Research</td>
</tr>
</tbody>
</table>

03:40 PM  | Posters and Coffee                        |
04:30 PM  | Contributed Talk 3                        |
04:50 PM  | Contributed Talk 4                        |
05:10 PM  | Panel                                     |

Machine Learning for the Developing World

Maria De-Arteaga, William Herlands

Room S-1, Fri Dec 08, 08:00 AM

Six billion people live in developing world countries. The unique development challenges faced by these regions have long been studied by researchers ranging from sociology to statistics and ecology to economics. With the emergence of mature machine learning methods in the past decades, researchers from many fields - including core machine learning - are increasingly turning to machine learning to study and address challenges in the developing world. This workshop is about delving into the intersection of machine learning and development research.

Machine learning present tremendous potential to development research and practice. Supervised methods can provide expert telemedicine decision support in regions with few resources; deep learning techniques can analyze satellite imagery to create novel economic indicators; NLP algorithms can preserve and translate obscure languages, some of which are only spoken. Yet, there are notable challenges with machine learning in the developing world. Data cleanliness, computational capacity, power availability, and internet accessibility are more limited than in developed countries. Additionally, the specific applications differ from what many machine learning researchers normally encounter. The confluence of machine learning's immense potential with the practical challenges posed by developing world settings has inspired a growing body of research at the intersection of machine learning and development research.

This one-day workshop is focused on machine learning for the developing world, with an emphasis on developing novel methods and technical applications that address core concerns of developing regions. We will consider a wide range of development areas including health, education, institutional integrity, violence mitigation, economics, societal analysis, and environment. From the machine learning perspective we are open to all methodologies with an emphasis on novel techniques inspired by particular use cases in the developing world.

Invited speakers will address particular areas of interest, while poster sessions and a guided panel discussion will encourage interaction between attendees. We wish to review the current approaches to machine learning in the developing world, and inspire new approaches and paradigms that can lay the groundwork for substantial innovation.
The nematode worm, *C. elegans*, provides a ready experimental system for reverse-engineering the nervous system, being one of the best studied animals in the life sciences. The neural connectome of *C. elegans* has been known for 30 years, providing the structural basis for building models of its neural information processing. Despite its small size, *C. elegans* exhibits complex behaviors, such as, locating food, mating partners and navigating its environment by integrating a plethora of environmental cues. Over the past years, the field has made an enormous progress in understanding some of the neural circuits that control sensory processing, decision making and locomotion. In laboratory, the crawling behavior of worms occurs mainly in 2D. This enables the use of machine learning tools to obtain quantitative behavioral descriptions of unprecedented accuracy. Moreover, neuronal imaging techniques have been developed so that the activity of nearly all nerve cells in the brain can be recorded in real time. Leveraging on these advancements, the community wide *C. elegans* OpenWorm project will make a realistic in silico simulation of a nervous system and the behavior it produces possible, for the first time.

The goal of this workshop is to gather researchers in neuroscience and machine learning together, to advance understanding of the neural information processing of the worm and to outline what challenges still lie ahead. We particularly aim to:

- Comprehensively, introduce the nervous system of *C. elegans*. We will discuss the state-of-the-art findings and potential future solutions for modeling its neurons and synapses, complete networks of neurons and the various behaviors of the worm.
- Identify main challenges and their solutions in behavioral and neural data extraction, such as imaging techniques, generation of time series data from calcium imaging records and high resolution behavioral data, as well as cell recognition, cell tracking and image segmentation.
- Explore machine learning techniques for interpretation of brain data, such as time series analysis, feature extraction methods, complex network analysis, complex nonlinear systems analysis, large-scale parameter optimization methods, and representation learning.
- Get inspirations from this well-understood brain to design novel network architectures, control algorithms and neural processing units.

We have invited leading neuroscientists, machine learning scientists and interdisciplinary experts, to address these main objectives of the workshop, in the form of Keynote talks and a panel discussion. We also invite submissions of 4-page papers for posters, spotlight presentations and contributed talks, and offer travel awards.

Topics of interests are: Deep learning applications in nervous system data analysis, neural circuits analysis, behavior modeling, novel computational approaches and algorithms for brain data interpretations, brain simulation platforms, optimization algorithms for nonlinear systems, applications of machine learning methods to brain data and cell biology, complex network analysis, cell modeling, cell recognition and tracking, dynamic modeling of neural circuits and genetic regulatory networks.

The workshop’s webpage: https://sites.google.com/site/wwnip2017/

---

### Workshop on Worm’s Neural Information Processing (WNIP)

**Ramin Hasani, Manuel Zimmer, Stephen Larson, Radu Grosu**

**Room S-3, Fri Dec 08, 08:00 AM**

A fundamental Challenge in neuroscience is to understand the elemental computations and algorithms by which brains perform information processing. This is of great significance to biologists, as well as, to engineers and computer scientists, who aim at developing energy efficient and intelligent solutions for the next generation of computers and autonomous devices. The benefits of collaborations between these fields are reciprocal, as brain-inspired computational algorithms and devices not only advance engineering, but also assist neuroscientists by conforming their models and making novel predictions. A large impediment toward such an efficient interaction is still the complexity of brains. We thus propose that the study of small model organisms should pioneer these efforts.

The nematode worm, *C. elegans*, provides a ready experimental system
turning locomotion. In contrast to global stereotypy, variability is evident
6D space, which is globally composed of three sets of cyclic trajectories
apply this approach to C. elegans and show that the dynamics lie on a
unfolded as a combination of postures and their short-time dynamics. We
behavioral phase space in which the instantaneous state is smoothly
ignores important variability within each motif. Here, we develop a
Abstract 2: Capturing the continuous complexity of natural behavior in Workshop on Worm's Neural Information Processing (WNIP), Stephens 09:15 AM

While animal behavior is often quantified through discrete motifs, this is only an approximation to fundamentally continuous dynamics and ignores important variability within each motif. Here, we develop a behavioral phase space in which the instantaneous state is smoothly unfolded as a combination of postures and their short-time dynamics. We apply this approach to C. elegans and show that the dynamics lie on a 6D space, which is globally composed of three sets of cyclic trajectories that form the animal’s basic behavioral motifs: forward, backward and turning locomotion. In contrast to global stereotypy, variability is evident by the presence of locally-unstable dynamics for each set of cycles. Across the full phase space we show that the Lyapunov spectrum is symmetric with positive, chaotic exponents driving variability balanced by negative, dissipative exponents driving stereotypy. The symmetry of the spectrum holds for different environments and for human walking, suggesting a general condition of motor control. Finally, we use the reconstructed phase space to analyze the complexity of the dynamics along the worm’s body and find evidence for multiple, spatially-separate oscillators driving C. elegans locomotion.

Abstract 3: Neuronal analysis of value-based decision making in C. elegans in Workshop on Worm’s Neural Information Processing (WNIP), Lockery 09:45 AM

Decision making is a central function of the brain and the focus of intensive study in neuroscience, psychology, and economics. Value-based decision making (e.g., ‘which fragrance do you prefer?’ not ‘which smells more like roses?’) guides significant, sometimes life-changing, choices yet its neuronal basis is poorly understood. Research into this question would be accelerated by the introduction of genetically tractable invertebrates with small nervous systems, like the Drosophila and C. elegans.

We have recently shown that the nematode C. elegans makes value-based decisions. This was done using a formal economic method – the Generalized Axiom of Revealed Preference (GARP). The basis of the method is to establish that the subject's choices are internally consistent with respect to transitivity (A > B > C ⇒ A > C). In the wild, C. elegans feeds on a variety of bacteria and learns to prefer the more nutritious species. We tested worms on a set of decisions between a high quality species and a low quality species at a range of relative concentrations and found the worm’s choices to be 100% transitive, the necessary and sufficient condition for value-based decision making. Further, we found that the olfactory neuron AWC, known to be activated by the sudden absence of food, is required for intact food choice behavior. Surprisingly, however, we found that AWC is also activated by the switch from high quality food to low quality food, even when the two foods are at the same concentration. Thus, food value may be represented at the level of individual olfactory neurons.

We are now investigating the neural mechanisms of choice transitivity. C. elegans selects food sources utilizing klinotaxis, a chemotaxis strategy during locomotion in which the worm’s head bends more deeply on the side of preferred food. The chemosensory neurons, interneurons, and motor neurons of a candidate circuit for klinotaxis have been identified. Extrapolating from our findings with respect to AWC, we have developed a model of the circuit in which distinct chemo-sensory neuron types encode food quality and quantity during particular phases of head bending. Activation of downstream interneurons in the model is the weighted sum of these inputs in accordance with phase information. The model proposes that the signs and strengths of synaptic weights in the biological circuit are adjusted to ensure that subjective value is a monotonic function of the relative quantity of high and low quality food, a property that guarantees transitivity under GARP. Work in progress tests the model using calcium imaging, optogenetic activation, and ablations of each neuron in the circuit.

Abstract 6: Mechanisms and Functions of Neuronal Population Dynamics in C. elegans in Workshop on Worm’s Neural Information Processing (WNIP), Zimmer 11:00 AM

Populations of neurons in the brains of many different animals, ranging from invertebrates to primates, typically coordinate their activities to generate low dimensional and transient activity dynamics, an operational principle serving many neuronal functions like sensory coding, decision
making and motor control. However, the mechanism that bind individual neurons to global population states are not yet known. Are population dynamics driven by a smaller number of pacemaker neurons or are they an emergent property of neuronal networks? What are the features in global network architecture that support coordinated network wide dynamics?

In order to address these problems, we study neuronal population dynamics in C. elegans. We recently developed a calcium imaging approach to record the activity of nearly all neuron in the worm brain in real time and at single cell resolution. We show that brain activity of C. elegans is dominated by brain wide coordinated population dynamics involving a large fraction of interneurons and motor neurons. The activity patterns of these neuronal ensembles recur in an orderly and cyclical fashion. In subsequent experiments, we characterized these brain dynamics functionally and found that they represent action commands and their assembly into a typical action sequence of these animals: forward crawling – backward crawling – turning.

Deciphering the mechanisms underlying neuronal population dynamics is key to understanding the principal computations performed by neuronal networks in the brains of animals, and perhaps will inspire the design of novel machine learning algorithms for robotic control. In this talk, I will discuss three of our approaches to uncover these mechanisms:

First, using graph theory, we aim to identify the key features of neuronal network architecture that support functional dynamics. We found that rich club neurons, i.e. highly interconnected network hubs contribute most to brain dynamics. However, simple measures of synaptic connectivity (e.g. connection strength) failed to predict functional interactions between these neurons; unlike higher order network statistics that measure the similarity in synaptic input patterns.

We next performed systematic perturbations by interrogation of rich club neurons via transgenic neuronal inhibition tools. Using whole brain imaging in combination with computational analysis methods we found that upon inhibition of critical hubs, leading to a disintegration of the network, most other individual neurons remain vigorously active, however the global coordination across neurons was abolished. Based on these results we hypothesize that neuronal population dynamics are an emergent property of neuronal networks.

Finally, we aim to recapitulate C. elegans brain dynamics in silico. Here, we generate neuronal network simulations based on deterministic and stochastic biophysical models of neurons and synapses, at multiscale levels of abstraction. We then adopt a genetic algorithm for neuronal circuit parameter optimization, to find the best matches between simulations and measured calcium dynamics. This approach enables us to test our hypotheses, and to predict unknown properties of neural circuits important for brain dynamics.

Abstract 7: From salt navigation in Caenorhabditis elegans to robot navigation in urban environments, or: the role of sensory computation in balancing exploration and exploitation during animal search in Workshop on Worm’s Neural Information Processing (WNIP), Cohen 11:30 AM

Effective spatial navigation is essential for the survival of animals. Navigation, or the search for favorable conditions, is fundamentally an adaptive behavior that can depend on the changing environment, the animal's past history of success and failure and its internal state. C. elegans implements combinations of systematic and stochastic navigational strategies that are modulated by plasticity across a range of time scales. Here, we combine experiments and computational modeling to characterise adaptation in gustatory and nociceptive salt sensing neurons and construct a simulation framework in which animals can navigate a virtual environment. Our model, and simulations on a variety of smooth, rugged or complex landscapes, suggest that these different forms of sensory adaptation combine to dynamically modulate navigational strategies, giving rise to effective exploration and navigation of the environment. Inspired by this compact and elegant sensory circuit, we present a robotic simulation framework, capable of robustly searching for landmarks in a toy simulation environment.

Abstract 9: Using Network Control Principles to Probe the Structure and Function of Neuronal Connectomes in Workshop on Worm’s Neural Information Processing (WNIP), Schafer 12:15 PM

William R. Schafer1, Gang Yan2, 3, Petra E. Vértes4, Emma K. Towlson3, Yee Lian Chew1, Denise S. Walker1, & Albert-László Barabási3
1Division of Neurobiology, MRC Laboratory of Molecular Biology, Cambridge Biomedical Campus, Francis Crick Avenue, Cambridge CB2 0QH, UK.
2School of Physics Science and Engineering, Tongji University, Shanghai 200092, China.
3Center for Complex Network Research and Department of Physics, Northeastern University, Boston, Massachusetts 02115, USA.
4Department of Psychiatry, Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge CB2 0SZ, UK.

Large-scale efforts are underway to map the neuronal connectomes of many animals, from flies to humans. However, even for small connectomes, such as that of C. elegans, it has been difficult to relate the structure of neuronal wiring patterns to the function of neural circuits. Recent theoretical studies have suggested that control theory might provide a framework to understand structure-function relationships in complex biological networks, including neuronal connectomes. To test this hypothesis experimentally, we have used the complete neuronal connectome of C. elegans to identify neurons predicted to affect the controllability of the body muscles and assess the effect of ablating these neurons on locomotor behavior. We identified 12 neural classes whose removal from the connectome reduced the structural controllability of the body neuromusculature, one of which was the uncharacterized PDB motorneuron. Consistent with the control theory prediction, ablation of PDB had a specific effect on locomotion, altering the dorsoventral polarity of large turns. Control analysis also predicted that three members of the DD motorneuron class (DD4, DD5 and DD6) are individually required for body muscle controllability, while more anterior DDs (DD1, DD2 and DD3) are not. Indeed, we found that ablation of DD4 or DD5, but not DD2 or DD3, led to abnormalities in posterior body movements, again consistent with control theory predictions. We are currently using the control framework to probe other parts of the C. elegans connectome, and are developing more sophisticated approaches behavioral analysis in order to more precisely relate ablation phenotypes to specific muscle groups. We anticipate that the control framework validated by this work may have application in the analysis of larger neuronal connectomes and other complex networks.

Abstract 11: Biological Neurons Are Different From Neural Networks: Simulating C.elegans in an open science project in Workshop on Worm’s Neural Information Processing (WNIP), Larson 02:00 PM

The membrane potential of a biological neuron is considered to be one of the most important properties to understand its dynamic state. While the action potential or discrete “spike” featureof mammalian neurons has been emphasized as an information bearing signal, biological evidence
exists that even without action potentials, neurons process information and give rise to different behavioral states. Nowhere is this more evident than in the nematode worm C. elegans, where its entire nervous system of 302 neurons, despite a lack of action potentials, organizes complex behaviors such as mating, predator avoidance, location of food sources, and many others.

For thirty years, the C. elegans nervous system has remained the only adult animal that has had its nervous system connectivity mapped at the level of individual synapses and gap junctions. As part of the international open science collaboration known as OpenWorm, we have built a simulation framework, known as c302, that enables us to assemble the known connectivity and other biological data of the C. elegans nervous system into a Hodgkin-Huxley-based simulation that can be run in the NEURON simulation engine.

Using a physical simulation of the C. elegans body, known as Sibernetic, we have injected simple sinusoidal activation patterns of the muscle cells of the C. elegans and produced simple crawling and swimming behavior. With the goal of producing the same simple sinusoids in the muscle cells, we have used c302 to select a subnetwork from the full C. elegans nervous system and used machine learning techniques to fit dynamic parameters that are underspecified by the data. Our preliminary results still leave many important biological features out, but initially demonstrate that it is possible to make motor neurons produce sinusoidal activity patterns in the muscles as used in the physical simulation.

In this talk I will discuss these initial results and discuss future directions for a better understanding of the information processing underlying the C. elegans' nervous system.

Abstract 14: Evolving Neural Circuits for Behavior: C. elegans Locomotion in Workshop on Worm's Neural Information Processing (WNIP), Izquierdo 03:30 PM

One of the grand scientific challenges of this century is to understand how behavior is grounded in the interaction between an organism's brain, its body, and its environment. Although a lot of attention and resources are focused on understanding the human brain, I will argue that the study of simpler organisms is an ideal place to begin to address this challenge. I will introduce the nematode worm Caernohabditis elegans, with just 302 neurons, the only fully-reconstructed connectome at the cellular level, and a rich behavioral repertoire that we are still discovering. I will describe a computational approach to address such grand challenge. I will lay out some of the advantages of expressing our understanding in equations and computational models rather than just words. I will describe our unique methodology for exploring the unknown biological parameters of the model through the use of evolutionary algorithms. We train the neural networks on what they should do, with little or no instructions on how to do it. The effort is then to analyze and understand the evolved solutions as a way to generate novel, often unexpected, hypotheses. As an example, I will focus on how the rhythmic pattern is both generated and propagated along the body during locomotion.

Synergies in Geometric Data Analysis

Marina Meila, Frederic Chazal

Room S-5, Fri Dec 08, 08:00 AM

This two-day workshop will bring together researchers from the various subdisciplines of Geometric Data Analysis, such as manifold learning, topological data analysis, shape analysis, will showcase recent progress in this field and will establish directions for future research. The focus will be on high dimensional and big data, and on mathematically founded methodology. Specific aims =========== One aim of this workshop is to build connections between Topological Data Analysis on one side and Manifold Learning on the other. This is starting to happen, after many years of more or less separate evolution of the two fields. The moment has been reached when the mathematical, statistical and algorithmic foundations of both areas are mature enough -- it is now time to lay the foundations for joint topological and differential geometric understanding of data, and this workshop will explicitly focus on this process. The second aim is to bring GDA closer to real applications. We see the challenge of real problems and real data as a motivator for researchers to explore new research questions, to reframe and expand the existing theory, and to step out of their own sub-area. In particular, for people in GDA to see TDA and ML as one. The impact of GDA in practice also depends on having scalable implementations of the most current results in theory. This workshop will showcase the GDA tools which achieve this and initiate a collective discussion about the tools that need to be built. We intend this workshop to be a forum for researchers in all areas of Geometric Data Analysis. Through the tutorials, we are reaching out to the wider NIPS audience, to the many potential users of Geometric Data Analysis, to make them aware of the state of the art in GDA, and of the tools available. Last but not least, we hope that the scientists invited will bring these methods back to their communities.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Presenters</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:10 AM</td>
<td>TBA</td>
<td>Harer</td>
</tr>
<tr>
<td>09:10 AM</td>
<td>Estimating the Reach of a Manifold</td>
<td>Aamari</td>
</tr>
<tr>
<td>09:10 AM</td>
<td>Multiscale geometric feature extraction</td>
<td>Chandler</td>
</tr>
<tr>
<td>09:40 AM</td>
<td>Multiscale geometric feature extraction</td>
<td>Polonik</td>
</tr>
<tr>
<td>10:10 AM</td>
<td>Poster spotlights</td>
<td></td>
</tr>
<tr>
<td>10:15 AM</td>
<td>A dual framework for low rank tensor completion</td>
<td></td>
</tr>
<tr>
<td>10:15 AM</td>
<td>Maximum likelihood estimation of Riemannian metrics from Euclidean data</td>
<td></td>
</tr>
<tr>
<td>10:15 AM</td>
<td>Persistent homology of KDE filtration of Rips complexes</td>
<td></td>
</tr>
<tr>
<td>10:15 AM</td>
<td>Functional data analysis using a topological summary statistic</td>
<td></td>
</tr>
<tr>
<td>10:15 AM</td>
<td>Parallel multi-scale reduction of persistent homology</td>
<td>Mendoza Smith</td>
</tr>
<tr>
<td>02:00 PM</td>
<td>Multiscale characterization of molecular dynamics</td>
<td>Clementi</td>
</tr>
</tbody>
</table>
Furthermore, whereas the massively parallelised. Simulations on synthetic examples show that the approach gives an alternative method by which barcodes are partly revealed for multiple scales simultaneously and further refined as the algorithm progresses. Specifically, our numerical experiments show that for a Vietoris-Rips filtration with \(10^4\) simplices, the essential topological information can be estimated with 95% precision in two iterations and that the reduction completed to within 1% in about ten iterations of our algorithm as opposed to nearly approximately eight thousand iterations for traditional methods.

Abstract 12: Discussion: Geometric Data Analysis in Synergies in Geometric Data Analysis, Chazal, Meila 05:00 PM

One aim of this workshop is to build connections between Topological Data Analysis on one side and Manifold Learning on the other. The moment has been reached when the mathematical, statistical and algorithmic foundations of both areas are mature enough -- it is now time to lay the foundations for joint topological and differential geometric understanding of data, and this discussion will explicitly focus on this process.

The second aim is to bring GDA closer to real applications. We see the challenge of real problems and real data as a motivator for researchers to explore new research questions, to reframe and expand the existing theory, and to step out of their own sub-area.

Abstract 13: Topological Data Analysys with GUDHI and scalable manifold learning and clustering with megaman in Synergies in Geometric Data Analysis, Rouvreau, Meila 08:00 AM

Presentation and demo of the Gudhi library for Topological Data Analysis, followed by a presentation of the magaman package.

The aim of the presentations will be give an introduction for beginners into the practical side of GDA, and to give an overview of the software capabilities. The presenters will leave ample time for questions and will be available during poster sessions for more detailed discussions and demos.

http://gudhi.gforge.inria.fr/
http://github.com/mmp2/megaman

Abstract 17: Geometric Data Analysis software in Synergies in Geometric Data Analysis, Rouvreau, Meila 10:50 AM

We invite the GDA community to discuss the goods, the bads and the ways forward in the software for GDA.

[NOTE THE START TIME: 10 minutes before the official end of the break]

**Competition track**

**Sergio Escalera, Markus Weimer**

**Room S-7, Fri Dec 08, 08:00 AM**

This is the first NIPS edition on “NIPS Competitions”. We received 23 competition proposals related to data-driven and live competitions on different aspects of NIPS. Proposals were reviewed by several qualified researchers and experts in challenge organization. Five top-scored...
competitions were accepted to be run and present their results during the NIPS 2017 Competition track day. Evaluation was based on the quality of data, problem interest and impact, promoting the design of new models, and a proper schedule and managing procedure. Below, you can find the five accepted competitions. Organizers and participants in these competitions will be invited to present their work to this workshop, to be held on December 8th. Accepted competitions:

- The Conversational Intelligence Challenge
- Classifying Clinically Actionable Genetic Mutations
- Learning to Run
- Human-Computer Question Answering Competition
- Adversarial Attacks and Defences

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Presenters</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:15 AM</td>
<td>Opening</td>
<td>Escalera, Weimer</td>
</tr>
<tr>
<td>08:30 AM</td>
<td>AI XPRIZE Milestone award</td>
<td>Banifatemi, Cooper</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>Competition I: Adversarial Attacks and Defenses</td>
<td>Kurakin, Goodfellow, Bengio, Zhao</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Coffee Break and Competition I</td>
<td></td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Competition II: Learning to Run</td>
<td>Kidzinski, Ong, Mohanty, Fries, Hicks</td>
</tr>
<tr>
<td>12:15 PM</td>
<td>Lunch</td>
<td></td>
</tr>
<tr>
<td>01:15 PM</td>
<td>DeepArt competition</td>
<td>Ecker, Gatys, Bethge</td>
</tr>
<tr>
<td>01:30 PM</td>
<td>Competition III: The Conversational Intelligence Challenge</td>
<td>Burtsev, Lowe, Serban, Bengio, Rudnicky, Black, Prabhunoye, Rodichev, Smetanin, Fedorenko, Lee, HONG, Lee, Kim, Gontier, Saito</td>
</tr>
<tr>
<td>03:00 PM</td>
<td>Coffee break and Competition III: The Conversational Intelligence Challenge posters</td>
<td></td>
</tr>
<tr>
<td>03:30 PM</td>
<td>Competition IV: Classifying Clinically Actionable Genetic Mutations</td>
<td>Huerga, Grigorenko, Das, Thorbergsson, Kan, Zhang, Chen, Li, Kumar, Gao</td>
</tr>
<tr>
<td>05:15 PM</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>05:45 PM</td>
<td>Competition V: Human-Computer Question Answering</td>
<td>Ying, Daume III, He, Iyyer, Rodriguez</td>
</tr>
<tr>
<td>07:30 PM</td>
<td>Closing</td>
<td>Escalera, Weimer</td>
</tr>
<tr>
<td>07:30 PM</td>
<td>Dinner</td>
<td></td>
</tr>
</tbody>
</table>

Abstract 3: Competition I: Adversarial Attacks and Defenses in Competition track, Kurakin, Goodfellow, Bengio, Zhao 09:00 AM

- * Introduction into adversarial examples. Invited speaker, Dawn Song
- * Overview of the competition, Alexey Kurakin, Ian Goodfellow
- * Winner of attack competition, Yinpeng Dong, Fangzhou Liao, Tianyu Pang
- * Winner of defense competition, Yinpeng Dong, Fangzhou Liao, Tianyu Pang
- * 2nd place defense competition, Cihang Xie

Abstract 5: Competition II: Learning to Run in Competition track, Kidzinski, Ong, Mohanty, Fries, Hicks 10:30 AM

- * Overview of the Challenge
- * Keynote: Carmichael Ong
- * Challenge logistics (CrowdAI platform)
- * Run top submissions on the final test environment
- * Talks from top participants
- * AWS & NVIDIA sponsored prize ceremony
- * Keynote: Sergey Levine

Abstract 7: DeepArt competition in Competition track, Ecker, Gatys, Bethge 01:15 PM

- Competition review, results, and award ceremony

Abstract 8: Competition III: The Conversational Intelligence Challenge in Competition track, Burtsev, Lowe, Serban, Bengio, Rudnicky, Black, Prabhunoye, Rodichev, Smetanin, Fedorenko, Lee, HONG, Lee, Kim, Gontier, Saito 01:30 PM

- * Overview of the Challenge
- * Awarding prize
- * Short presentation by winning team
- * Inspirational talk
- * Panel Discussion

Abstract 12: Competition V: Human-Computer Question Answering in Competition track, Ying, Daume III, He, Iyyer, Rodriguez 05:45 PM

- * Overview of the Challenge
- * Awarding prize
- * Short presentation by winning team
- * Inspirational talk
- * Panel Discussion
Machine Learning and Computer Security

Jacob Steinhardt, Nicolas Papernot, Bo Li, Chang Liu, Percy Liang, Dawn Song

S-4, Fri Dec 08, 08:00 AM

While traditional computer security relies on well-defined attack models and proofs of security, a science of security for machine learning systems has proven more elusive. This is due to a number of obstacles, including (1) the highly varied angles of attack against ML systems, (2) the lack of a clearly defined attack surface (because the source of the data analyzed by ML systems is not easily traced), and (3) the lack of formal definitions of security that are appropriate for ML systems. At the same time, security of ML systems is of great importance due to the recent trend of using ML systems as a line of defense against malicious behavior (e.g., network intrusion, malware, and ransomware), as well as the prevalence of ML systems as parts of sensitive and valuable software systems (e.g., sentiment analyzers for predicting stock prices). This workshop will bring together experts from the computer security and machine learning communities in an attempt to highlight recent work in this area, as well as to clarify the foundations of secure ML and chart out important directions for future work and cross-community collaborations.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00 AM</td>
<td>Strategic and Societal Implications of ML</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Security in Practice</td>
</tr>
<tr>
<td>11:15 AM</td>
<td>Security in FinTech</td>
</tr>
<tr>
<td>01:15 PM</td>
<td>Vulnerabilities of Neural Nets</td>
</tr>
<tr>
<td>02:45 PM</td>
<td>Security and Fairness</td>
</tr>
<tr>
<td>03:30 PM</td>
<td>Security of Chat Bots</td>
</tr>
<tr>
<td>04:00 PM</td>
<td>Verification of ML Systems</td>
</tr>
</tbody>
</table>

Learning on Distributions, Functions, Graphs and Groups

Florence d’Alché-Buc, Krikamol Muandet, Bharath Sriperumbudur, Zoltán Szabó

Seaside Ballroom, Fri Dec 08, 08:00 AM

The increased variability of acquired data has recently pushed the field of machine learning to extend its scope to non-standard data including for example functional (Ferraty & Vieu, 2006; Wang et al., 2015), distributional (Póczos et al., 2013), graph, or topological data (Carlsson, 2009; Vitaliy). Successful applications span across a wide range of disciplines such as healthcare (Zhou et al., 2013), action recognition from iPod/iPhone accelerometer data (Sun et al., 2013), causal inference (Lopez-Paz et al., 2015), bioinformatics (Kondor & Pan, 2016; Kusano et al., 2016), cosmology (Ravanbakhsh et al., 2016; Law et al., 2017), acoustic-to-articulatory speech inversion (Kadri et al., 2016), network inference (Brouard et al., 2016), climate research (Szabó et al., 2016), and ecological inference (Fiaxman et al., 2015).

Leveraging the underlying structure of these non-standard data types often leads to significant boost in prediction accuracy and inference performance. In order to achieve these compelling improvements, however, numerous challenges and questions have to be addressed: (i) choosing an adequate representation of the data, (ii) constructing appropriate similarity measures (inner product, norm or metric) on these representations, (iii) efficiently exploiting their intrinsic structure such as multi-scale nature or invariances, (iv) designing affordable computational schemes (relying e.g., on surrogate losses), (v) understanding the computational-statistical tradeoffs of the resulting algorithms, and (vi) exploring novel application domains.

The goal of this workshop is (i) to discuss new theoretical considerations and applications related to learning with non-standard data, (ii) to explore future research directions by bringing together practitioners with various domain expertise and algorithmic tools, and theoreticians interested in providing sound methodology, (iii) to accelerate the advances of this recent area and application arsenal.

We encourage submissions on a variety of topics, including but not limited to:
- Novel applications for learning on non-standard objects
- Learning theory/algorithms on distributions
- Topological and geometric data analysis
- Functional data analysis
- Multi-task learning, structured output prediction, and surrogate losses
- Vector-valued learning (e.g., operator-valued kernel)
- Gaussian processes
- Learning on graphs and networks
- Group theoretic methods and invariances in learning
- Learning with non-standard input/output data
- Large-scale approximations (e.g. sketching, random Fourier features, hashing, Nyström method, inducing point methods), and statistical-computational efficiency tradeoffs

References:


 Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Surname</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30 AM</td>
<td>Cortes</td>
<td>Rational kernels.</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>Lacoste-Julien</td>
<td>On Structured Prediction Theory with Calibrated Convex Surrogate Losses</td>
</tr>
<tr>
<td>09:30 AM</td>
<td>Differently Private Database Release via Kernel Mean Embeddings.</td>
<td></td>
</tr>
<tr>
<td>09:50 AM</td>
<td>Bayesian Distribution Regression.</td>
<td></td>
</tr>
<tr>
<td>10:10 AM</td>
<td>Bayesian Distribution Regression (poster).</td>
<td></td>
</tr>
<tr>
<td>10:10 AM</td>
<td>The Weighted Kendall Kernel (poster).</td>
<td></td>
</tr>
<tr>
<td>10:10 AM</td>
<td>On Kernel Methods for Covariates that are Rankings (poster).</td>
<td></td>
</tr>
<tr>
<td>10:10 AM</td>
<td>When is Network Lasso Accurate: The Vector Case (poster).</td>
<td></td>
</tr>
<tr>
<td>10:10 AM</td>
<td>Algorithim and Statistical Aspects of Linear Regression without Correspondence (poster).</td>
<td></td>
</tr>
<tr>
<td>10:10 AM</td>
<td>Graph based Feature Selection for Structured High Dimensional Data (poster). Zhang</td>
<td></td>
</tr>
<tr>
<td>10:10 AM</td>
<td>Large Scale Graph Learning from Smooth Signals (poster).</td>
<td></td>
</tr>
<tr>
<td>10:10 AM</td>
<td>Post Selection Inference with Maximum Mean Discrepancy (poster).</td>
<td></td>
</tr>
<tr>
<td>10:10 AM</td>
<td>Differently Private Database Release via Kernel Mean Embeddings (poster).</td>
<td></td>
</tr>
<tr>
<td>10:10 AM</td>
<td>Convolutional Layers based on Directed Multi-Graphs (poster). Arodz</td>
<td></td>
</tr>
<tr>
<td>10:10 AM</td>
<td>Kernels on Fuzzy Sets: an Overview (poster). Guevara Diaz</td>
<td></td>
</tr>
<tr>
<td>10:10 AM</td>
<td>The Geometric Block Model (poster). Pal</td>
<td></td>
</tr>
<tr>
<td>10:10 AM</td>
<td>Worst-case vs. Average-case Design for Estimation from Fixed Pairwise Comparisons (poster). Hou</td>
<td></td>
</tr>
</tbody>
</table>
Learning from Conditional Distributions via Dual Embeddings (poster). Song

When is Network Lasso Accurate: The Vector Case. Tran

Worst-case vs. Average-case Design for Estimation from Fixed Pairwise Comparisons.

The Weighted Kendall Kernel.

On Kernel Methods for Covariates that are Rankings.

Lunch Break

Learning on topological and geometrical structures of data. Fukumizu

Operator-valued kernels and their application to functional data analysis. Kadri

Poster Session II & Coffee

Distribution Regression and its Applications. Poczos

TBA. Kondor

Abstracts (28):

Abstract 1: Rational kernels. in Learning on Distributions, Functions, Graphs and Groups, Cortes 08:30 AM

This talk presents an overview of a powerful framework for learning with sequences or distributions over sequences: rational kernels. It will describe the favourable properties of rational kernels, including their efficient computation, their flexibility and graph representations, and highlight their use in a number of applications in text and speech processing and computational biology, including for learning sequence kernels.

Abstract 2: On Structured Prediction Theory with Calibrated Convex Surrogate Losses. in Learning on Distributions, Functions, Graphs and Groups, Lacoste-Julien 09:00 AM

We provide novel theoretical insights on structured prediction in the context of efficient convex surrogate loss minimization with consistency guarantees. For any task loss, we construct a convex surrogate that can be optimized via stochastic gradient descent and we prove tight bounds on the so-called “calibration function” relating the excess surrogate risk to the actual risk. In contrast to prior related work, we carefully monitor the effect of the exponential number of classes in the learning guarantees as well as on the optimization complexity. As an interesting consequence, we formalize the intuition that some task losses make learning harder than others, and that the classical 0-1 loss is ill-suited for general structured prediction.

This (https://arxiv.org/abs/1703.02403) is joint work with Anton Osokin and Francis Bach.

Abstract 3: Differentially Private Database Release via Kernel Mean Embeddings. in Learning on Distributions, Functions, Graphs and Groups, 09:30 AM

Authors: Matej Balog, Ilya Tolstikhin, Bernhard Schölkopf.

Abstract 4: Bayesian Distribution Regression. in Learning on Distributions, Functions, Graphs and Groups, 09:50 AM

Authors: Ho Chung Leon Law, Dougal J. Sutherland, Dino Sejdinovic, Seth Flaxman.

Abstract 5: Bayesian Distribution Regression (poster). in Learning on Distributions, Functions, Graphs and Groups, 10:10 AM

Authors: Ho Chung Leon Law, Dougal J Sutherland, Dino Sejdinovic, Seth Flaxman. Poster session continues at 14:50 - 15:50.

Abstract 6: The Weighted Kendall Kernel (poster). in Learning on Distributions, Functions, Graphs and Groups, 10:10 AM

Authors: Yunlong Jiao, Jean-Philippe Vert. Poster session continues at 14:50 - 15:50.

Abstract 7: On Kernel Methods for Covariates that are Rankings (poster). in Learning on Distributions, Functions, Graphs and Groups, 10:10 AM

Authors: Horia Mania, Aaditya Ramdas, Martin Wainwright, Michael Jordan, Benjamin Recht. Poster session continues at 14:50 - 15:50.

Abstract 8: When is Network Lasso Accurate: The Vector Case (poster). in Learning on Distributions, Functions, Graphs and Groups, 10:10 AM

Authors: Nguyen Quang Tran, Alexander Jung, Saeed Basirian. Poster session continues at 14:50 - 15:50.

Abstract 9: Algorithmic and Statistical Aspects of Linear Regression without Correspondence (poster). in Learning on Distributions, Functions, Graphs and Groups, 10:10 AM

Authors: Daniel Hsu, Kevin Shi, Xiaorui Sun. Poster session continues at 14:50 - 15:50.

Abstract 10: Graph based Feature Selection for Structured High Dimensional Data (poster). in Learning on Distributions, Functions, Graphs and Groups, Zhang 10:10 AM


Abstract 11: Large Scale Graph Learning from Smooth Signals (poster). in Learning on Distributions, Functions, Graphs and Groups, 10:10 AM

Authors: Vassilis Kalofolias, Nathanael Perraudin. Poster session continues at 14:50 - 15:50.

Abstract 12: Learning from Graphs with Structural Variation (poster). in Learning on Distributions, Functions, Graphs and Groups, Holm,
Abstract 13: Post Selection Inference with Maximum Mean Discrepancy (poster). in Learning on Distributions, Functions, Graphs and Groups, 10:10 AM


Abstract 14: Differentially Private Database Release via Kernel Mean Embeddings (poster). in Learning on Distributions, Functions, Graphs and Groups, 10:10 AM

Authors: Matej Balog, Ilya Tolstikhin, Bernhard Schölkopf. Poster session continues at 14:50 - 15:50.

Abstract 15: Convolutional Layers based on Directed Multi-Graphs (poster). in Learning on Distributions, Functions, Graphs and Groups, 10:10 AM

Author: Tomasz Arodz. Poster session continues at 14:50 - 15:50.

Abstract 16: Kernels on Fuzzy Sets: an Overview (poster). in Learning on Distributions, Functions, Graphs and Groups, Diaz 10:10 AM

Author: Jorge Luis Guevara Diaz. Poster session continues at 14:50 - 15:50.

Abstract 17: The Geometric Block Model (poster). in Learning on Distributions, Functions, Graphs and Groups, Pal 10:10 AM


Abstract 18: Worst-case vs. Average-case Design for Estimation from Fixed Pairwise Comparisons (poster). in Learning on Distributions, Functions, Graphs and Groups, 10:10 AM


Abstract 19: Squared Earth Mover’s Distance Loss for Training Deep Neural Networks on Ordered-Classes (poster). in Learning on Distributions, Functions, Graphs and Groups, Hou 10:10 AM

Authors: Le Hou, Chen-Ping Yu, Dimitris Samaras. Poster session continues at 14:50 - 15:50.

Abstract 20: Learning from Conditional Distributions via Dual Embeddings (poster). in Learning on Distributions, Functions, Graphs and Groups, Song 10:10 AM

Authors: Bo Dai, Niao He, Yunpeng Pan, Byron Boots, Le Song. Poster session continues at 14:50 - 15:50.

Abstract 21: When is Network Lasso Accurate: The Vector Case. in Learning on Distributions, Functions, Graphs and Groups, Tran 11:00 AM

Authors: Nguyen Quang Tran, Alexander Jung, Saeed Basirian.

Abstract 22: Worst-case vs. Average-case Design for Estimation from Fixed Pairwise Comparisons. in Learning on Distributions, Functions, Graphs and Groups, 11:20 AM

Authors: Ashwin Pananjady, Cheng Mao, Vidya Muthukumar, Martin Wainwright, Thomas Courtade.

Abstract 23: The Weighted Kendall Kernel. in Learning on Distributions, Functions, Graphs and Groups, 11:40 AM

Authors: Yunlong Jiao, Jean-Philippe Vert.

Abstract 24: On Kernel Methods for Covariates that are Rankings. in Learning on Distributions, Functions, Graphs and Groups, 12:00 PM

Authors: Horia Mania, Aaditya Ramdas, Martin Wainwright, Michael Jordan, Benjamin Recht.

Abstract 25: Learning on topological and geometrical structures of data. in Learning on Distributions, Functions, Graphs and Groups, Fukumizu 01:50 PM

Topological data analysis (TDA) is a recent methodology for extracting topological and geometrical features from complex geometric data structures. Persistent homology, a new mathematical notion proposed by Edelsbrunner (2002), provides a multiscale descriptor for the topology of data, and has been recently applied to a variety of data analysis. In this talk I will introduce a machine learning framework of TDA by combining persistent homology and kernel methods. As an expression of persistent homology, persistence diagrams are widely used to express the lifetimes of generators of homology groups. While they serve as a compact representation of data, it is not straightforward to apply standard data analysis to persistence diagrams, since they consist of a set of points in 2D space expressing the lifetimes. We introduce a method of kernel embedding of the persistence diagrams to obtain their vector representation, which enables one to apply any kernel methods in topological data analysis, and propose a persistence weighted Gaussian kernel as a suitable kernel for vectorization of persistence diagrams. Some theoretical properties including Lipschitz continuity of the embedding are also discussed. I will also present applications to change point detection and time series analysis in the field of material sciences and biochemistry.

Abstract 27: Operator-valued kernels and their application to functional data analysis. in Learning on Distributions, Functions, Graphs and Groups, Kadri 02:20 PM

Positive semidefinite operator-valued kernel generalizes the well-known notion of reproducing kernel, and is a main concept underlying many kernel-based vector-valued learning algorithms. In this talk I will give a brief introduction to learning with operator-valued kernels, discuss current challenges in the field, and describe convenient schemes to overcome them. I will overview our recent work on learning with functional data in the case where both attributes and labels are functions. In this setting, a set of rigorously defined infinite-dimensional operator-valued kernels that can be valuably applied when the data are functions is described, and a learning scheme for nonlinear functional data analysis is introduced. The methodology is illustrated through speech and audio signal processing experiments.
Abstract 29: Distribution Regression and its Applications. in Learning on Distributions, Functions, Graphs and Groups, Poczos
03:50 PM

The most common machine learning algorithms operate on finite-dimensional vectorial feature representations. In many applications, however, the natural representation of the data consists of distributions, sets, and other complex objects rather than finite-dimensional vectors. In this talk we will review machine learning algorithms that can operate directly on these complex objects. We will discuss applications in various scientific problems including estimating the cosmological parameters of our Universe, dynamical mass measurements of galaxy clusters, finding anomalous events in fluid dynamics, and estimating phenotypes in agriculturally important plants.

Abstract 30: TBA. in Learning on Distributions, Functions, Graphs and Groups, Kondor 04:20 PM

TBA