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

Schedule Highlights

Dec. 8, 2017

101 A, **Learning in the Presence of Strategic Behavior** *Haghtalab, Mansour, Roughgarden, Syrgkanis, Wortman Vaughan*

101 B, **Visually grounded interaction and language** *Strub, de Vries, Das, Kottur, Lee, Malinowski, Pietquin, Parikh, Batra, Courville, Mary*

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

102 C, **Synergies in Geometric Data Analysis (TWO DAYS)** *Meila, Chazal, Chen*

102 C, **6th Workshop on Automated Knowledge Base Construction (AKBC)** *Pujara, Chen, Dalvi Mishra, Rocktäschel*

103 A+B, **Competition track** *Escalera, Weimer*

104 A, **Machine Learning for Health (ML4H) - What Parts of Healthcare are Ripe for Disruption by Machine Learning Right Now?** *Beam, Beam, Fiterau, Fiterau, Schulam, Schulam, Fries, Fries, Hughes, Hughes, Wiltschko, Wiltschko, Snoek, Snoek, Antropova, Antropova, Ranganath, Ranganath, Jedynak, Jedynak, Naumann, Naumann, Dalca, Dalca, Dalca, Dalca, Althoff, Althoff, ASTHANA, ASTHANA, Tandon, Tandon, Kandola, Kandola, Ratner, Ratner, Kale, Kale, Shalit, Shalit, Ghassemi, Ghassemi, Kohane, Kohane*

104 B, **Acting and Interacting in the Real World: Challenges in Robot Learning** *Posner, Hadsell, Riedmiller, Wulfmeier, Paul*

104 C, **Deep Learning for Physical Sciences** *Baydin, Prabhat, Cranmer, Wood*

201 A, **Machine Learning for Audio Signal Processing (ML4Audio)** *Purwins, Sturm, Plumbley*

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

202, **Machine Deception** *Goodfellow, Hwang, Goodman, Rodriguez*

203, **Discrete Structures in Machine Learning** *Singer, Bilmes, Krause, Jegelka, Karbasi*

204, **Transparent and interpretable Machine Learning in Safety Critical Environments** *Tosi, Vellido, Álvarez*

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

Grand Ballroom B, **Conversational AI - today's practice and tomorrow's potential** *Geramifard, Williams, Heck, Glass, Bordes, Young, Tesauro*

Hall A, **OPT 2017: Optimization for Machine Learning** *Sra, J. Reddi, Agarwal, Recht*

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

Hyatt Hotel, Regency Ballroom A+B+C, **Extreme Classification: Multi-class & Multi-label Learning in Extremely Large Label Spaces** *Varma, Kloft, Dembczynski*

Hyatt Hotel, Regency Ballroom D+E+F+H, **Learning on Distributions, Functions, Graphs and Groups** *d'Alché-Buc, Muandet, Sriperumbudur, Szabó*

Hyatt Hotel, Seaview Ballroom, **Machine Learning for Creativity and Design** *Eck, Ha, Eslami, Dieleman, Fiebrink, Elliott*

Hyatt Hotel, Shoreline, **Machine Learning and Computer Security** *Steinhardt, Papernot, Li, Liu, Liang, Song*

S1, **ML Systems Workshop @ NIPS 2017** *Lakshmiratan, Bird, Sen, Ré, Li, Gonzalez, Crankshaw*

S4, **Machine Learning for Molecules and Materials** *Chmiela, Hernández-Lobato, Schütt, Aspuru-Guzik, Tkatchenko, Ramsundar, von Lilienfeld, Kusner, Tsuda, Paige, Müller*

S5, **Workshop on Worm's Neural Information Processing (WNIP)** *Hasani, Zimmer, Larson, Grosu*

S7, **Machine Learning for the Developing World** *De-Arteaga, Herlands*

Seaside Ballroom, **Advances in Approximate Bayesian Inference** *Ruiz, Mandt, Zhang, McInerney, Tran, Broderick, Titsias, Blei, Welling*

Dec. 9, 2017

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

101 B, **Deep Learning at Supercomputer Scale** *Elsen, Hafner, Stone, Saeta*

102 A+B, **Machine Learning on the Phone and other Consumer Devices** *Aradhye, Quinonero Candela, Prasad*

102 C, **Synergies in Geometric Data Analysis (2nd day)** *Meila, Chazal, Chen*

103 A+B, **Medical Imaging meets NIPS** *Glocker, Konukoglu, Lombaert, Bhatia*

103 C, **Workshop on Prioritising Online Content** *Shawe-Taylor, Pontil, Cesa-Bianchi, Yilmaz, Watkins, Riedel, Grobelnik*

104 A, **Cognitively Informed Artificial Intelligence: Insights From Natural Intelligence** *Mozer, Lake, Yu*

104 B, **Machine Learning in Computational Biology** *Zou, Kundaje, Quon, Fusi, Mostafavi*

104 C, **The future of gradient-based machine learning software & techniques** *Wiltschko, van Merriënboer, Lamblin*

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

201 B, **Aligned Artificial Intelligence** *Hadfield-Menell, Steinhardt, Duvenaud, Krueger, Dragan*

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

203, **Learning Disentangled Features: from Perception to Control** *Denton, Narayanaswamy, Kulkarni, Lee, Bouchacourt, Tenenbaum, Pfau*

204, **BigNeuro 2017: Analyzing brain data from nano to macroscale** *Dyer, Kiar, Gray Roncal, , Koerding, Vogelstein*

Grand Ballroom A, **Hierarchical Reinforcement Learning** *Barto, Precup, Mannor, Schaul, Fox, Florensa*

Grand Ballroom B, **Learning with Limited Labeled Data: Weak Supervision and Beyond** *Augenstein, Bach, Belilovsky, Blaschko,* *Lampert, Oyallon, Platanios, Ratner, Ré*

Hall A, **Deep Learning: Bridging Theory and Practice** *Arora, Raghu, Salakhutdinov, Schmidt, Vinyals*

Hall C, **Bayesian Deep Learning** *Gal, Hernández-Lobato, Louizos, Wilson, Kingma, Ghahramani, Murphy, Welling*

Hyatt Beacon Ballroom D+E+F+H, **Workshop on Meta-Learning** *Calandra, Hutter, Larochelle, Levine*

Hyatt Hotel, Regency Ballroom A+B+C, **Interpreting, Explaining and Visualizing Deep Learning - Now what ?** *Müller, Vedaldi, Hansen, Samek, Montavon*

Hyatt Hotel, Seaview Ballroom, **Optimal Transport and Machine Learning** *Bousquet, Cuturi, Peyré, Sha, Solomon*

Hyatt Hotel, Shoreline, **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*

S1, **Machine Learning Challenges as a Research Tool** *Guyon, Viegas, Escalera, Abernethy*

S4, **Emergent Communication Workshop** *Foerster, Mordatch, Lazaridou, Cho, Kiela, Abbeel*

S7, **Bayesian optimization for science and engineering** *Martinez-Cantin, Hernández-Lobato, Gonzalez*

Seaside Ballroom, **Teaching Machines, Robots, and Humans** *Cakmak, Rafferty, Singla, Zhu, Zilles*

Dec. 8, 2017

Learning in the Presence of Strategic Behavior

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

101 A, 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 it once learned about. 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.

Schedule

09:00 AM	**(Invited Talk) Yiling Chen: Learning in Strategic Data Environments.**	*Chen*
09:45 AM	**Strategic Classification from Revealed Preferences**	*Dong*
10:00 AM	**Learning in Repeated Auctions with Budgets: Regret Minimization and Equilibrium**	*Gur*
10:15 AM	**Spotlights**	*Podimata, Zuo, Feng, Kim*
11:00 AM	**(Invited Talk) Eva Tardos: Online learning with partial information for players in games.**	*Tardos*
11:45 AM	**(Invited Talk) Mehryar Mohri: Regret minimization against strategic buyers.**	*Mohri*
12:30 PM	**Lunch Break**	
01:50 PM	**(Invited Talk) Percy Liang: Learning with Adversaries and Collaborators**	*Liang*
02:35 PM	**Spotlights**	*Kangasrääsiö, Everett, Liang, Cai, Wu, Muthukumar, Schmit*
03:00 PM	**Poster Session & Coffee break**	
03:30 PM	**(Invited Talk) Alex Peysakhovich: Towards cooperative AI**	*Peysakhovich*
04:15 PM	**Statistical Tests of Incentive Compatibility in Display Ad Auctions**	*Munoz*

04:30 PM	**Optimal Economic Design through Deep Learning**	*Parkes*
04:45 PM	**Learning Against Non-Stationary Agents with Opponent Modeling & Deep Reinforcement Learning**	*Everett*

Abstracts (12):

Abstract 1: **(Invited Talk) Yiling Chen: Learning in Strategic Data Environments. in Learning in the Presence of Strategic Behavior**, *Chen* 09:00 AM

We live in a world where activities and interactions are recorded as data: food consumption, workout activities, buying and selling products, sharing information and experiences, borrowing and lending money, and exchanging excess resources. Scientists use the rich data of these activities to understand human social behavior, generate accurate predictions, and make policy recommendations. Machine learning traditionally take such data as given, often treating them as independent samples from some unknown statistical distribution. However, such data are possessed or generated by potentially strategic people in the context of specific interaction rules. Hence, what data become available depends on the interaction rules. For example, people with sensitive medical conditions may not reveal their medical data in a survey but could be willing to share them when compensated; crowd workers may not put in a good-faith effort in completing a task if they know that the requester cannot verify the quality of their contributions. In this talk, I argue that a holistic view that jointly considers data acquisition and learning is important. I will discuss two projects. The first project considers acquiring data from strategic data holders who have private cost for revealing their data and then learning from the acquired data. We provide a risk bound on learning, analogous to classic risk bounds, for situations when agents' private costs can correlate with their data in arbitrary ways. The second project leverages techniques in learning to design a mechanism for obtaining high-quality data from strategic data holders. The mechanism has a strong incentive property: it is a dominant strategy for each agent to truthfully reveal their data even if we have no ground truth to directly evaluate their contributions.

This talk is based on joint works with Jacob Abernethy, Chien-Ju Ho, Yang Liu, and Bo Waggoner.

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, Zuo, Feng, 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.

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

4. Budget Management Strategies in Repeated Auctions. Uniqueness and Stability of System Equilibria. Santiago Balseiro, Anthony Kim, Mohammad Mahdian and Vahab Mirrokni.

5. Dynamic Second Price Auctions with Low Regret. Vahab Mirrokni, Renato Paes Leme, Rita Ren and Song Zuo.

Abstract 5: **(Invited Talk) Eva Tardos: Online learning with partial information for players in games. in Learning in the Presence of Strategic Behavior**, *Tardos* 11:00 AM

Learning has been adopted as a general behavioral model for players in repeated games. Learning offers a way that players can adopt to (possibly changing) environment. Learning guarantees high social welfare in many games (including traffic routing as well as online auctions), even when the game or the population of players is dynamically changing. The rate at which the game can change depends on the speed of convergence of the learning algorithm. If players observe all other participants, which such full information feedback classical learning algorithms offer very fast convergence. However, such full information feedback is often not available, and the convergence of classical algorithms with partial feedback is much good. In this talk we develop a black-box approach for learning where the learner observes as feedback only losses of a subset of the actions. The simplicity and black box nature of the approach allows us to use of this faster learning rate as a behavioral assumption in games. Talk based on joint work with Thodoris Lykouris and Karthik Sridharan.

Abstract 6: **(Invited Talk) Mehryar Mohri: Regret minimization against strategic buyers. in Learning in the Presence of Strategic Behavior**, *Mohri* 11:45 AM

This talk presents an overview of several recent algorithms for regret minimization against strategic buyers in the context of posted-price auctions, which are crucial for revenue optimization in online advertisement.

Joint work with Andres Munoz Medina.

Abstract 8: **(Invited Talk) Percy Liang: Learning with Adversaries and Collaborators in Learning in the Presence of Strategic Behavior**, *Liang* 01:50 PM

We argue that the standard machine learning paradigm is both too weak and too string. First, we show that current systems for image classification and reading comprehension are vulnerable to adversarial attacks, suggesting that existing learning setups are inadequate to produce systems with robust behavior. Second, we show that in an interactive learning setting where incentives are aligned, a system can learn a simple natural language from a user from scratch, suggesting that much more can be learned under a cooperative setting.

Abstract 9: **Spotlights in Learning in the Presence of Strategic Behavior**, *Kangasrääsiö, Everett, Liang, Cai, Wu, Muthukumar, Schmit* 02:35 PM

1. Multi-Agent Counterfactual Regret Minimization for Partial-Information Collaborative Games. Matthew Hartley, Stephan Zheng and Yisong Yue.

2. Inference of Strategic Behavior based on Incomplete Observation Data. Antti Kangasrääsiö and Samuel Kaski.

3. Inferring agent objectives at different scales of a complex adaptive system. Dieter Hendricks, Adam Cobb, Richard Everett, Jonathan Downing and Stephen Roberts.

4. A Reinforcement Learning Framework for Eliciting High Quality Information. Zehong Hu, Yang Liu, Yitao Liang and Jie Zhang.

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

6. Competing Bandits: Learning under Competition. Yishay Mansour, Aleksandrs Slivkins and Zhiwei Steven Wu.

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

8. Learning with Abandonment. Ramesh Johari and Sven Schmit.

Abstract 11: **(Invited Talk) Alex Peysakhovich: Towards cooperative AI in Learning in the Presence of Strategic Behavior**, *Peysakhovich* 03:30 PM

Social dilemmas are situations where individuals face a temptation to increase their payoffs at a cost to total welfare. Importantly, social dilemmas are ubiquitous in real world interactions. We show how to modify modern reinforcement learning methods to construct agents that act in ways that are simple to understand, begin by cooperating, try to avoid being exploited, and forgiving (try to return to mutual cooperation). Such agents can maintain cooperation in Markov social dilemmas with both perfect and imperfect information. Our construction does not require training methods beyond a modification of self-play, thus if an environment is such that good strategies can be constructed in the zero-sum case (eg. Atari) then we can construct agents that solve social dilemmas in this environment.

Abstract 12: **Statistical Tests of Incentive Compatibility in Display Ad Auctions in Learning in the Presence of Strategic Behavior**, *Munoz* 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 user 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**, *Everett* 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

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

101 B, 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 - Univeristy of Berkeley
Chen Yu - University of Indiana

Schedule

08:30 AM	**Welcome!**	
08:45 AM	**Visually Grounded Language: Past, Present, and Future...**	*Mooney*
09:30 AM	**Connecting high-level semantics with low-level vision**	*Fidler*
10:15 AM	**Break + Poster (1)**	*Chaplot, MA, Brodeur, Matsuo, Kobayashi, Shinagawa, Yoshino, Guo, Murdoch, Mysore Sathyendra, Ricks, Zhang, Peterson, Zhang, Mironenco, Anderson, Johnson, Yoo, Barzdins, Zaidi, Andrews, Witteveen, OOTA, Vijayaraghavan, Wang, Zhu, Liepins, Quinn, Raj, Cartillier, Chu, Caballero, Obermeyer*
10:40 AM	**The interface between vision and language in the human brain?**	*Gallant*
11:25 AM	**Towards Embodied Question Answering**	*Parikh*
02:00 PM	**Dialogue systems and RL: interconnecting language, vision and rewards**	*Pietquin*
02:45 PM	**Accepted Papers**	
03:15 PM	**Break + Poster (2)**	
03:40 PM	**Grounded Language Learning in a Simulated 3D World**	*Hill*
04:25 PM	**How infant learn to speak by interacting with the visual world?**	*Yu*
05:10 PM	**Panel Discussion**	*Hill, Pietquin, Gallant, Mooney, Fidler, Yu, Parikh*

Advances in Modeling and Learning Interactions from Complex Data

Gautam Dasarathy, Mladen Kolar, Richard Baraniuk

102 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

08:00 AM	**Opening Remarks**	*Dasarathy*
08:05 AM	**Community Detection and Invariance to Distribution**	*Bresler*
08:40 AM	**Edge Exchangeable Temporal Network Models**	*Ng*

09:00 AM	**A Data-Driven Sparse-Learning Approach to Model Reduction in Chemical Reaction Networks**	*HARIRCHI*
09:20 AM	**Poster Spotlights**	
10:00 AM	**Coffee Break + Posters**	
11:00 AM	**Your dreams may come true with MTP2**	*Uhler*
11:35 AM	**Estimating Mixed Memberships with Sharp Eigenvector Deviations**	*Mao*
12:00 PM	**Lunch Break**	
02:00 PM	**Recovering Latent Causal Relations from Times Series Data**	*Kiyavash*
02:35 PM	**Learning High-Dimensional DAGs: Provable Statistical Guarantees and Scalable Approximation**	
04:00 PM	**Conditional Densities and Efficient Models in Infinite Exponential Families**	*Gretton*
04:35 PM	**The Expxorcist: Nonparametric Graphical Models Via Conditional Exponential Densities**	*Suggala*
04:55 PM	**Mathematical and Computational challenges in Reconstructing Evolution**	*Warnow*

Abstracts (9):

Abstract 2: **Community Detection and Invariance to Distribution in Advances in Modeling and Learning Interactions from Complex Data**, *Bresler* 08:05 AM

We consider the problem of recovering a hidden community of size K from a graph where edges between members of the community have label X drawn i.i.d. according to P and all other edges have labels drawn i.i.d. according to Q. The information limits for this problem were characterized by Hajek-Wu-Xu in 2016 in terms of the KL-divergence between P and Q. We complement their work by showing that for a broad class of distributions P and Q the computational difficulty is also determined by the KL-divergence. We additionally show how to reduce general P and Q to the case P = Ber(p) and Q = Ber(q) and vice versa, giving a direct computational equivalence (up to polynomial time).

Abstract 3: **Edge Exchangeable Temporal Network Models in Advances in Modeling and Learning Interactions from Complex Data**, *Ng* 08:40 AM

We propose a dynamic edge exchangeable network model that can capture sparse connections observed in real temporal networks, in contrast to existing models which are dense. The model achieved superior link prediction accuracy when compared to a dynamic variant of the blockmodel, and is able to extract interpretable time-varying community structures. In addition to sparsity, the model accounts for the effect of social influence on vertices' future behaviours. Compared to the dynamic blockmodels, our model has a smaller latent space. The compact latent space requires a smaller number of parameters to be estimated in variational inference and results in a computationally friendly inference algorithm.

Abstract 4: **A Data-Driven Sparse-Learning Approach to Model Reduction in Chemical Reaction Networks in Advances in Modeling and Learning Interactions from Complex Data**, *HARIRCHI* 09:00 AM

In this paper, we propose an optimization-based sparse learning approach to iden- tify the set of most influential reactions in a chemical reaction network. This reduced set of reactions is then employed to construct a reduced chemical reaction mechanism, which is relevant to chemical interaction network modeling. The problem of identifying influential reactions is first formulated as a mixed-integer quadratic program, and then a relaxation method is leveraged to reduce the compu- tational complexity of our approach. Qualitative and quantitative validation of the sparse encoding approach demonstrates that the model captures important network structural properties with moderate computational load.

Abstract 7: **Your dreams may come true with MTP2 in Advances in Modeling and Learning Interactions from Complex Data**, *Uhler* 11:00 AM

We study maximum likelihood estimation for exponential families that are multivariate totally positive of order two (MTP2). Such distributions appear in the context of ferromagnetism in the Ising model and various latent models, as for example Brownian motion tree models used in phylogenetics. We show that maximum likelihood estimation for MTP2 exponential families is a convex optimization problem. For quadratic exponential families such as Ising models and Gaussian graphical models, we show that MTP2 implies sparsity of the underlying graph without the need of a tuning parameter. In addition, we characterize a subgraph and a supergraph of Gaussian graphical models under MTP2. Moreover, we show that the MLE always exists even in the high-dimensional setting. These properties make MTP2 constraints an intriguing alternative to methods for learning sparse graphical models such as the graphical lasso.

Abstract 8: **Estimating Mixed Memberships with Sharp Eigenvector Deviations in Advances in Modeling and Learning Interactions from Complex Data**, *Mao* 11:35 AM

Real world networks often have nodes belonging to multiple communities. We consider the detection of overlapping communities under the popular Mixed Membership Stochastic Blockmodel (MMSB). Using the inherent geometry of this model, we link the inference of overlapping communities to the problem of finding corners in a noisy rotated and scaled simplex, for which consistent algorithms exist. We use this as a building block for our algorithm to infer the community memberships of each node, and prove its consistency. As a byproduct of our analysis, we derive sharp row-wise eigenvector deviation bounds, and provide a cleaning step that improves the performance drastically for sparse networks. We also propose both necessary and sufficient conditions for identifiability of the model, while existing methods typically present sufficient conditions. The empirical performance of our method is shown using simulated and real datasets scaling up to 100,000 nodes.

Abstract 10: **Recovering Latent Causal Relations from Times Series Data in Advances in Modeling and Learning Interactions from Complex Data**, *Kiyavash* 02:00 PM

Discovering causal relationships from data is a challenging problem that is exacerbated when some of the variables of interests are latent. In this talk, we discuss the problem of learning the support of transition matrix between random processes in a Vector Autoregressive (VAR) model from samples when a subset of the processes are latent. It is well known that ignoring the effect of the latent processes may lead to very different estimates of the influences even among observed processes. We are not only interested in identifying the influences among the observed processes, but also aim at learning those between the latent ones, and those from the latent to the observed ones. We show that the support of transition matrix among the observed processes and lengths of all latent paths between any two observed processes can be identified successfully under some conditions on the VAR model. Our results apply to both non-Gaussian and Gaussian cases, and experimental results on various synthetic and real-world datasets validate our theoretical findings.

Abstract 12: **Conditional Densities and Efficient Models in Infinite Exponential Families in Advances in Modeling and Learning Interactions from Complex Data**, *Gretton* 04:00 PM

The exponential family is one of the most powerful and widely used classes of models in statistics. A method was recently developed to fit this model when the natural parameter and sufficient statistic are infinite dimensional, using a score matching approach. The infinite exponential family is a natural generalisation of the finite case, much like the Gaussian and Dirichlet processes generalise their respective finite modfels.
In this talk, I'll describe two recent results which make this model more applicable in practice, by reducing the computational burden and improving performance for high-dimensional data. The firsrt is a Nytsrom-like approximation to the full solution. We prove that this approximate solution has the same consistency and convergence rates as the full-rank solution (exactly in Fisher distance, and nearly in other distances), with guarantees on the degree of cost and storage reduction. The second result is a generalisation of the model family to the conditional case, again with consistency guarantees. In experiments, the conditional model generally outperforms a competing approach with consistency guarantees, and is competitive with a deep conditional density model on datasets that exhibit abrupt transitions and heteroscedasticity.

Abstract 13: **The Expxorcist: Nonparametric Graphical Models Via Conditional Exponential Densities in Advances in Modeling and Learning Interactions from Complex Data**, *Suggala* 04:35 PM

Non-parametric multivariate density estimation faces strong statistical and computational bottlenecks, and the more practical approaches impose near-parametric assumptions on the form of the density functions. In this paper, we leverage recent developments to propose a class of non-parametric models which have very attractive computational and statistical properties.

Abstract 14: **Mathematical and Computational challenges in Reconstructing Evolution in Advances in Modeling and Learning Interactions from Complex Data**, *Warnow* 04:55 PM

Reconstructing evolutionary histories is a basic step in much biological discovery, as well as in historical linguistics and other domains. Inference methods based on mathematical models of evolution have been used to make substantial advances, including in understanding the early origins of life, to predicting protein structures and functions, and to addressing questions such as "Where did the Indo-European languages begin?" In this talk, I will describe the current state of the art in phylogeny estimation in these domains, what is understood from a mathematical perspective, and identify fascinating open problems where novel mathematical research - drawing from graph theory, algorithms, and probability theory - is needed. This talk will be accessible to mathematicians, computer scientists, and probabilists, and does not require any knowledge in biology.

Synergies in Geometric Data Analysis (TWO DAYS)

Marina Meila, Frederic Chazal, yuchaz Chen

102 C, 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 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 expliecitly 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. Trough the tutorials, we are reaching out to the wider NIPS audience, to the many potential users of 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

08:10 AM	**Supervised learning of labeled pointcloud differences via cover-tree entropy reduction**	*Harer*
09:10 AM	**Estimating the Reach of a Manifold**	*Aamari*
09:40 AM	**Multiscale geometric feature extraction**	*Polonik*
10:10 AM	**Poster spotlights**	
10:15 AM	**Maximum likelihood estimation of Riemannian metrics from Euclidean data**	*Arvanitidis*
10:15 AM	**A dual framework for low rank tensor completion**	*Nimishakavi*
10:15 AM	**Parallel multi-scale reduction of persistent homology**	*Mendoza Smith*
10:30 AM	**Coffee break**	
11:00 AM	**Persistent homology of KDE filtration of Rips complexes**	*Shin, Rinaldo*
11:30 AM	**Characterizing non-linear dimensionality reduction methods using Laplacian-like operators**	*Ting*
01:30 PM	**Poster session I**	
02:00 PM	**Multiscale characterization of molecular dynamics**	*Clementi*
03:30 PM	**Functional Data Analysis using a Topological Summary Statistic: the Smooth Euler Characteristic Transform,**	*Crawford*
04:00 PM	**Consistent manifold representation for TDA**	*Sauer*
05:00 PM	**Discussion: Geometric Data Analysis**	*Chazal, Meila*
08:30 AM	**Topological Data Analisys with GUDHI and scalable manifold learning and clustering with megaman**	*Rouvreau, Meila*
08:50 AM	**Introduction to the R package TDA**	*KIM*
09:20 AM	**Riemannian metric estimation and the problem of isometric embedding**	*Joncas*
09:50 AM	**Ordinal distance comparisons: from topology to geometry**	*von Luxburg*
10:20 AM	**Cofee break**	
10:50 AM	**Geometric Data Analysis software**	
11:30 AM	**Poster session II**	
02:00 PM	**Modal-sets, and density-based Clustering**	*Kpotufe*
02:30 PM	**A Note on Community Trees in Networks**	*Chen*
03:30 PM	**Beyond Two-sample-tests: Localizing Data Discrepancies in High-dimensional Spaces**	*Cazals*

Abstracts (15):

Abstract 1: **Supervised learning of labeled pointcloud differences via cover-tree entropy reduction in Synergies in Geometric Data Analysis (TWO DAYS)**, *Harer* 08:10 AM

We introduce a new algorithm, called CDER, for supervised machine learning that merges the multi-scale geometric properties of Cover Trees with the information-theoretic properties of entropy. CDER applies to a training set of labeled pointclouds embedded in a common Euclidean space. If typical pointclouds corresponding to distinct labels tend to differ at any scale in any sub-region, CDER can identify these differences in linear time, creating a set of distributional coordinates which act as a feature extraction mechanism for supervised learning. We describe the use of CDER both directly on point clouds and on persistence diagrams.

Abstract 2: **Estimating the Reach of a Manifold in Synergies in Geometric Data Analysis (TWO DAYS)**, *Aamari* 09:10 AM

Various problems in manifold estimation make use of a quantity called the reach, denoted by τM, which is a measure of the regularity of the manifold. This paper is the first investigation into the problem of how to estimate the reach. First, we study the geometry of the reach through an approximation perspective. We derive new geometric results on the reach for submanifolds without boundary. An estimator ˆτ of τM is proposed in a framework where tangent spaces are known, and bounds assessing its efficiency are derived. In the case of i.i.d. random point cloud X n , τˆ (X n) is showed to achieve uniform expected loss bounds over a C3-like model. Finally, we obtain upper and lower bounds on the minimax rate for estimating the reach.

Abstract 4: **Poster spotlights in Synergies in Geometric Data Analysis (TWO DAYS)**, 10:10 AM

Each poster presenter will have approximatively 1.5 minutes to advertise their posters.
We encourage all poster presenters to put up their posters at the beginning of the workshop.

Abstract 6: **A dual framework for low rank tensor completion in Synergies in Geometric Data Analysis (TWO DAYS)**, *Nimishakavi* 10:15 AM

We propose a novel formulation of the low-rank tensor completion problem that is based on the duality theory and a particular choice of low-rank regularizer. This low-rank regularizer along with the dual perspective provides a simple characterization of the solution to the tensor completion problem. Motivated by large-scale setting, we next derive a rank-constrained reformulation of the proposed optimization problem, which is shown to lie on the Riemannian spectrahedron manifold. We exploit the versatile Riemannian optimization framework to develop computationally efficient conjugate gradient and trust-region algorithms. The experiments confirm the benefits of our choice of regularization and the proposed algorithms outperform state-of-the-art algorithms on several real-world data sets in different applications.

Abstract 7: **Parallel multi-scale reduction of persistent homology in Synergies in Geometric Data Analysis (TWO DAYS)**, *Mendoza Smith* 10:15 AM

Persistent homology is a mathematical formalism based on algebraic topology and is central to Topological Data Analysis (TDA). Its paradigm consists in estimating the topological features of a shape embedded in an Euclidean space from a point-cloud sampled from it. The estimation is done at multiple scales by reducing a, so-called, boundary matrix which is a sparse binary matrix that encodes a simplicial complex filtration built from the point-cloud. The reduction process is similar to Gaussian elimination and represents an important computational bottleneck in the pipeline. To improve the scalability of the TDA framework, several strategies to accelerate it have been proposed. Herein, we present a number of structural dependencies in boundary matrices and use them to design a novel parallel reduction algorithm. In particular, we show that this structural information: (i) makes part of the reduction immediately apparent, (ii) decreases the total number of column operations required for reduction, (iii) gives a framework for which the process can be massively parallelised. Simulations on synthetic examples show that the computational burden can be conducted in a small fraction of the number of iterations needed by traditional methods. Moreover, whereas the traditional methods reveal barcodes sequentially from a filtration order, this 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 9: **Persistent homology of KDE filtration of Rips complexes in Synergies in Geometric Data Analysis (TWO DAYS)**, *Shin, Rinaldo* 11:00 AM

When we observe a point cloud in the Euclidean space, the persistent homology of the upper level sets filtration of the density is one of the most important tools to understand topological features of the data generating distribution. The persistent homology of KDEs (kernel density estimators) for the density function is a natural way to estimate the target quantity. In practice, however, calculating the persistent homology of KDEs on d-dimensional Euclidean spaces requires to approximate the ambient space to a grid, which could be computationally inefficient when the dimension of the ambient space is high or topological features are in different scales. In this abstract, we consider the persistent homologies of KDE filtrations on Rips complexes as alternative estimators. We show consistency results for both the persistent homology of the upper level sets filtration of the density and its simplified version. We also describe a novel methodology to construct an asymptotic confidence set based on the bootstrap procedure. Unlike existing procedures, our method does not heavily rely on grid-approximations, scales to higher dimensions, and is adaptive to heterogeneous topological features.

Abstract 10: **Characterizing non-linear dimensionality reduction methods using Laplacian-like operators in Synergies in Geometric Data Analysis (TWO DAYS)**, *Ting* 11:30 AM

(note: the talk is 30 mins, but the server has problems with 12:00 noon)

We examine a number of non-linear dimensionality reduction techniques including Laplacian Eigenmaps, LLE, MVU, HLLE, LTSA, and t-SNE. In each case we show that the non-linear embedding can be characterized by a Laplacian or Laplacian-like operator. By comparing the resulting operators, one can uncover the similarities and differences between the methods. For example, HLLE and LTSA can be shown to be asymptotically identical, and whilst maximum variance unfolding (MVU) can be shown to generate a Laplacian, the behavior of the Laplacian is completely different from that generated by Laplacian Eigenmaps. We discuss the implications of this characterization for generating new non-linear dimensionality reduction methods and smoothness penalties.

Abstract 13: **Functional Data Analysis using a Topological Summary Statistic: the Smooth Euler Characteristic Transform, in Synergies in Geometric Data Analysis (TWO DAYS)**, *Crawford* 03:30 PM

Lorin Crawford1,2,3,†, Anthea Monod4,†, Andrew X. Chen4, Sayan Mukherjee5,6,7,8, and Rau ■l Rabad ■an4

Abstract 15: **Discussion: Geometric Data Analysis in Synergies in Geometric Data Analysis (TWO DAYS)**, *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 16: **Topological Data Analisys with GUDHI and scalable manifold learning and clustering with megaman in Synergies in Geometric Data Analysis (TWO DAYS)**, *Rouvreau, Meila* 08:30 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: **Introduction to the R package TDA in Synergies in Geometric Data Analysis (TWO DAYS)**, *KIM* 08:50 AM

We present a short tutorial and introduction to using the R package TDA, which provides some tools for Topological Data Analysis. In particular, it includes implementations of functions that, given some data, provide topological information about the underlying space, such as the distance function, the distance to a measure, the kNN density estimator, the kernel density estimator, and the kernel distance. The salient topological features of the sublevel sets (or superlevel sets) of these functions can be quantified with persistent homology. We provide an R interface for the efficient algorithms of the C++ libraries GUDHI, Dionysus, and PHAT, including a function for the persistent homology of the Rips filtration, and one for the persistent homology of sublevel sets (or superlevel sets) of arbitrary functions evaluated over a grid of points. The significance of the features in the resulting persistence diagrams can be analyzed with functions that implement the methods discussed in Fasy, Lecci, Rinaldo, Wasserman, Balakrishnan, and Singh (2014), Chazal, Fasy, Lecci, Rinaldo, and Wasserman (2014c) and Chazal, Fasy, Lecci, Michel, Rinaldo, and Wasserman (2014a). The R package TDA also includes the implementation of an algorithm for density clustering, which allows us to identify the spatial organization of the probability mass associated to a density function and visualize it by means of a dendrogram, the cluster tree.

Abstract 21: **Geometric Data Analysis software in Synergies in Geometric Data Analysis (TWO DAYS)**, 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]

Abstract 23: **Modal-sets, and density-based Clustering in Synergies in Geometric Data Analysis (TWO DAYS)**, *Kpotufe* 02:00 PM

Modes or Modal-sets are points or regions of space where the underlying data density is locally-maximal. They are relevant in problems such as clustering, outlier detection, or can simply serve to identify salient structures in high-dimensional data (e.g. point-cloud data from medical imaging, astronomy, etc).

In this talk we will argue that modal-sets, as general extremal surfaces, yield more stable clustering than usual modes (extremal points) of a density. For one, modal-sets can be consistently estimated, at non-trivial convergence rates, despite the added complexity of unknown surface-shape and dimension. Furthermore, modal-sets neatly dovetail into existing approaches that cluster data around point-modes, yielding competitive, yet more stable clustering on a range of real-world problems.

Abstract 24: **A Note on Community Trees in Networks in Synergies in Geometric Data Analysis (TWO DAYS)**, *Chen* 02:30 PM

We introduce the concept of community trees that summarizes topological structures within a network. A community tree is a tree structure representing clique communities from the clique percolation method (CPM). The community tree also generates a persistent diagram. Community trees and persistent diagrams reveal topological structures of the underlying networks and can be used as visualization tools. We study the stability of community trees and derive a quantity called the total star number (TSN) that presents an upper bound on the change of community trees. Our findings provide a topological interpretation for the stability of communities generated by the CPM.

Abstract 25: **Beyond Two-sample-tests: Localizing Data Discrepancies in High-dimensional Spaces in Synergies in Geometric Data Analysis (TWO DAYS)**, *Cazals* 03:30 PM

https://hal.inria.fr/hal-01159235/document

6th Workshop on Automated Knowledge Base Construction (AKBC)

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

102 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 foundations for search and discovery applications. A similarly abundant set of knowledge systems have been developed at top universities such as Stanford (DeepDive), Carnegie Mellon (NELL), the University of Washington (OpenIE), the University of Mannheim (DBpedia), and the Max Planck Institut Informatik (YAGO, WebChild), among others. Our workshop serves as a forum for researchers working on knowledge base construction in both academia and industry. With this year’s workshop we would like to continue the successful tradition of the previous five AKBC workshops. AKBC fills a unique need in the field, bringing together industry leaders and academic researchers. Our workshop is focused on stellar invited talks from high-profile speakers who identify the pressing research areas where current methods fall short and propose visionary approaches that will lead to the next generation of knowledge bases. Our workshop prioritizes a participatory environment where attendees help identify the most promising research, contribute to surveys on 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

09:00 AM	**Challenges and Innovations in Building a Product Knowledge Graph**	*Dong*

09:30 AM	**End-to-end Learning for Broad Coverage Semantics: SRL, Coreference, and Beyond**	*Zettlemoyer*
10:00 AM	**Graph Convolutional Networks for Extracting and Modeling Relational Data**	*Titov*
10:30 AM	**Poster Session - Session 1**	
11:30 AM	**Learning Hierarchical Representations of Relational Data**	*Nickel*
12:00 PM	**Reading and Reasoning with Neural Program Interpreters**	*Riedel*
02:00 PM	**Multimodal KB Extraction and Completion**	*Singh*
02:30 PM	**Go for a Walk and Arrive at the Answer: Reasoning Over Knowledge Bases with Reinforcement Learning**	
02:45 PM	**Multi-graph Affinity Embeddings for Multilingual Knowledge Graphs**	
03:00 PM	**A Study of Automatically Acquiring Explanatory Inference Patterns from Corpora of Explanations: Lessons from Elementary Science Exams**	
03:15 PM	**NELL: Lessons and Future Directions**	*Mitchell*
03:45 PM	**Poster Session - Session 2**	*Rawat, Joulin, Jansen, Lee, Chen, Xu, Verga, Juba, Dumitrache, Jat, Logan, Sridhar, Yang, Das, Pezeshkpour, Monath*
04:45 PM	**Discussion Panel and Debate**	
05:45 PM	**Closing Remarks**	

Abstracts (7):

Abstract 1: **Challenges and Innovations in Building a Product Knowledge Graph in 6th Workshop on Automated Knowledge Base Construction (AKBC)**, *Dong* 09:00 AM

Knowledge graphs have been used to support a wide range of applications and enhance search results for multiple major search engines, such as Google and Bing. At Amazon we are building a Product Graph, an authoritative knowledge graph for all products in the world. The thousands of product verticals we need to model, the vast number of data sources we need to extract knowledge from, the huge volume of new products we need to handle every day, and the various applications in Search, Discovery, Personalization, Voice, that we wish to support, all present big challenges in constructing such a graph.

In this talk we describe four scientific directions we are investigating in building and using such a graph, namely, harvesting product knowledge from the web, hands-off-the-wheel knowledge integration and cleaning, human-in-the-loop knowledge learning, and graph mining and graph-enhanced search. This talk will present our progress to achieve near-term goals in each direction, and show the many research opportunities towards our moon-shot goals.

Abstract 2: **End-to-end Learning for Broad Coverage Semantics: SRL, Coreference, and Beyond in 6th Workshop on Automated Knowledge Base Construction (AKBC)**, *Zettlemoyer* 09:30 AM

Deep learning with large supervised training sets has had significant impact on many research challenges, from speech recognition to machine translation. However, applying these ideas to problems in computational semantics has been difficult, at least in part due to modest dataset sizes and relatively complex structured prediction tasks.

In this talk, I will present two recent results on end-to-end deep learning for classic challenge problems in computational semantics: semantic role labeling and coreference resolution. In both cases, we will introduce relative simple deep neural network approaches that use no preprocessing (e.g. no POS tagger or syntactic parser) and achieve significant performance gains, including over 20% relative error reductions when compared to non-neural methods. I will also discuss our first steps towards scaling the amount of data such methods can be trained on by many orders of magnitude, including semi-supervised learning via contextual word embeddings and supervised learning through crowdsourcing. Our hope is that these advances, when combined, will enable very high quality semantic analysis in any domain from easily gathered supervision.

Abstract 3: **Graph Convolutional Networks for Extracting and Modeling Relational Data in 6th Workshop on Automated Knowledge Base Construction (AKBC)**, *Titov* 10:00 AM

Graph Convolutional Networks (GCNs) is an effective tool for modeling graph structured data. We investigate their applicability in the context of both extracting semantic relations from text (specifically, semantic role labeling) and modeling relational data (link prediction). For semantic role labeling, we introduce a version of GCNs suited to modeling syntactic dependency graphs and use them as sentence encoders. Relying on these linguistically-informed encoders, we achieve the best reported scores on standard benchmarks for Chinese and English. For link prediction, we propose Relational GCNs (RGCNs), GCNs developed specifically to deal with highly multi-relational data, characteristic of realistic knowledge
bases. By explicitly modeling neighbourhoods of entities, RGCNs accumulate evidence over multiple inference steps in relational graphs and yield competitive results on standard link prediction benchmarks.

Joint work with Diego Marcheggiani, Michael Schlichtkrull, Thomas Kipf, Max Welling, Rianna van den Berg and Peter Bloem.

Abstract 5: **Learning Hierarchical Representations of Relational Data in 6th Workshop on Automated Knowledge Base Construction (AKBC)**, *Nickel* 11:30 AM

Representation learning has become an invaluable approach for making statistical inferences from relational data. However, while complex relational datasets often exhibit a latent hierarchical structure, state-of-the-art embedding methods typically do not account for this property. In this talk, I will introduce a novel approach to learning such hierarchical representations of symbolic data by embedding them into hyperbolic space -- or more precisely into an n-dimensional Poincaré ball. I will discuss how the underlying hyperbolic geometry allows us to learn parsimonious representations which simultaneously capture hierarchy and similarity. Furthermore, I will show that Poincaré embeddings can outperform Euclidean embeddings significantly on data with latent hierarchies, both in terms of representation capacity and in terms of generalization ability.

Abstract 6: **Reading and Reasoning with Neural Program Interpreters in 6th Workshop on Automated Knowledge Base Construction (AKBC)**, *Riedel* 12:00 PM

We are getting better at teaching end-to-end neural models how to answer questions about content in natural language text. However, progress has been mostly restricted to extracting answers that are directly stated in text. In this talk, I will present our work towards teaching machines not only to read, but also to reason with what was read and to do this in a interpretable and controlled fashion. Our main hypothesis is that this can be achieved by the development of neural abstract machines that follow the blueprint of program interpreters for real-world programming languages. We test this idea using two languages: an imperative (Forth) and a declarative (Prolog/Datalog) one. In both cases we implement differentiable interpreters that can be used for learning reasoning patterns. Crucially, because they are based on interpretable host languages, the interpreters also allow users to easily inject prior knowledge and inspect the learnt patterns. Moreover, on tasks such as math word problems and relational reasoning our approach compares favourably to state-of-the-art methods.

Abstract 7: **Multimodal KB Extraction and Completion in 6th Workshop on Automated Knowledge Base Construction (AKBC)**, *Singh* 02:00 PM

Existing pipelines for constructing KBs primarily support a restricted set of
data types, such as focusing on the text of the documents when extracting information, ignoring the various modalities of evidence that we regularly encounter, such as images, semi-structured tables, video, and audio. Similarly, approaches that reason over incomplete and uncertain KBs are limited to basic entity-relation graphs, ignoring the diversity of data types that are useful for relational reasoning, such as text, images, and numerical attributes. In this work, we present a novel AKBC pipeline that takes the first steps in combining textual and relational evidence with other sources like numerical, image, and tabular data. We focus on two tasks: single entity attribute extraction from documents and relational knowledge graph completion. For each, we introduce new datasets that contain multimodal information, propose benchmark evaluations, and develop models that build upon advances in deep neural encoders for different data types.

Abstract 11: **NELL: Lessons and Future Directions in 6th Workshop on Automated Knowledge Base Construction (AKBC)**, *Mitchell* 03:15 PM

The Never Ending Language Learner (NELL) research project has produced a computer program that has been running continuously since January 2010, learning to build a large knowledge base by extracting structured beliefs (e.g., PersonFoundedCompany(Gates,Microsoft), BeverageServedWithBakedGood(tea,crumpets)) from unstructured text on the web. This talk will provide an update on new NELL research results, reflect on the lessons learned from this effort, and discuss specific challenges for future systems that attempt to build large knowledge bases automatically.

Competition track

Sergio Escalera, Markus Weimer

103 A+B, 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

08:15 AM	**Opening**	*Escalera, Weimer*
08:30 AM	**AI XPRIZE Milestone award**	*Banifatemi, khalaf, McGregor*
09:00 AM	**Competition I: Adversarial Attacks and Defenses**	*Kurakin, Goodfellow, Bengio, Zhao, Dong, Pang, Liao, Xie, Ganesh, Elibol*
10:00 AM	**Coffee Break and Competition I: Adversarial Attacks and Defenses posters**	
10:30 AM	**Competition II: Learning to Run**	*Kidzi■ski, Ong, Mohanty, Fries, Hicks, Zheng, Yuan, Plis*
12:15 PM	**Lunch**	
01:15 PM	**DeepArt competition**	*Ecker, Gatys, Bethge*
01:30 PM	**Competition III: The Conversational Intelligence Challenge**	*Burtsev, Lowe, Serban, Bengio, Rudnicky, Black, Prabhumoye, Rodichev, Smetanin, Fedorenko, Lee, HONG, Lee, Kim, Gontier, Saito, Gershfeld, Burachenok*

03:00 PM	**Coffee break and Competition III: The Conversational Intelligence Challenge posters**	
03:30 PM	**Competition IV: Classifying Clinically Actionable Genetic Mutations**	*Huerga, Grigorenko, Das, Thorbergsson, Kan, Zhang, Chen, Li, Kumar, Gao*
05:15 PM	**Break**	
05:45 PM	**Competition V: Human-Computer Question Answering**	*Ying, Daume III, He, Iyyer, Rodriguez*
07:30 PM	**Closing**	*Escalera, Weimer*
07:30 PM	**Dinner**	

Abstracts (6):

Abstract 2: **AI XPRIZE Milestone award in Competition track**, *Banifatemi, khalaf, McGregor* 08:30 AM

- Overview of the Challenge
- Overview of top 10 competitors
- Presentation by top 2 competitors
- Winners recognition on stage
- Keynote: XPRIZE scientific Advisor

Abstract 3: **Competition I: Adversarial Attacks and Defenses in Competition track**, *Kurakin, Goodfellow, Bengio, Zhao, Dong, Pang, Liao, Xie, Ganesh, Elibol* 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**, *Kidzi■ski, Ong, Mohanty, Fries, Hicks, Zheng, Yuan, Plis* 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, Prabhumoye, Rodichev, Smetanin, Fedorenko, Lee, HONG, Lee, Kim, Gontier, Saito, Gershfeld, Burachenok* 01:30 PM

* Overview of the Challenge
* Awarding prize
* Short presentation by winning team
* Inspirational talk: Alexander Rudnicky
Abstract: Current automatic systems use limited models of conversation with humans. Historically research in dialog systems has focused on goal-directed interaction for which objective measures of success could be defined. More recently attention has turned to so-called open-domain conversation chatbots, similar to social exchanges between people but for which measures of success are not yet agreed upon. On reflection, natural human communication is mostly a combination of task, social and likely other levels. The challenge is to understand how to manage such communication in a way that reflects human capabilities.
* Panel Discussion. Moderated by Julian Serban. Panelists: Alexander Rudnicky, Dilek Hakkani-Tur, Jianfeng Gao and Joelle Pineau.

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

* Overview of the Challenge
* Talks from participants
* Competition against human team
* Q&A with organizers, experts, developers

Machine Learning for Health (ML4H) - What Parts of Healthcare are Ripe for Disruption by Machine Learning Right Now?

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

104 A, 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

08:00 AM	**Welcome and opening remarks**	
08:20 AM	**Keynote: Zak Kohane, Harvard DBMI**	*Kohane*
08:55 AM	**Jennifer Chayes, Microsoft Research New England**	*Chayes*
09:20 AM	**Keynote: Susan Murphy, U. Michigan**	*Murphy*
09:55 AM	**Contributed spotlights**	
10:20 AM	**Coffee break and Poster Session I**	*Khandwala, Gallant, Way, Raghu, Shen, Gasimova, Bozkurt, Boag, Lopez-Martinez, Bodenhofer, Nasiri GhoshehBolagh, Guo, Kurz, Pillay, Perros, Chen, Yahi, Sushil, Purushotham, Tutubalina, Virdi, Schulz, Weisenthal, Srikishan, Veli■kovi■, Ahuja, Miller, Craig, Ji, Dabek, Pou-Prom, Zhang, Kalyanam, Weng, Bhat, Chen, Kohl, Gao, Zhu, Poh, Urteaga, Honoré, De Palma, Al-Shedivat, Rajpurkar, McDermott, Chen, Sui, Lee, Cheng, Fang, Hussain, Furlanello, Waks, Chougrad, Kjellstrom, Doshi-Velez, Fruehwirt, Zhang, Hu, Chen, Park, Mikelsons, Dakka, Hyland, chevaleyre, Lee, Giro-i-Nieto, Kale, Hughes, Erion, Mehra, Zame, Trajanovski, Chakraborty, Peterson, Srivastava, Jin, Tejeda Lemus, Ray, Madl, Futoma, Gong, Ahmad, Lei, Legros*
10:50 AM	**Invited clinical panel**	*Velazquez, Priest, strigo*
11:50 AM	**Keynote II: Fei-Fei Li, Stanford**	*Fei-Fei*
01:30 PM	**Interactive panel**	
02:30 PM	**Jill Mesirov, UCSD**	*Mesirov*
02:55 PM	**Greg Corrado, Google**	*Corrado*
03:20 PM	**Coffee break and Poster Session II**	*Kane, Haque, Papalexakis, Guibas, Li, Arias, Nalisnick, Smyth, Rudzicz, Zhu, Willke, Elhadad, Raffauf, Suresh, Varma, Yue, Rudovic, Foschini, Ahmad, Haq, Maggio, Jurman, Parbhoo, Bashivan, Islam, Musolesi, Wu, Ratner, Dunnmon, Esteban, Galstyan, Ver Steeg, Khachatrian, Górriz, van der Schaar, Nemchenko, Patwardhan, Tandon*
03:50 PM	**Award session + A word from our affiliates**	
04:10 PM	**Mihaela Van Der Schaar, Oxford**	
04:35 PM	**Networking Break**	
05:00 PM	**Jure Leskovec, Stanford**	*Leskovec*

05:25 PM	**Keynote: Atul Butte**	*Butte*
06:00 PM	**Closing Remarks**	

Abstracts (3):

Abstract 7: **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* 10:50 AM

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

Abstract 9: **Interactive panel in Machine Learning for Health (ML4H) - What Parts of Healthcare are Ripe for Disruption by Machine Learning Right Now?**, 01:30 PM

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

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?**, 03:50 PM

Award session and a word from affiliates
- Eunho Yang, KAIST, Korea
- Sung-ju Hwang,UNIST, Korea
representing AItrics, DeepMind, IBM, Google, MSR

Acting and Interacting in the Real World: Challenges in Robot Learning

Ingmar Posner, Raia Hadsell, Martin Riedmiller, Markus Wulfmeier, Rohan Paul

104 B, Fri Dec 08, 08:00 AM

In recent years robotics has made significant strides towards applications of real value to the public domain. Robots are now increasingly expected to work for and alongside us in complex, dynamic environments. Machine learning has been a key enabler of this success, particularly in the realm of robot perception where, due to substantial overlap with the machine vision community, methods and training data can be readily leveraged.

Recent advances in reinforcement learning and learning from demonstration — geared towards teaching agents how to act — provide a tantalising glimpse at a promising future trajectory for robot learning. Mastery of challenges such as the Atari suite and AlphaGo build significant excitement as to what our robots may be able to do for us in the future. However, this success relies on the ability of learning cheaply, often within the confines of a virtual environment, by trial and error over as many episodes as required. This presents a significant challenge for embodied systems acting and interacting in the real world. Not only is there a cost (either monetary or in terms of execution time) associated with a particular trial, thus limiting the amount of training data obtainable, but there also exist safety constraints which make an exploration of the state space simply unrealistic: teaching a real robot to cross a real road via reinforcement learning for now seems a noble yet somewhat far fetched goal. A significant gulf therefore exists between prior art on teaching agents to act and effective approaches to real-world robot learning. This, we posit, is one of the principal impediments at the moment in advancing real-world robotics science.

In order to bridge this gap researchers and practitioners in robot learning have to address a number of key challenges to allow real-world systems to be trained in a safe and data-efficient manner. This workshop aims to bring together experts in reinforcement learning, learning from demonstration, deep learning, field robotics and beyond to discuss what the principal challenges are and how they might be addressed. With a particular emphasis on data efficient learning, of particular interest will be contributions in representation learning, curriculum learning, task transfer, one-shot learning, domain transfer (in particular from simulation to real-world tasks), reinforcement learning for real world applications, learning from demonstration for real world applications, knowledge learning from observation and interaction, active concept acquisition and learning causal models.

Schedule

08:50 AM	**Welcome**	
09:00 AM	**Pieter Abbeel: Reducing Data Needs for Real-World Reinforcement Learning**	
09:30 AM	**Pierre Sermanet: Self-Supervised Imitation**	*Sermanet*
11:30 AM	**Raquel Urtasun: Deep Learning for Self-Driving Cars**	*Urtasun*
02:00 PM	**Jitendra Malik: Vision for Manipulation and Navigation**	*Malik*
02:30 PM	**Martial Hebert: Reducing supervision**	*Hebert*
03:00 PM	**Poster Session**	*Bruce, Quillen, Rakicevic, Chua, Schenck, Chien, Babaeizadeh, Wichers, yan, Wohlhart, Ibarz, Konolige*

Abstracts (6):

Abstract 2: **Pieter Abbeel: Reducing Data Needs for Real-World Reinforcement Learning in Acting and Interacting in the Real World: Challenges in Robot Learning**, 09:00 AM

Reinforcement learning and imitation learning have seen success in many domains, including autonomous helicopter flight, Atari, simulated locomotion, Go, robotic manipulation. However, sample complexity of these methods remains very high. In this talk I will present several ideas towards reducing sample complexity: (i) Hindsight Experience Replay, which infuses learning signal into (traditionally) zero-reward runs, and is compatible with existing off-policy algorithms; (ii) Some recent advances in Model-based Reinforcement Learning, which achieve 100x sample complexity gain over the more widely studied model-free methods; (iii) Meta-Reinforcement Learning, which can significantly reduce sample complexity by building off other skills acquired in the past; (iv) Domain Randomization, a simple idea that can often enable training fully in simulation, yet still recover policies that perform well in the real world.

Abstract 3: **Pierre Sermanet: Self-Supervised Imitation in Acting and Interacting in the Real World: Challenges in Robot Learning**, *Sermanet* 09:30 AM

We propose a self-supervised approach for learning representations and robotic behaviors entirely from unlabeled videos recorded from multiple viewpoints. We study how these representations can be used in two robotic imitation settings: imitating object interactions from videos of humans, and imitating human poses. Imitation of human behavior requires a viewpoint-invariant representation that captures the relationships between end-effectors (hands or robot grippers) and the environment, object attributes, and body pose. We train our representations using a triplet loss, where multiple simultaneous viewpoints of the same observation are attracted in the embedding space, while being repelled from temporal neighbors which are often visually similar but functionally different. This signal causes our model to discover attributes that do not change across viewpoint, but do change across time, while ignoring nuisance variables such as occlusions, motion blur, lighting and background. We demonstrate that this representation can be used by a robot to directly mimic human poses without an explicit correspondence, and that it can be used as a reward function within a reinforcement learning algorithm. While representations are learned from an unlabeled collection of task-related videos, robot behaviors such as pouring are learned by watching a single 3rd-person demonstration by a human. Reward functions obtained by following the human demonstrations under the learned representation enable efficient reinforcement learning that is practical for real-world robotic systems. Video results, open-source code and dataset are available at https://sermanet.github.io/imitate

Abstract 4: **Raquel Urtasun: Deep Learning for Self-Driving Cars in Acting and Interacting in the Real World: Challenges in Robot Learning**, *Urtasun* 11:30 AM

Raquel Urtasun is the Head of Uber ATG Toronto. She is also an Associate Professor in the Department of Computer Science at the University of Toronto, a Raquel Urtasun is the Head of Uber ATG Toronto. She is also an Associate Professor in the Department of Computer Science at the University of Toronto, a Canada Research Chair in Machine Learning and Computer Vision and a co-founder of the Vector Institute for AI. Prior to this, she was an Assistant Professor at the Toyota Technological Institute at Chicago (TTIC), an academic computer science institute affiliated with the University of Chicago. She was also a visiting professor at ETH Zurich during the spring semester of 2010. She received her Bachelors degree from Universidad Publica de Navarra in 2000, her Ph.D. degree from the Computer Science department at Ecole Polytechnique Federal de Lausanne (EPFL) in 2006 and did her postdoc at MIT and UC Berkeley. She is a world leading expert in machine perception for self-driving cars. Her research interests include machine learning, computer vision, robotics and remote sensing. Her lab was selected as an NVIDIA NVAIL lab. She is a recipient of an NSERC EWR Steacie Award, an NVIDIA Pioneers of AI Award, a Ministry of Education and Innovation Early Researcher Award, three Google Faculty Research Awards, an Amazon Faculty Research Award, a Connaught New Researcher Award and two Best Paper Runner up Prize awarded at the Conference on Computer Vision and Pattern Recognition (CVPR) in 2013 and 2017 respectively. She is also an Editor of the International Journal in Computer Vision (IJCV) and has served as Area Chair of multiple machine learning and vision conferences (i.e., NIPS, UAI, ICML, ICLR, CVPR, ECCV).

Abstract 5: **Jitendra Malik: Vision for Manipulation and Navigation in Acting and Interacting in the Real World: Challenges in Robot Learning**, *Malik* 02:00 PM

I will describe recent results from my group on visually guided manipulation and navigation. We are guided considerably by insights from human development and cognition. In manipulation, our work is based on object-oriented task models acquired by experimentation. In navigation, we show the benefits of architectures based on cognitive maps and landmarks.

Abstract 6: **Martial Hebert: Reducing supervision in Acting and Interacting in the Real World: Challenges in Robot Learning**, *Hebert* 02:30 PM

A key limitation, in particular for computer vision tasks, is their reliance on vast amounts of strongly supervised data. This limits scalability, prevents rapid acquisition of new concepts, and limits adaptability to new tasks or new conditions. To address this limitation, I will explore ideas in learning visual models from limited data. The basic insight behind all of these ideas is that it is possible to learn from a large corpus of vision tasks how to learn models for new tasks with limited data, by representing the way visual models vary across tasks, also called model dynamics. The talk will also show examples from common visual classification tasks.

Abstract 7: **Poster Session in Acting and Interacting in the Real World: Challenges in Robot Learning**, *Bruce, Quillen, Rakicevic, Chua, Schenck, Chien, Babaeizadeh, Wichers, yan, Wohlhart, Ibarz, Konolige* 03:00 PM

Spotlights:
Deep Object-Centric Representations for Generalizable Robot Learning
<
Coline Devin>

Using Simulation and Domain Adaptation to Improve Efficiency of Deep Robotic Grasping

Learning Deep Composable Maximum-Entropy Policies for Real-World Robotic Manipulation

SE3-Pose-Nets: Structured Deep Dynamics Models for Visuomotor Control

Learning Flexible and Reusable Locomotion Primitives for a Microrobot

Policy Search using Robust Bayesian Optimization

Learning Robotic Assembly from CAD

Learning Robot Skill Embeddings

Self-Supervised Visual Planning with Temporal Skip Connections

Overcoming Exploration in Reinforcement Learning with Demonstrations

Deep Reinforcement Learning for Vision-Based Robotic Grasping

Soft Value Iteration Networks for Planetary Rover Path Planning

Posters:
One-Shot Visual Imitation Learning via Meta-Learning

One-Shot Reinforcement Learning for Robot Navigation with Interactive Replay < Jake Bruce; Niko Suenderhauf; Piotr Mirowski; Raia Hadsell; Michael Milford >

Bayesian Active Edge Evaluation on Expensive Graphs < Sanjiban Choudhury >

Sim-to-Real Transfer of Accurate Grasping with Eye-In-Hand Observations and Continuous Control < Mengyuan Yan; Iuri Frosio*; Stephen Tyree; Kautz Jan >

Learning Robotic Manipulation of Granular Media < Connor Schenck*; Jonathan Tompson; Dieter Fox; Sergey Levin>

End-to-End Learning of Semantic Grasping < Eric Jang >

Self-supervised Deep Reinforcement Learning with Generalized Computation Graphs for Robot Navigation

Efficient Robot Task Learning and Transfer via Informed Search in Movement Parameter Space < Nemanja Rakicevic*; Kormushev Petar >

Metrics for Deep Generative Models based on Learned Skills

Unsupervised Hierarchical Video Prediction < Nevan wichers*; Dumitru Erhan; Honglak Lee >

Imitation from Observation: Learning to Imitate Behaviors from Raw Video via Context Translation

Domain Randomization and Generative Models for Robotic Grasping

Learning to Grasp from Vision and Touch

Neural Network Dynamics Models for Control of Under-actuated Legged Millirobots

On the Importance of Uncertainty for Control with Deep Dynamics Models

Increasing Sample-Efficiency via Online Meta-Learning

Stochastic Variational Video Prediction

(Author information copied from CMT please contact the workshop organisers under nips17robotlearning@gmail.com for any changes)

Deep Learning for Physical Sciences

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

104 C, 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

08:50 AM	**Introduction and opening remarks**	
09:00 AM	**Invited talk 1: Deep recurrent inverse modeling for radio astronomy and fast MRI imaging**	*Welling*
09:40 AM	**Contributed talk 1: Neural Message Passing for Jet Physics**	*Henrion*
10:00 AM	**Contributed talk 2: A Foray into Using Neural Network Control Policies For Rapid Switching Between Beam Parameters in a Free Electron Laser**	*Edelen*
10:20 AM	**Poster session 1 and coffee break**	*Hagge, McGregor, Stoye, Pham, Hong, Farbin, Seo, Zoghbi, George, Fort, Farrell, Pajot, Pearson, McCarthy, Germain, Anderson, Lezcano Casado, Mudigonda, Nachman, de Oliveira, Jing, Li, Kim, Gebhard, Zahavy*
11:00 AM	**Invited talk 2: Adversarial Games for Particle Physics**	*Louppe*

11:40 AM	**Contributed talk 3: Implicit Causal Models for Genome-wide Association Studies**	*Tran*
12:00 PM	**Contributed talk 4: Graphite: Iterative Generative Modeling of Graphs**	*Grover*
12:20 PM	**Sponsor presentation: Intel Nervana**	*Tang*
12:25 PM	**Lunch break**	
02:00 PM	**Invited talk 3: Learning priors, likelihoods, or posteriors**	*Murray*
02:40 PM	**Contributed talk 5: Deep Learning for Real-time Gravitational Wave Detection and Parameter Estimation with Real LIGO Data**	*George*
03:00 PM	**Poster session 2 and coffee break**	*McGregor, Hagge, Stoye, Pham, Hong, Farbin, Seo, Zoghbi, George, Fort, Farrell, Pajot, Pearson, McCarthy, Germain, Anderson, Lezcano Casado, Mudigonda, Nachman, de Oliveira, Jing, Li, Kim, Gebhard, Zahavy*
04:00 PM	**Invited talk 4: A machine learning perspective on the many-body problem in classical and quantum physics**	*Carrasquilla*
04:40 PM	**Invited talk 5: Quantum Machine Learning**	*von Lilienfeld*
05:20 PM	**Contributed talk 6: Physics-guided Learning of Neural Networks: An Application in Lake Temperature Modeling**	*Karpatne*
05:40 PM	**Panel session**	*Murray, Welling, Carrasquilla, von Lilienfeld, Louppe, Cranmer*
06:40 PM	**Closing remarks**	

Machine Learning for Audio Signal Processing (ML4Audio)

Hendrik Purwins, Bob L. Sturm, Mark Plumbley

201 A, Fri Dec 08, 08:00 AM

Abstracts and full papers: http://media.aau.dk/smc/ml4audio/

Audio signal processing is currently undergoing a paradigm change, where data-driven machine learning is replacing hand-crafted feature design. This has led some to ask whether audio signal processing is still useful in the "era of machine learning." There are many challenges, new and old, including the interpretation of learned models in high dimensional spaces, problems associated with data-poor domains, adversarial examples, high computational requirements, and research driven by companies using large in-house datasets that is ultimately not reproducible.

ML4Audio aims to promote progress, systematization, understanding, and convergence of applying machine learning in the area of audio signal processing. Specifically, we are interested in work that demonstrates novel applications of machine learning techniques to audio data, as well as methodological considerations of merging machine learning with audio signal processing. We seek contributions in, but not limited to, the following topics:
- audio information retrieval using machine learning;
- audio synthesis with given contextual or musical constraints using machine learning;
- audio source separation using machine learning;
- audio transformations (e.g., sound morphing, style transfer) using machine learning;
- unsupervised learning, online learning, one-shot learning, reinforcement learning, and incremental learning for audio;
- applications/optimization of generative adversarial networks for audio;
- cognitively inspired machine learning models of sound cognition;
- mathematical foundations of machine learning for audio signal processing.

This workshop especially targets researchers, developers and musicians in academia and industry in the area of MIR, audio processing, hearing instruments, speech processing, musical HCI, musicology, music technology, music entertainment, and composition.

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)
Shlomo 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 SVAIL)
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

08:00 AM	**Overture**	*Purwins*
08:15 AM	**Acoustic word embeddings for speech search**	*Livescu*
08:45 AM	**Learning Word Embeddings from Speech**	*Glass, Chung*
09:05 AM	**Multi-Speaker Localization Using Convolutional Neural Network Trained with Noise**	*Chakrabarty, Habets*
09:25 AM	**Adaptive Front-ends for End-to-end Source Separation**	*Venkataramani, Smaragdis*
09:45 AM	**Poster Session Speech: source separation, enhancement, recognition, synthesis**	*Zarar, Fakoor, Dumpala, Kim, Smaragdis, Dubey, Ko, Sakti, Wang, Guo, Kenyon, Tjandra, Tax, Lee*
11:00 AM	**Learning and transforming sound for interactive musical applications**	*Marchini*
11:30 AM	**Compact Recurrent Neural Network based on Tensor Train for Polyphonic Music Modeling**	*Sakti*
11:50 AM	**Singing Voice Separation using Generative Adversarial Networks**	*Choi, Lee*
12:10 PM	**Audio Cover Song Identification using Convolutional Neural Network**	*Chang, Lee*
12:30 PM	**Lunch Break**	
01:30 PM	**Polyphonic piano transcription using deep neural networks**	*Eck*
02:00 PM	**Deep learning for music recommendation and generation**	*Dieleman*
02:30 PM	**Exploring Ad Effectiveness using Acoustic Features**	*Prockup, Vahabi*
03:00 PM	**Poster Session Music and environmental sounds**	*Nieto, Pons, Raj, Tax, Elizalde, Nam, Kumar*
04:00 PM	**Sight and sound**	*Freeman*
04:30 PM	**k-shot Learning of Acoustic Context**	*de Vries*
04:50 PM	**Towards Learning Semantic Audio Representations from Unlabeled Data**	*Jansen*
05:10 PM	**Cost-sensitive detection with variational autoencoders for environmental acoustic sensing**	*Li, Roberts*
05:30 PM	**Break**	
05:45 PM	**Panel: Machine learning and audio signal processing: State of the art and future perspectives**	*Hochreiter, Li, Livescu, Mandal, Nieto, Slaney, Purwins*

Abstracts (18):

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.

Karen Livescu is an Associate Professor at TTI-Chicago. She completed her PhD and post-doc in electrical engineering and computer science at MIT and her Bachelor's degree in Physics at Princeton University. Karen's main research interests are at the intersection of speech and language processing and machine learning. Her recent work includes multi-view representation learning, segmental neural models, acoustic word embeddings, and automatic sign language recognition. She is a member of the IEEE Spoken Language Technical Committee, an associate editor for IEEE Transactions on Audio, Speech, and Language Processing, and a technical co-chair of ASRU 2015 and 2017.

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

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

The problem of multi-speaker localization is formulated 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

(+ Jonah Casebeer)
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.

Abstract 6: **Poster Session Speech: source separation, enhancement, recognition, synthesis in Machine Learning for Audio Signal Processing (ML4Audio)**, *Zarar, Fakoor, Dumpala, Kim, Smaragdis, Dubey, Ko, Sakti, Wang, Guo, Kenyon, Tjandra, Tax, Lee* 09:45 AM

Poster abstracts and full papers: http://media.aau.dk/smc/ml4audio/

SPEECH SOURCE SEPARATION
*Lijiang Guo and Minje Kim. Bitwise Source Separation on Hashed Spectra: An Efficient Posterior Estimation Scheme Using Partial Rank Order Metrics
*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

AUTOMATIC SPEECH RECOGNITION
*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 Kopparapu. 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 Rabiee and Soo-Young Lee. Emotional End-to-End Neural Speech Synthesizer

Abstract 7: **Learning and transforming sound for interactive musical applications in Machine Learning for Audio Signal Processing (ML4Audio)**, *Marchini* 11:00 AM

Recent developments in object recognition (especially convolutional neural networks) led to a new spectacular application: image style transfer. But what would be the music version of style transfer? In the flow-machine project, we created diverse tools for generating audio tracks by transforming prerecorded music material. Our artists integrated these tools in their composition process and produced some pop tracks. I present some of those tools, with audio examples, and give an operative definition of music style transfer as an optimization problem. Such definition allows for an efficient solution which renders possible a multitude of musical applications: from composing to live performance.

Marco Marchini works at Spotify in the Creator Technology Research Lab, Paris. His mission is bridging the gap between between creative artists and intelligent technologies.
Previously, he worked as research assistant for the Pierre-and-Marie-Curie University at the Sony Computer Science Laboratory of Paris and worked for the Flow Machine project. His previous academic research also includes unsupervised music generation and ensemble performance analysis, this research was carried out during my M.Sc. and Ph.D. at the Music Technology Group (DTIC, Pompeu Fabra University). He has a double degree in Mathematics from Bologna University.

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

(+Andros Tjandra, Satoshi Nakamura)
This paper introduces a novel compression method for recurrent neural networks (RNNs) based on Tensor Train (TT) format. The objective in this work are to reduce the number of parameters in RNN and maintain their expressive power. The key of our approach is to represent 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.

Abstract 9: **Singing Voice Separation using Generative Adversarial Networks in Machine Learning for Audio Signal Processing (ML4Audio)**, *Choi, Lee* 11:50 AM

(+Ju-heon Lee)
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, Lee* 12:10 PM

(+Juheon Lee, Sankeun Choe)
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 12: **Polyphonic piano transcription using deep neural networks in Machine Learning for Audio Signal Processing (ML4Audio)**, *Eck* 01:30 PM

I'll discuss the problem of transcribing polyphonic piano music with an emphasis on generalizing to unseen instruments. We optimize for two objectives. We first predict pitch onset events and then conditionally predict pitch at the frame level. I'll discuss the model architecture, which combines CNNs and LSTMs. I'll also discuss challenges faced in robust piano transcription, such as obtaining enough data to train a good model I'll also provide some demos and links to working code. This collaboration was led by Curtis Hawthorne, Erich Elsen and Jialin Song (https://arxiv.org/abs/1710.11153).

Douglas Eck works at the Google Brain team on the Magenta project, an effort to generate music, video, images and text using machine intelligence.
He also worked on music search and recommendation for Google Play Music. I was an Associate Professor in Computer Science at University of Montreal in the BRAMS research center. He also worked on music performance modeling.

Abstract 13: **Deep learning for music recommendation and generation in Machine Learning for Audio Signal Processing (ML4Audio)**, *Dieleman* 02:00 PM

The advent of deep learning has made it possible to extract high-level information from perceptual signals without having to specify manually and explicitly how to obtain it; instead, this can be learned from examples. This creates opportunities for automated content analysis of musical audio signals. In this talk, I will discuss how deep learning techniques can be used for audio-based music recommendation. I will also discuss my ongoing work on music generation in the raw waveform domain with WaveNet.

Sander Dieleman is a Research Scientist at DeepMind in London, UK, where he has worked on the development of AlphaGo and WaveNet. He was previously a PhD student at Ghent University, where he conducted research on feature learning and deep learning techniques for learning hierarchical representations of musical audio signals. During his PhD he also developed the Theano-based deep learning library Lasagne and won solo and team gold medals respectively in Kaggle's "Galaxy Zoo" competition and the first National Data Science Bowl. In the summer of 2014, he interned at Spotify in New York, where he worked on implementing audio-based music recommendation using deep learning on an industrial scale.

Abstract 14: **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.

Matt Prockup is currently a scientist at Pandora working on methods and tools for Music Information Retrieval at scale. He recently received his Ph.D. in Electrical Engineering from Drexel University. His research interests span a wide scope of topics including audio signal processing, recommender systems, machine learning, and human computer interaction. He is also an avid drummer, percussionist, and composer.

Puya - Hossein Vahabi is a senior research scientist at Pandora working on Audio/Video Computational Advertising. Before Pandora, he was a research scientist at Yahoo Labs. He has a PhD in CS, and he has been a research associate of the Italian National Research for many years. He has a PhD in CS, and his background is on computational advertising, graph mining and information retrieval.

Abstract 15: **Poster Session Music and environmental sounds in Machine Learning for Audio Signal Processing (ML4Audio)**, *Nieto, Pons, Raj, Tax, Elizalde, Nam, Kumar* 03:00 PM

Poster abstracts and full papers: http://media.aau.dk/smc/ml4audio/

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 Maaløe 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 16: **Sight and sound in Machine Learning for Audio Signal Processing (ML4Audio)**, *Freeman* 04:00 PM

William T. Freeman is the Thomas and Gerd Perkins Professor of Electrical Engineering and Computer Science at MIT. His current research interests include motion rerendering, computational photography, and learning for vision. He received outstanding paper awards at computer vision or machine learning conferences in 1997, 2006, 2009 and 2012, and recently won "test of time" awards for papers written in 1991 and 1995. Previous research topics include steerable filters and pyramids, the generic viewpoint assumption, color constancy, bilinear models for separating style and content, and belief propagation in networks with loops. He holds 30 patents.

Abstract 17: **k-shot Learning of Acoustic Context in Machine Learning for Audio Signal Processing (ML4Audio)**, *de Vries* 04:30 PM

(+Ivan Bocharov, Tjalling Tjalkens)
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

(+ Manoj Plakal, Ratheet Pandya, Daniel P. W. Ellis, Shawn Hershey, Jiayang Liu, R. Channing Moore, Rif A. Saurous)
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 constraints to sample 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

(+ Ivan Kiskin, Davide Zilli, Marianne Sinka, Henry Chan, Kathy Willis)
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 hyper-parameters of classifiers and the structure of variational autoencoders are selected in a principled Neyman- Pearson framework. We examine the performance of the proposed approach using a dataset from the HumBug project1 which aims to detect the presence of mosquitoes using sound collected by simple embedded devices.

Abstract 21: **Panel: Machine learning and audio signal processing: State of the art and future perspectives in Machine Learning for Audio Signal Processing (ML4Audio)**, *Hochreiter, Li, Livescu, Mandal, Nieto, Slaney, Purwins* 05:45 PM

How can end-to-end audio processing be further optimized? How can an audio processing system be built that generalizes across domains, in particular different languages, music styles, or acoustic environments? How can complex musical hierarchical structure be learned? How can we use machine learning to build a music system that is able to react in the same way an improvisation partner would?
Can we build a system that could put a composer in the role of a perceptual engineer?

Sepp Hochreiter (Johannes Kepler University Linz, http://www.bioinf.jku.at/people/hochreiter/)
Bo Li (Google, https://research.google.com/pubs/BoLi.html)
Karen Livescu (Toyota Technological Institute at Chicago, http://ttic.uchicago.edu/~klivescu/)
Arindam Mandal (Amazon Alexa, https://scholar.google.com/citations?user=tV1hW0YAAAAJ&hl=en)
Oriol Nieto (Pandora, http://urinieto.com/about/)
Malcolm Slaney (Google, http://www.slaney.org/malcolm/pubs.html)
Hendrik Purwins (Aalborg University Copenhagen, http://personprofil.aau.dk/130346?lang=en)

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

George H Chen, Devavrat Shah, Christina Lee

201 B, 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

09:00 AM	**A millennium of nearest neighbor methods – an introduction to the NIPS nearest neighbor workshop 2017**	*Chen*
09:20 AM	**Analyzing Robustness of Nearest Neighbors to Adversarial Examples**	*Chaudhuri*
10:05 AM	**New perspective from Blackwell's "comparisons of experiments" on generative adversarial networks**	*Oh*
11:00 AM	**Data-dependent methods for similarity search in high dimensions**	*Indyk*
11:45 AM	**Fast k-Nearest Neighbor Search via Prioritized DCI**	*Li*
01:45 PM	**How to stop worrying and learn to love Nearest Neighbors**	*Efros*
02:30 PM	**A Platform for Data Science**	*Doshi*
02:50 PM	**Poster Session (encompasses coffee break)**	*Chen, Balle, Lee, frosio, Malik, Kautz, Li, Sugiyama, Carreira-Perpinan, Raziperchikolaei, Tulabandhula, Noh, Yu*
04:15 PM	**The Boundary Forest Algorithm**	*Yedidia*
05:00 PM	**Iterative Collaborative Filtering for Sparse Matrix Estimation**	*Lee*

Abstracts (10):

Abstract 1: **A millennium of nearest neighbor methods – an introduction to the NIPS nearest neighbor workshop 2017 in Nearest Neighbors for Modern Applications with Massive Data: An Age-old Solution with New Challenges**, *Chen* 09:00 AM

I give a brief history of nearest neighbor (NN) methods, starting from Alhazen's “Book of Optics” in the 11th century, and leading up to present time. Surprisingly, the most general nonasymptotic results on NN prediction only recently emerged in 2014 in the work of Chaudhuri and Dasgupta. Turning toward “user-friendly” theory with an eye toward practitioners, I mention recent guarantees (2013-2015) in the contemporary applications of time series forecasting, online collaborative filtering, and medical image segmentation — in all three cases, clustering structure enables successful NN prediction.

Abstract 2: **Analyzing Robustness of Nearest Neighbors to Adversarial Examples in Nearest Neighbors for Modern Applications with Massive Data: An Age-old Solution with New Challenges**, *Chaudhuri* 09:20 AM

Motivated by applications such as autonomous vehicles, test-time attacks via adversarial examples have received a great deal of recent attention. In this setting, an adversary is capable of making queries to a classifier, and perturbs a test example by a small amount in order to force the classifier to report an incorrect label. While a long line of work has explored a number of attacks, not many reliable defenses are known, and there is an overall lack of general understanding about the foundations of designing machine learning algorithms robust to adversarial examples.

In this talk, we take a step towards addressing this challenging question by introducing a new theoretical framework, analogous to bias-variance theory, which we can use to tease out the causes of vulnerability. We apply our framework to a simple classification algorithm: nearest neighbors, and analyze its robustness to adversarial examples. Motivated by our analysis, we propose a modified version of the nearest neighbor algorithm, and demonstrate both theoretically and empirically that it has superior robustness to standard nearest neighbors.

Abstract 3: **New perspective from Blackwell's "comparisons of experiments" on generative adversarial networks in Nearest Neighbors for Modern Applications with Massive Data: An Age-old Solution with New Challenges**, *Oh* 10:05 AM

We bring the tools from Blackwell's seminal result on comparing two stochastic experiments, to shine new lights on the modern applications of great interest: generative adversarial networks (GAN). Binary hypothesis testing is at the center of GANs, and we propose new data processing inequalities that allows us to discover new algorithms to combat mode collapse, provide sharper analyses, and provide simpler proofs. This leads to a new architecture to handle one of the major challenges in GAN known as ``mode collapse''; the lack of diversity in the samples generated by the learned generators. The hypothesis testing view of GAN allows us to analyze the new architecture and show that it encourages generators with no mode collapse. Experimental results show that the proposed architecture can learn to generate samples with diversity that is orders of magnitude better than competing approaches, while being simpler. For this talk, I will assume no prior background on GANs.

Abstract 4: **Data-dependent methods for similarity search in high dimensions in Nearest Neighbors for Modern Applications with Massive Data: An Age-old Solution with New Challenges**, *Indyk* 11:00 AM

I will give an overview of recent research on designing provable data-dependent approximate similarity search methods for high-dimensional data. Unlike many approaches proposed in the algorithms literature, the new methods lead to data structures that adapt to the data while retaining correctness and running time guarantees. The high-level principle behind those methods is that every data set (even a completely random one) has some structure that can be exploited algorithmically. This interpolates between basic "data-oblivious" approaches (e.g., those based on simple random projections) that are analyzable but do not adapt easily to the structure of the data, and basic "data-dependent" methods that exploit specific structure of the data, but might not perform well if this structure is not present.

Concretely, I will cover two recent lines of research:
- Data-dependent Locality-Sensitive Hashing: approximate nearest neighbor search methods that provably improve over the best possible data-oblivious LSH algorithms (covers work from SODA'14, STOC'15 and NIPS'15).
- Data-dependent metric compression: algorithms that represent sets of points using a small number of bits while approximately preserving the distances up to higher accuracy than previously known (covers work from SODA'17 and NIPS'17).

Abstract 5: **Fast k-Nearest Neighbor Search via Prioritized DCI in Nearest Neighbors for Modern Applications with Massive Data: An Age-old Solution with New Challenges**, *Li* 11:45 AM

Most exact methods for k-nearest neighbour search suffer from the curse of dimensionality; that is, their query times exhibit exponential dependence on either the ambient or the intrinsic dimensionality. Dynamic Continuous Indexing (DCI) [19] offers a promising way of circumventing the curse and successfully reduces the dependence of query time on intrinsic dimensionality from exponential to sublinear. In this paper, we propose a variant of DCI, which we call Prioritized DCI, and show a remarkable improvement in the dependence of query time on intrinsic dimensionality. In particular, a linear increase in intrinsic dimensionality, or equivalently, an exponential increase in the number of points near a query, can be mostly counteracted with just a linear increase in space. We also demonstrate empirically that Prioritized DCI significantly outperforms prior methods. In particular, relative to Locality-Sensitive Hashing (LSH), Prioritized DCI reduces the number of distance evaluations by a factor of 14 to 116 and the memory consumption by a factor of 21.

Abstract 6: **How to stop worrying and learn to love Nearest Neighbors in Nearest Neighbors for Modern Applications with Massive Data: An Age-old Solution with New Challenges**, *Efros* 01:45 PM

Nearest neighbors is an algorithm everyone loves to hate. It's too trivial, too brute-force, doesn't offer any insights. In this talk, I will try to make you fall in love with the humble nearest neighbor. First, I will discuss the use of nearest neighbor as an exceptionally useful baseline to guard against our field's natural bias in favor of elegant algorithms over data. Then, I will discuss some scenarios when nearest neighbors is particularly useful in practice.

Abstract 7: **A Platform for Data Science in Nearest Neighbors for Modern Applications with Massive Data: An Age-old Solution with New Challenges**, *Doshi* 02:30 PM

We describe an extensible data science platform based on non-parametric statistical methods such as nearest neighbors. This platform was borne out of the need for us to provide scalable and flexible predictive analytics solutions for retailers and the federal government. In this talk, I will describe the formalism associated with the platform, discuss its performance on benchmark datasets out-of-the-box and show a live demo for building a recommendation system and for detecting geo-spatial anomalies in maritime domain.

Abstract 8: **Poster Session (encompasses coffee break) in Nearest Neighbors for Modern Applications with Massive Data: An Age-old Solution with New Challenges**, *Chen, Balle, Lee, frosio, Malik, Kautz, Li, Sugiyama, Carreira-Perpinan, Raziperchikolaei, Tulabandhula, Noh, Yu* 02:50 PM

- Learning Supervised Binary Hashing without Binary Code Optimization (Miguel Carreira-Perpinan; Ramin Raziperchikolaei)
- Sub-linear Privacy-preserving Search with Unsecured Server and Semi-honest Parties (Beidi Chen)
- On Nearest Neighbors in Non Local Means Denoising (Iuri Frosio; Kautz Jan)
- Fast k-Nearest Neighbour Search via Prioritized DCI (Ke Li; Jitendra Malik)
- Generative Local Metric Learning for Nearest Neighbor Methods (Yung-Kyun Noh; Masashi Sugiyama; Daniel D Lee)
- Private Document Classification in Federated Databases (Phillipp Schoppmann; Adria Gascon; Borja Balle)
- Optimizing Revenue Over Data-Driven Assortments (Deeksha Sinha; Theja Tulabandula)
- Fast Distance Metric Learning for Multi-Task Large Margin Nearest Neighbor Classification (Adams Yu)

Abstract 9: **The Boundary Forest Algorithm in Nearest Neighbors for Modern Applications with Massive Data: An Age-old Solution with New Challenges**, *Yedidia* 04:15 PM

I explain the Boundary Forest algorithm, a simple, fast, and incremental online algorithm that can be used in both supervised and unsupervised settings, for classification, regression, or retrieval problems. As a classification algorithm, its generalization performance is at least as good as K-nearest-neighbors, while being able to respond to queries very quickly. It maintains trees of examples that let it train and respond to test queries in a time logarithmic with the number of stored examples, which is typically very much less than the number of training examples.

Abstract 10: **Iterative Collaborative Filtering for Sparse Matrix Estimation in Nearest Neighbors for Modern Applications with Massive Data: An Age-old Solution with New Challenges**, *Lee* 05:00 PM

The sparse matrix estimation problem consists of estimating the distribution of an n-by-n matrix Y , from a sparsely observed single instance of this matrix where the entries of Y are independent random variables. This captures a wide array of problems; special instances include matrix completion in the context of recommendation systems, graphon estimation, and community detection in (mixed membership) stochastic block models. Inspired by classical collaborative filtering for recommendation systems, we propose a novel iterative, collaborative filtering style algorithm for matrix estimation in this generic setting. Under model assumptions of uniform sampling, bounded entries, and finite spectrum, we provide bounds on the the mean squared error (MSE) of our estimator and show improved sample complexity.

Machine Deception

Ian Goodfellow, Tim Hwang, Bryce Goodman, Mikel Rodriguez

202, Fri Dec 08, 08:00 AM

Machine deception refers to the capacity for machine learning systems to manipulate human and machine agents into believing, acting upon or otherwise accepting false information. The development of machine deception has had a long, foundational and under-appreciated impact on shaping research in the field of artificial intelligence. Thought experiments such as Alan Turing's eponymous "Turing test" - where an automated system attempts to deceive a human judge into believing it is a human interlocutor, or Searle's "Chinese room" - in which a human operator attempts to imbue the false impression of consciousness in a machine, are simultaneously exemplars of machine deception and some of the most famous and influential concepts in the field of AI. As the field of machine learning advances, so too does machine deception seem poised to give rise to a host of practical opportunities and concerns. Machine deception can have many benign and beneficial applications. Chatbots designed to mimic human agents offer technical support and even provide therapy at a cost and scale that may not be otherwise achievable. On the other hand, the rise of techniques that leverage bots and other autonomous agents to manipulate and shape political speech online, has put machine deception in the political spotlight and raised fundamental questions regarding the ability to preserve truth in the digital domain. These concerns are amplified by recent demonstrations of machine learning techniques that synthesize hyper-realistic manipulations of audio and video. The proposed workshop will bring together research at the forefront of machine deception, including: Machine-machine deception: Where a machine agent deceives another machine agent, e.g. the use of "bot farms" that automate posting on social media platforms to manipulate content ranking algorithms or evolutionary networks to generate images that "fool" deep neural networks. Human-machine deception: Where a human agent deceives a machine agent, e.g. the use of human "troll farms" to manipulate content ranking algorithms or use of adversarial examples to exploit fragility of autonomous systems (e.g. stop sign sticker for self driving cars or printed eye-glasses for facial recognition). Machine-human deception: Where a machine agent is leveraged to deceives a human agent, e.g. the use of GANs to produce realistic manipulations of audio and video content. Although the workshop will primarily focus on the technical aspects of machine deception, submissions from the fields of law, policy, social sciences and psychology will also be encouraged. It is envisaged that this interdisciplinary forum will both shine a light on what is possible given state of the art tools today, and provide instructive guidance for both technologists and policy-makers going forward.

Schedule

10:00 AM	**Generating Natural Adversarial Examples**	*Zhao*
10:30 AM	**Adversarial Patch**	*Gilmer*
11:30 AM	**Interpretation of Neural Networks is Fragile**	*Abid*
12:00 PM	**Generative Models for Spear Phishing Posts on Social Media**	*Tully, Seymour*
01:30 PM	**CycleGAN, a Master of Steganography**	*Chu*
02:00 PM	**Machine Against Machine: Minimax-Optimal Attacks and Defenses**	*Hamm*
03:00 PM	**A3T: Adversarially Augmented Adversarial Training**	*Baratin, Lacoste-Julien, Bengio, Erraqabi*
03:30 PM	**Thermometer Encoding: One Hot way to resist Adversarial Examples**	

Discrete Structures in Machine Learning

Yaron Singer, Jeff A Bilmes, Andreas Krause, Stefanie Jegelka, Amin Karbasi

203, Fri Dec 08, 08:00 AM

Traditionally, machine learning has been focused on methods where objects reside in continuous domains. The goal of this workshop is to advance state-of-the-art methods in machine learning that involve discrete structures.

Models with ultimately discrete solutions play an important role in machine learning. At its core, statistical machine learning is concerned with making inferences from data, and when the underlying variables of the data are discrete, both the tasks of model inference as well as predictions using the inferred model are inherently discrete algorithmic problems. Many of these problems are notoriously hard, and even those that are theoretically tractable become intractable in practice with abundant and steadily increasing amounts of data. As a result, standard theoretical models and off-the-shelf algorithms become either impractical or intractable (and in some cases both).

While many problems are hard in the worst case, the problems of practical interest are often much more well-behaved, and have the potential to be modeled in ways that make them tractable. Indeed, many discrete problems in machine learning can possess beneficial structure; such structure has been an important ingredient in many successful (approximate) solution strategies. Examples include submodularity, marginal polytopes, symmetries and exchangeability.

Machine learning, algorithms, discrete mathematics and combinatorics 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.

Schedule

09:15 AM	**Andrea Montanari**	
10:00 AM	**Spotlight session I**	*Thaker, Korba, Lenail*
11:00 AM	**David Tse**	
12:00 PM	**Lunch**	
01:30 PM	**Bobby Kleinberg**	
02:45 PM	**Spotlight session III**	*Kazemi, Salehi*
03:00 PM	**Coffee Break and Posters**	
04:00 PM	**Nina Balcan**	
05:00 PM	**Contributed talk**	
05:15 PM	**Poster session**	

Transparent and interpretable Machine Learning in Safety Critical Environments

Alessandra Tosi, Alfredo Vellido, Mauricio A. Álvarez

204, Fri Dec 08, 08:00 AM

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

08:50 AM	**Opening remarks**	*Tosi, Vellido, Álvarez*
09:10 AM	**Invited talk: Is interpretability and explainability enough for safe and reliable decision making?**	*Saria*
09:45 AM	**Invited talk: The Role of Explanation in Holding AIs Accountable**	*Doshi-Velez*
10:20 AM	**Contributed talk: Beyond Sparsity: Tree-based Regularization of Deep Models for Interpretability**	*Wu, Parbhoo, Doshi-Velez*
10:30 AM	**Coffe break 1**	
11:00 AM	**Invited talk: Challenges for Transparency**	*Weller*
11:20 AM	**Contributed talk: Safe Policy Search with Gaussian Process Models**	*Polymenakos, Roberts*
11:30 AM	**Poster spotlights**	*Kuwajima, Tanaka, Liang, Komorowski, Que, Drumond, Raghu, Celi, Göpfert, Ross, Tan, Caruana, Lou, Kumar, Taylor, Poursabzi-Sangdeh, Wortman Vaughan, Wallach*
12:00 PM	**Poster session part I**	
12:30 PM	**Lunch break**	
02:00 PM	**Invited talk: When the classifier doesn't know: optimum reject options for classification.**	*Hammer*
02:35 PM	**Contributed talk: Predict Responsibly: Increasing Fairness by Learning To Defer Abstract**	*Madras, Zemel, Pitassi*
02:45 PM	**Contributed talk: Deep Motif Dashboard: Visualizing and Understanding Genomic Sequences Using Deep Neural Networks**	*Lanchantin, Singh, Wang*
02:55 PM	**Best paper prize announcement**	
03:00 PM	**Coffe break and Poster session part II**	
03:30 PM	**Invited talk: Robot Transparency as Optimal Control**	*Dragan*
04:05 PM	**Invited talk 6**	*Amodei*
04:40 PM	**Panel discussion**	
05:20 PM	**Final remarks**	*Tosi, Vellido, Álvarez*
05:30 PM	**End of workshop**	

Abstracts (8):

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.

Abstract 3: **Invited talk: The Role of Explanation in Holding AIs Accountable in Transparent and interpretable Machine Learning in Safety Critical Environments**, *Doshi-Velez* 09:45 AM

As AIs are used in more common and consequential situations, it is important that we find ways to take advantage of our computational capabilities while also holding the creators of these systems accountable. In this talk, I'll start out by sharing some of the challenges associated with deploying AIs in healthcare, and how interpretability or explanation is an essential tool in this domain. Then I'll speak more broadly about the role of explanation in holding AIs accountable under the law (especially in the context of current regulation around AIs). In doing so, I hope to spark discussions about how we, as a machine learning community, believe that our work should be regulated.

Abstract 4: **Contributed talk: Beyond Sparsity: Tree-based Regularization of Deep Models for Interpretability in Transparent and interpretable Machine Learning in Safety Critical Environments**, *Wu, Parbhoo, Doshi-Velez* 10:20 AM

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.

Abstract 7: **Contributed talk: Safe Policy Search with Gaussian Process Models in Transparent and interpretable Machine Learning in Safety Critical Environments**, *Polymenakos, Roberts* 11:20 AM

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.

Abstract 8: **Poster spotlights in Transparent and interpretable Machine Learning in Safety Critical Environments**, *Kuwajima, Tanaka, Liang, Komorowski, Que, Drumond, Raghu, Celi, Göpfert, Ross, Tan, Caruana, Lou, Kumar, Taylor, Poursabzi-Sangdeh, Wortman Vaughan, Wallach* 11:30 AM

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

Abstract 12: **Contributed talk: Predict Responsibly: Increasing Fairness by Learning To Defer Abstract in Transparent and interpretable Machine Learning in Safety Critical Environments**, *Madras, Zemel, Pitassi* 02:35 PM

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: **Contributed talk: 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

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.

Abstract 16: **Invited talk: Robot Transparency as Optimal Control in Transparent and interpretable Machine Learning in Safety Critical Environments**, *Dragan* 03:30 PM

In this talk, we will formalize transparency as acting in a dynamical system or MDP in which we augment the physical state with the human's belief about the robot. We will characterize the dynamics model in this MDP, and show that approximate solutions lead to cars that drive in a way that is easier to anticipate, robots that come up with instructive demonstrations of their task knowledge, manipulator arms that clarify their intent, and navigation robots that clarify their future task plans.

Lastly, we will briefly explore robots that express more interesting properties like the their level of confidence in their task, or the weight of an object they are carrying.

NIPS 2017 Time Series Workshop

Vitaly Kuznetsov, Oren Anava, Scott Yang, Azadeh Khaleghi

Grand Ballroom A, 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

09:00 AM	**Introduction to Time Series Workshop**	
09:15 AM	**Marco Cuturi: Soft-DTW, a differentiable loss for time series data**	
10:00 AM	**Learning theory and algorithms for shapelets and other local features. Daiki Suehiro, Kohei Hatano, Eiji Takimoto, Shuji Yamamoto, Kenichi Bannai and Akiko Takeda.**	
10:15 AM	**Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. Yaguang Li, Rose Yu, Cyrus Shahabi and Yan Liu.**	*Li*
10:30 AM	**Morning Coffee Break**	
11:00 AM	**Panel discussion featureing Marco Cuturi (ENSAE / CREST), Claire Monteleoni (GWU), Karthik Sridharan (Cornell), Firdaus Janoos (Two Sigma) and Matthias Seeger (Amazon)**	
11:45 AM	**Poster Session**	*Zand, Tu, Lee, Covert, Hernandez, Ebrahimzadeh, Slawinska, Supratak, Lu, Alberg, Shen, Yeo, Pao, Chai, Agarwal, Giannakis, Amjad*
12:30 PM	**Lunch**	
02:30 PM	**DISCOVERING ORDER IN UNORDERED DATASETS: GENERATIVE MARKOV NETWORKS. Yao-Hung Hubert Tsai, Han Zhao, Nebojsa Jojic and Ruslan Salakhutdinov.**	
02:45 PM	**Vitaly Kuznetsov: Kaggle web traffic time series forecasting competition: results and insights**	
03:30 PM	**Afternoon Coffee Break**	
04:00 PM	**Skip RNN: Learning to Skip State Updates in Recurrent Neural Networks. Víctor Campos, Brendan Jou, Xavier Giró-I-Nieto, Jordi Torres and Shih-Fu Chang.**	
04:15 PM	**Karthik Sridharan: Online learning, Probabilistic Inequalities and the Burkholder Method**	
05:00 PM	**Scalable Joint Models for Reliable Event Prediction. Hossein Soleimani, James Hensman and Suchi Saria.**	
05:15 PM	**Claire Monteleoni: Algorithms for Climate Informatics: Learning from spatiotemporal data with both spatial and temporal non-stationarity**	
06:00 PM	**An Efficient ADMM Algorithm for Structural Break Detection in Multivariate Time Series. Alex Tank, Emily Fox and Ali Shojaie.**	*Tank*
06:15 PM	**Conclusion and Awards**	

Abstracts (5):

Abstract 2: **Marco Cuturi: Soft-DTW, a differentiable loss for time series data in NIPS 2017 Time Series Workshop**, 09:15 AM

I will present in this talk a modification of the dynamic time warping distance which is, unlike the original quantity, differentiable in all of its inputs. As a result, that alternative distance can be used naturally as a learning loss to learn with datasets of time series, to produce means, clusters or structured prediction where the goal is to forecast entire time series.

Abstract 7: **Poster Session in NIPS 2017 Time Series Workshop**, *Zand, Tu, Lee, Covert, Hernandez, Ebrahimzadeh, Slawinska, Supratak, Lu, Alberg, Shen, Yeo, Pao, Chai, Agarwal, Giannakis, Amjad* 11:45 AM

Feel free to enjoy posters at lunch time as well!

Víctor Campos, Brendan Jou, Xavier Giró-I-Nieto, Jordi Torres and Shih-Fu Chang. Skip RNN: Learning to Skip State Updates in Recurrent Neural Networks.

Yao-Hung Hubert Tsai, Han Zhao, Nebojsa Jojic and Ruslan Salakhutdinov. DISCOVERING ORDER IN UNORDERED DATASETS: GENERATIVE MARKOV NETWORKS.

Yaguang Li, Rose Yu, Cyrus Shahabi and Yan Liu. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting.

Alex Tank, Emily Fox and Ali Shojaie. An Efficient ADMM Algorithm for Structural Break Detection in Multivariate Time Series.

Hossein Soleimani, James Hensman and Suchi Saria. Scalable Joint Models for Reliable Event Prediction.

Daiki Suehiro, Kohei Hatano, Eiji Takimoto, Shuji Yamamoto, Kenichi Bannai and Akiko Takeda. Learning theory and algorithms for shapelets and other local features.

Tao-Yi Lee, Yuh-Jye Lee, Hsing-Kuo Pao, You-Hua Lin and Yi-Ren Yeh. Elastic Motif Segmentation and Alignment of Time Series for Encoding and Classification.

Yun Jie Serene Yeo, Kian Ming A. Chai, Weiping Priscilla Fan, Si Hui Maureen Lee, Junxian Ong, Poh Ling Tan, Yu Li Lydia Law and Kok-Yong Seng. DP Mixture of Warped Correlated GPs for Individualized Time Series Prediction.

Anish Agarwal, Muhammad Amjad, Devavrat Shah and Dennis Shen. Time Series Forecasting = Matrix Estimation.

Rose Yu, Stephan Zheng, Anima Anandkumar and Yisong Yue. Long-term Forecasting using Tensor-Train RNNs.

Pranamesh Chakraborty, Chinmay Hegde and Anuj Sharma. Trend Filtering in Network Time Series with Applications to Traffic Incident Detection.

Jaleh Zand and Stephen Roberts. MiDGaP: Mixture Density Gaussian Processes.

Dimitrios Giannakis, Joanna Slawinska, Abbas Ourmazd and Zhizhen Zhao. Vector-Valued Spectral Analysis of Space-Time Data.

Ruofeng Wen, Kari Torkkola and Balakrishnan Narayanaswamy. A Multi-Horizon Quantile Recurrent Forecaster.

Alessandro Davide Ialongo, Mark van der Wilk and Carl Edward Rasmussen. Closed-form Inference and Prediction in Gaussian Process State-Space Models.

Hao Liu, Haoli Bai, Lirong He and Zenglin Xu. Structured Inference for Recurrent Hidden Semi-markov Model.

Petar Veli■kovi■, Laurynas Karazija, Nicholas Lane, Sourav Bhattacharya, Edgar Liberis, Pietro Lio, Angela Chieh, Otmane Bellahsen and Matthieu Vegreville. Cross-modal Recurrent Models for Weight Objective Prediction from Multimodal Time-series Data.

Kun Tu, Bruno Ribeiro, Ananthram Swami and Don Towsley. Temporal Clustering in time-varying Networks with Time Series Analysis.

Shaojie Bai, J. Zico Kolter and Vladlen Koltun. Convolutional Sequence Modeling Revisited.

Apurv Shukla, Se-Young Yun and Daniel Bienstock. Non-Stationary Streaming PCA.

Kun Zhao, Takayuki Osogami and Rudy Raymond. Fluid simulation with dynamic Boltzmann machine in batch manner.

Anderson Zhang, Miao Lu, Deguang Kong and Jimmy Yang. Bayesian Time Series Forecasting with Change Point and Anomaly Detection.

Akara Supratak, Steffen Schneider, Hao Dong, Ling Li and Yike Guo. Towards Desynchronization Detection in Biosignals.

Rudy Raymond, Takayuki Osogami and Sakyasingha Dasgupta. Dynamic Boltzmann Machines for Second Order Moments and Generalized Gaussian Distributions.

Itamar Ben-Ari and Ravid Shwartz-Ziv. Sequence modeling using a memory controller extension for LSTM.

Neil Dhir and Adam Kosiorek. Bayesian delay embeddings for dynamical systems.

Aleksander Wieczorek and Volker Roth. Time Series Classification with Causal Compression.

Daniel Hernandez, Liam Paninski and John Cunningham. Variational inference for latent nonlinear dynamics.

Alex Tank, Ian Covert, Nick Foti, Ali Shojaie and Emily Fox. An Interpretable and Sparse Neural Network Model for Nonlinear Granger Causality Discovery.

John Alberg and Zachary Lipton. Improving Factor-Based Quantitative Investing by Forecasting Company Fundamentals.

Achintya Kr. Sarkar and Zheng-Hua Tan. Time-Contrastive Learning

Based DNN Bottleneck Features for Text-Dependent Speaker Verification.

Ankit Gandhi, Vineet Chaoji and Arijit Biswas. Modeling Customer Time Series for Age Prediction.

Zahra Ebrahimzadeh and Samantha Kleinberg. Multi-Scale Change Point Detection in Multivariate Time Series.

Abstract 8: **Lunch in NIPS 2017 Time Series Workshop**, 12:30 PM

Lunch on your own

Abstract 13: **Karthik Sridharan: Online learning, Probabilistic Inequalities and the Burkholder Method in NIPS 2017 Time Series Workshop**, 04:15 PM

Online learning is a framework that makes minimal assumptions about the sequence of instances provided to a learner. This makes online learning an excellent framework for dealing with sequences of instances that vary with time. In this talk, we will look at inherent connections between online learning, certain Probabilistic Inequalities and the so called Burkholder Method. We will see how one can derive new, optimal, adaptive online learning algorithms using the Burkholder Method via the connection with Probabilistic Inequalities. We will use this insight to help us get a step closer to what I shall term Plug-&-Play ML. That is, help us move a step towards building machine learning systems automatically.

Abstract 15: **Claire Monteleoni: Algorithms for Climate Informatics: Learning from spatiotemporal data with both spatial and temporal non-stationarity in NIPS 2017 Time Series Workshop**, 05:15 PM

Climate Informatics is emerging as a compelling application of machine learning. This is due in part to the urgent nature of climate change, and its many remaining uncertainties (e.g. how will a changing climate affect severe storms and other extreme weather events?). Meanwhile, progress in climate informatics is made possible in part by the public availability of vast amounts of data, both simulated by large-scale physics-based models, and observed. Not only are time series at the crux of the study of climate science, but also, by definition, climate change implies non-stationarity. In addition, much of the relevant data is spatiotemporal, and also varies over location. In this talk, I will discuss our work on learning in the presence of spatial and temporal non-stationarity, and exploiting local dependencies in time and space. Along the way, I will highlight open problems in which machine learning, including deep learning methods, may prove fruitful.

Conversational AI - today's practice and tomorrow's potential

Alborz Geramifard, Jason Williams, Larry Heck, Jim Glass, Antoine Bordes, Steve Young, Gerald Tesauro

Grand Ballroom B, 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

09:00 AM	**Opening (Alborz)**	*Geramifard*
09:10 AM	**Invited Talk - Satindar Singh**	*Singh*
09:50 AM	**An Ensemble Model with Ranking for Social Dialogue**	
10:10 AM	**A Deep Reinforcement Learning Chatbot**	
10:30 AM	**Coffee Break / Hangup Posters**	
11:00 AM	**Invited Talk - Joelle Pineau**	*Pineau*
11:40 AM	**Multi-Domain Adversarial Learning for Slot Filling in Spoken Language Understanding**	
12:00 PM	**Lunch Break**	
01:30 PM	**Invited Talk: Amazon (Ashwin)**	*Ram*
02:10 PM	**Invited Talk - Apple Talk (Blaise)**	*Thomson*
02:50 PM	**One Minute Poster Spotlight**	*Zhao, khatri, Joshi, Fazel-Zarandi, Gupta, Li, Cercas Curry, Hedayatnia, Zhou, Venkatesh, Mi, Cheng*
03:05 PM	**Coffee Break / Poster Session**	
03:30 PM	**Invited Talk: Microsoft (Asli and Jianfeng)**	*Gao*
04:10 PM	**Invited Talk: Google (Matt and Dilek)**	*Hakkani-Tur*

04:50 PM	**Poster Session**	
05:20 PM	**Invited Talk: IBM (David, Rudolf)**	*Kadlec, Nahamoo*
06:00 PM	**Panel by organizers**	
06:50 PM	**Closing Remarks (Jason)**	*Williams*

Abstracts (1):

Abstract 17: **Panel by organizers in Conversational AI - today's practice and tomorrow's potential**, 06:00 PM

Q&A

OPT 2017: Optimization for Machine Learning

Suvrit Sra, Sashank J. Reddi, Alekh Agarwal, Benjamin Recht

Hall A, Fri Dec 08, 08:00 AM

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 (Vandenberghe), 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

08:50 AM	**Opening Remarks**	
09:00 AM	**Poster Session**	*Lau, Maly, Loizou, Kroer, Yao, Park, Kovacs, Yin, Zhukov, Lim, Barmherzig, Metaxas, Shi, Udwani, Brendel, Zhou, braverman, Liu, Golikov*
09:00 AM	**Invited Talk: Leon Bottou**	
09:45 AM	**Invited Talk: Yurii Nesterov**	
10:30 AM	**Coffee Break 1**	
11:00 AM	**Spotlight: Oracle Complexity of Second-Order Methods for Smooth Convex Optimization**	
11:15 AM	**Spotlight: Gradient Diversity: a Key Ingredient for Scalable Distributed Learning**	
11:30 AM	**Invited Talk: Francis Bach**	
12:15 PM	**Lunch Break**	
02:00 PM	**Invited Talk: Dimitri Bertsekas**	
02:45 PM	**Spotlight: Lower Bounds for Finding Stationary Points of Non-Convex, Smooth High-Dimensional Functions**	
03:00 PM	**Coffee Break 2**	
03:30 PM	**Invited Talk : Pablo Parrilo**	

04:15 PM	**Spotlight: Efficiently Optimizing over (Non-Convex) Cones via Approximate Projections**
04:30 PM	**Poster Session II**

Abstracts (1):

Abstract 1: **Opening Remarks in OPT 2017: Optimization for Machine Learning**, 08:50 AM

Opening Remarks by the Organizers.

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

08:30 AM	**Introductions**	*Toulis, Volfovsky*
08:45 AM	**Looking for a Missing Signal**	*Bottou*
09:20 AM	**Invited Talk**	*Brunskill*
10:00 AM	**Contributed Talk 1**	*Heckerman*
10:15 AM	**Contributed Talk 2**	*Mitrovic*
11:00 AM	**Invited Talk 3**	*Hahn*
11:35 AM	**Invited Talk 4**	*Sontag*
12:10 PM	**Poster session**	*Zaidi, Kurz, Heckerman, Lin, Riezler, Shpitser, Yan, Goudet, Deshpande, Pearl, Mitrovic, Vegetabile, Lee, Sachs, Mohan, Rose, Ramakers, Hassanpour, Baldi, Nabi, Hammarlund, Sherman, Lawrence, Jabbari, Semenova, Dimakopoulou, Gajane, Greiner, Zadik, Blocker, Xu, EL HAY, Jebara, Rostykus*
01:35 PM	**Contributed Talk 3**	*Shpitser*
01:50 PM	**Contributed Talk 4**	*Pearl*
02:05 PM	**Invited Talk 5**	*imbens*
03:30 PM	**Invited Talk**	*Yu*
04:05 PM	**Causal inference with machine learning**	
05:00 PM	**Causality and Machine Learning Challenge: Criteo Ad Placement Challenge**	

Abstracts (1):

Abstract 2: **Looking for a Missing Signal in From 'What If?' To 'What Next?' : Causal Inference and Machine Learning for Intelligent Decision Making**, *Bottou* 08:45 AM

We know how to spot object in images, but we must learn on more images than a human can see in a lifetime. We know how to translate text (somehow), but we must learn it on more text than a human can read in a lifetime. We know how to learn playing Atari games, but we must learn it by playing more games than any teenager can endure. The list is long. We can of course try to pin this inefficiently to some properties of our algorithms. However, we can also take the point of view that there is possibly a lot of signal in natural data that we simply do not exploit. I will report on two works in this direction. The first one establishes that something as simple as a collection of static images contains nontrivial information about the causal relations between the objects they represent. The second one, time permitting, shows how an attempt to discover such a structure in observational data led to a clear improvement of Generative Adversarial Networks.

Extreme Classification: Multi-class & Multi-label Learning in Extremely Large Label Spaces

Manik Varma, Marius Kloft, Krzysztof Dembczynski

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

Extreme classification is a rapidly growing research area focussing 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

09:00 AM	**Introduction by Manik Varma**	*Varma*
09:05 AM	**John Langford (MSR) on Dreaming Contextual Memory**	*Langford*
09:35 AM	**Ed Chi (Google) on Learned Deep Retrieval for Recommenders**	*Chi*
10:05 AM	**David Sontag (MIT) on Representation Learning for Extreme Multi-class Classification & Density Estimation**	
10:35 AM	**Coffee Break**	
11:00 AM	**Inderjit Dhillon (UT Austin & Amazon) on Stabilizing Gradients for Deep Neural Networks with Applications to Extreme Classification**	
11:30 AM	**Wei-cheng Chang (CMU) on Deep Learning Approach for Extreme Multi-label Text Classification**	
12:00 PM	**Lunch**	
01:30 PM	**Pradeep Ravikumar (CMU) on A Parallel Primal-Dual Sparse Method for Extreme Classification**	*Ravikumar*
02:00 PM	**Maxim Grechkin (UW) on EZLearn: Exploiting Organic Supervision in Large-Scale Data Annotation**	*Grechkin*
02:15 PM	**Sayantan Dasgupta (Michigan) on Multi-label Learning for Large Text Corpora using Latent Variable Model**	*Dasgupta*
02:30 PM	**Yukihiro Tagami (Yahoo) on Extreme Multi-label Learning via Nearest Neighbor Graph Partitioning and Embedding**	*Tagami*
03:00 PM	**Coffee Break**	
03:15 PM	**Mehryar Mohri (NYU) on Tight Learning Bounds for Multi-Class Classification**	*Mohri*

03:45 PM	**Ravi Ganti (Walmart Labs) on Exploiting Structure in Large Scale Bandit Problems**	*Ganti*
04:00 PM	**Hai S Le (WUSTL) on Precision-Recall versus Accuracy and the Role of Large Data Sets**	*Le*
04:15 PM	**Loubna Benabbou (EMI) on A Reduction Principle for Generalizing Bona Fide Risk Bounds in Multi-class Seeting**	*BENABBOU*
04:30 PM	**Marius Kloft (Kaiserslautern) on Generalization Error Bounds for Extreme Multi-class Classification**	*Kloft*

Learning on Distributions, Functions, Graphs and Groups

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

Hyatt Hotel, Regency Ballroom D+E+F+H, 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 (Flaxman 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:

Frédéric Ferraty and Philippe Vieu. Nonparametric Functional Data Analysis: Theory and Practice. Springer Series in Statistics, Springer-Verlag, 2006.

Jane-Ling Wang, Jeng-Min Chiou, and Hans-Georg Müller. Review of Functional Data Analysis. Annual Review of Statistics, 3:1-41, 2015.

Barnabás Póczos, Aarti Singh, Alessandro Rinaldo, Larry Wasserman. Distribution-free Distribution Regression. International Conference on AI and Statistics (AISTATS), PMLR 31:507-515, 2013.

Gunnar Carlsson. Topology and data. Bulletin of the American Mathematical Society, 46 (2): 255-308, 2009.

Vitaliy Kurlin. Research blog: http://kurlin.org/blog/.

Jiayu Zhou, Jun Liu, Vaibhav A. Narayan, and Jieping Ye. Modeling disease progression via multi-task learning. NeuroImage, 78:233-248, 2013.

Xu Sun, Hisashi Kashima, and Naonori Ueda. Large-scale personalized human activity recognition using online multitask learning. IEEE Transactions on Knowledge and Data Engine, 25:2551-2563, 2013.

David Lopez-Paz, Krikamol Muandet, Bernhard Schölkopf, and Ilya Tolstikhin. Towards a Learning Theory of Cause-Effect Inference. International Conference on Machine Learning (ICML), PMLR 37:1452-1461, 2015.

Risi Kondor, Horace Pan. The Multiscale Laplacian Graph Kernel. Advances in Neural Information Processing Systems (NIPS), 2982-2990, 2016.

Genki Kusano, Yasuaki Hiraoka, Kenji Fukumizu. Persistence weighted Gaussian kernel for topological data analysis. International Conference on Machine Learning (ICML), PMLR 48:2004-2013, 2016.

Siamak Ravanbakhsh, Junier Oliva, Sebastian Fromenteau, Layne Price, Shirley Ho, Jeff Schneider, Barnabás Póczos. Estimating Cosmological

Parameters from the Dark Matter Distribution. International Conference on Machine Learning (ICML), PMLR 48:2407-2416, 2016.

Ho Chung Leon Law, Dougal J. Sutherland, Dino Sejdinovic, Seth Flaxman. Bayesian Distribution Regression. Technical Report, 2017 (https://arxiv.org/abs/1705.04293).

Hachem Kadri, Emmanuel Duflos, Philippe Preux, Stéphane Canu, Alain Rakotomamonjy, and Julien Audiffren. Operator-valued kernels for learning from functional response data. Journal of Machine Learning Research, 17:1-54, 2016.

Céline Brouard, Marie Szafranski, and Florence d'Alché-Buc. Input output kernel regression: Supervised and semi-supervised structured output prediction with operator-valued kernels. Journal of Machine Learning Research, 17:1-48, 2016.

Zoltán Szabó, Bharath K. Sriperumbudur, Barnabás Póczos, Arthur Gretton. Learning Theory for Distribution Regression. Journal of Machine Learning Research, 17(152):1-40, 2016.

Seth Flaxman, Yu-Xiang Wang, and Alex Smola. Who supported Obama in 2012? Ecological inference through distribution regression. In ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 289-298, 2015.

Schedule

09:00 AM	**On Structured Prediction Theory with Calibrated Convex Surrogate Losses.**	*Lacoste-Julien*
09:30 AM	**Differentially Private Database Release via Kernel Mean Embeddings.**	
09:50 AM	**Bayesian Distribution Regression.**	
10:10 AM	**Differentially Private Database Release via Kernel Mean Embeddings (poster).**	
10:10 AM	**Post Selection Inference with Maximum Mean Discrepancy (poster).**	
10:10 AM	**The Weighted Kendall Kernel (poster).**	
10:10 AM	**When is Network Lasso Accurate: The Vector Case (poster).**	
10:10 AM	**Algorithmic and Statistical Aspects of Linear Regression without Correspondence (poster).**	
10:10 AM	**Graph based Feature Selection for Structured High Dimensional Data (poster).**	*Zhang*
10:10 AM	**Convolutional Layers based on Directed Multi-Graphs (poster).**	*Arodz*
10:10 AM	**Kernels on Fuzzy Sets: an Overview (poster).**	*Guevara Diaz*
10:10 AM	**The Geometric Block Model (poster).**	*Pal*
10:10 AM	**Worst-case vs. Average-case Design for Estimation from Fixed Pairwise Comparisons (poster).**	
10:10 AM	**Squared Earth Mover's Distance Loss for Training Deep Neural Networks on Ordered-Classes (poster).**	*Hou*
10:10 AM	**Learning from Conditional Distributions via Dual Embeddings (poster).**	*Song*
10:10 AM	**Large Scale Graph Learning from Smooth Signals (poster).**	
10:10 AM	**Learning from Graphs with Structural Variation (poster).**	*Holm, Nielsen*
10:10 AM	**Bayesian Distribution Regression (poster).**	
10:10 AM	**On Kernel Methods for Covariates that are Rankings (poster).**	
11:00 AM	**When is Network Lasso Accurate: The Vector Case.**	*Tran*
11:20 AM	**Worst-case vs. Average-case Design for Estimation from Fixed Pairwise Comparisons.**	
11:40 AM	**The Weighted Kendall Kernel.**	
12:00 PM	**On Kernel Methods for Covariates that are Rankings.**	
12:20 PM	**Lunch Break**	
01:50 PM	**Learning on topological and geometrical structures of data.**	*Fukumizu*
02:20 PM	**Operator-valued kernels and their application to functional data analysis.**	*Kadri*
02:50 PM	**Poster Session II & Coffee**	
03:50 PM	**Distribution Regression and its Applications.**	*Poczos*

04:20 PM	**Covariant Compositional Networks for Learning Graphs**	*Kondor*

Abstracts (27):

Abstract 1: **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 2: **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 3: **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 4: **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 5: **Post Selection Inference with Maximum Mean Discrepancy (poster). in Learning on Distributions, Functions, Graphs and Groups**, 10:10 AM

Authors: Denny Wu, Makoto Yamada, Ichiro Takeuchi, Kenji Fukumizu. 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: **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 8: **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 9: **Graph based Feature Selection for Structured High Dimensional Data (poster). in Learning on Distributions, Functions, Graphs and Groups**, *Zhang* 10:10 AM

Authors: Thosini K. Bamunu Mudiyanselage, Yanqing Zhang. Poster session continues at 14:50 - 15:50.

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

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

Abstract 11: **Kernels on Fuzzy Sets: an Overview (poster). in Learning on Distributions, Functions, Graphs and Groups**, *Guevara Diaz* 10:10 AM

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

Abstract 12: **The Geometric Block Model (poster). in Learning on Distributions, Functions, Graphs and Groups**, *Pal* 10:10 AM

Authors: Sainyam Galhotra, Arya Mazumdar, Soumyabrata Pal, Barna Saha. Poster session continues at 14:50 - 15:50.

Abstract 13: **Worst-case vs. Average-case Design for Estimation from Fixed Pairwise Comparisons (poster). in Learning on Distributions, Functions, Graphs and Groups**, 10:10 AM

Authors: Ashwin Pananjady, Cheng Mao, Vidya Muthukumar, Martin Wainwright, Thomas Courtade. Poster session continues at 14:50 - 15:50.

Abstract 14: **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 15: **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 16: **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 17: **Learning from Graphs with Structural Variation (poster). in Learning on Distributions, Functions, Graphs and Groups**, *Holm, Nielsen* 10:10 AM

Authors: Rune K. Nielsen, Aasa Feragen, Andreas Holm. Poster session continues at 14:50 - 15:50.

Abstract 18: **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 19: **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 20: **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 21: **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 22: **The Weighted Kendall Kernel. in Learning on Distributions, Functions, Graphs and Groups**, 11:40 AM

Authors: Yunlong Jiao, Jean-Philippe Vert.

Abstract 23: **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 persistence 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 26: **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'll 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 28: **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 29: **Covariant Compositional Networks for Learning Graphs in Learning on Distributions, Functions, Graphs and Groups**, *Kondor* 04:20 PM

Most existing neural networks for learning graphs deal with the issue of permutation invariance by conceiving of the network as a message passing scheme, where each node sums the feature vectors coming from its neighbors. We argue that this imposes a limitation on their representation power, and instead propose a new general architecture for representing objects consisting of a hierarchy of parts, which we call covariant compositional networks (CCNs). Here covariance means that the activation of each neuron must transform in a specific way under permutations, similarly to steerability in CNNs. We achieve covariance by making each activation transform according to a tensor representation of the permutation group, and derive the corresponding tensor aggregation rules that each neuron must implement. Experiments show that CCNs can outperform competing methods on some standard graph learning benchmarks.

Machine Learning for Creativity and Design

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

Hyatt Hotel, Seaview Ballroom, 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 (Deep Tingle). 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 nacesent 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

08:30 AM	**Welcome and Introduction**	
08:45 AM	**Invited Talk**	*Schmidhuber*
09:15 AM	**Invited Talk**	*Denton*
09:45 AM	**Invited Talk**	*Fiebrink*

10:15 AM	**GANosaic - Mosaic Creation with Generative Texture Manifolds**	*Jetchev, Bergmann, Seward*
10:20 AM	**TopoSketch: Drawing in Latent Space**	
10:25 AM	**Input parameterization for DeepDream**	
11:00 AM	**Invited Talk**	*Goodfellow*
11:30 AM	**Improvised Comedy as a Turing Test**	
12:00 PM	**Lunch**	
01:00 PM	**Invited Talk**	*Elgammal*
01:30 PM	**Hierarchical Variational Autoencoders for Music**	
02:00 PM	**Lexical preferences in an automated story writing system**	
02:30 PM	**ObamaNet: Photo-realistic lip-sync from text**	*Kumar, Sotelo, Kumar, de Brébisson*
03:00 PM	**Art / Coffee Break**	
03:30 PM	**Towards the High-quality Anime Characters Generation with Generative Adversarial Networks**	
03:35 PM	**Crowd Sourcing Clothes Design Directed by Adversarial Neural Networks**	*Osone, Kato, Sato, Muramatsu*
03:40 PM	**Paper Cubes: Evolving 3D characters in Augmented Reality using Recurrent Neural Networks**	*Fuste, Jongejan*
03:45 PM	**Open discussion**	
04:15 PM	**Generating Black Metal and Math Rock: Beyond Bach, Beethoven, and Beatles**	
04:15 PM	**Algorithmic composition of polyphonic music with the WaveCRF**	
04:15 PM	**SOMNIA: Self-Organizing Maps as Neural Interactive Art**	
04:15 PM	**Combinatorial Meta Search**	
04:15 PM	**Compositional Pattern Producing GAN**	
04:15 PM	**Generative Embedded Mapping Systems for Design**	
04:15 PM	**Exploring Audio Style Transfer**	
04:15 PM	**Imaginary Soundscape : Cross-Modal Approach to Generate Pseudo Sound Environments**	*Kajihara, Tokui*
04:15 PM	**Repeating and Remembering: GANs in an art context**	*Ridler*
04:15 PM	**Improvisational Storytelling Agents**	
04:15 PM	**Learning to Create Piano Performances**	
04:15 PM	**Neural Style Transfer for Audio Spectograms**	*Verma, Smith*
04:15 PM	**SocialML: machine learning for social media video creators**	
04:15 PM	**Artwork**	*Ambrosi, Erler, Salavon, Reimann-Dubbers, Barrat*
04:15 PM	**Deep Interactive Evolutionary Computation**	*Bontrager*
04:15 PM	**ASCII Art Synthesis with Convolutional Networks**	
04:15 PM	**Disentangled representations of style and content for visual art with generative adversarial networks**	
04:15 PM	**Sequential Line Search for Generative Adversarial Networks**	
04:15 PM	**AI for Fragrance Design**	*Segal*
04:15 PM	**Consistent Comic Colorization with Pixel-wise Background Classification**	*Choo, Kang*
04:15 PM	**Deep Learning for Identifying Potential Conceptual Shifts for Co-creative Drawing**	*Karimi*
04:15 PM	**The Emotional GAN: Priming Adversarial Generation of Art with Emotion**	*Amores Fernandez*

Machine Learning and Computer Security

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

Hyatt Hotel, Shoreline, 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 clear formal definitions of security that are appropriate for ML systems. At the same time, security of ML systems is of great import due 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

09:00 AM	**Opening Remarks**	*Song*
09:15 AM	**AI Applications in Security at Ant Financial**	*Qi*
09:45 AM	**A Word Graph Approach for Dictionary Detection and Extraction in DGA Domain Names**	*Pereira*
10:00 AM	**Practical Machine Learning for Cloud Intrusion Detection**	*Siva Kumar*
10:15 AM	**Poster Spotlights I**	*Na, Song, Sinha, Shin, Huang, Narodytska, Staib, Pei, Suya, Ghorbani, Buckman, Hein, Zhang, Qi, Tian, Du, Tsipras*
11:00 AM	**International Security and the AI Revolution**	*Dafoe*
11:30 AM	**BadNets: Identifying Vulnerabilities in the Machine Learning Model Supply Chain**	*Garg*
11:45 AM	**Poster Spotlights II**	
01:30 PM	**Defending Against Adversarial Examples**	*Goodfellow*
02:00 PM	**Provable defenses against adversarial examples via the convex outer adversarial polytope**	*Kolter*
02:15 PM	**Games People Play (With Bots)**	*Brinkman*
02:45 PM	**Synthesizing Robust Adversarial Examples**	*Ilyas, Athalye, Engstrom, Kwok*
03:00 PM	**Poster Session**	
03:45 PM	**Safety beyond Security: Societal Challenges for Machine Learning**	*Hardt*
04:15 PM	**Towards Verification of Deep Neural Networks**	*Barrett*

ML Systems Workshop @ NIPS 2017

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

S1, 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

08:45 AM	**Opening Remarks**	
09:00 AM	**Invited Talk: Ray: A distributed execution engine for emerging AI applications, Ion Stoica, UC Berkeley**	*Stoica*
09:20 AM	**Contributed Talk 1: The Case for Learning Database Indexes**	
09:40 AM	**Invited Talk: Federated Multi-Task Learning, Virginia Smith, Stanford University**	*Smith*
10:00 AM	**Poster Previews: 1 min lightning talks**	

11:30 AM	**Invited Talk: Accelerating Persistent Neural Networks at Datacenter Scale, Daniel Lo, Microsoft Research**	*Lo*
11:50 AM	**Contributed Talk 2: DLVM: A modern compiler framework for neural network DSLs**	
12:10 PM	**Lunch**	
01:20 PM	**Updates from Current ML Systems (TensorFlow, PyTorch, Caffe2, CNTK, MXNet, TVM, Clipper, MacroBase, ModelDB)**	*Monga, Chintala, Zhang, Chen, Crankshaw, Tai, Tulloch, Vartak*
02:50 PM	**Invited Talk: Machine Learning for Systems and Systems for Machine Learning, Jeff Dean, Google Brain**	*Dean*
03:20 PM	**Invited Talk: Creating an Open and Flexible ecosystem for AI models with ONNX, Sarah Bird, Dmytro Dzhulgakov, Facebook Research**	*Bird*
03:40 PM	**Posters and Coffee**	*Tristan, Lee, Dorogush, Hido, Terry, Siam, Nakada, Coleman, Ha, Zhang, Stooke, Meng, Kappler, Schwartz, Olston, Schelter, Sun, Kang, Hummer, Chung, Kraska, Ramchandran, Hynes, Boden, Kwak*
04:30 PM	**Contributed Talk 3: NSML: A Machine Learning Platform That Enables You to Focus on Your Models**	
04:50 PM	**Contributed Talk 4: DAWNBench: An End-to-End Deep Learning Benchmark and Competition**	
05:10 PM	**Panel**	*Gibson, Gonzalez, Langford, Song*

Machine Learning for Molecules and Materials

Stefan Chmiela, Jose Miguel Hernández-Lobato, Kristof T. Schütt, Alan Aspuru-Guzik, Alexandre Tkatchenko, Bharath Ramsundar, Anatole von Lilienfeld, Matt Kusner, Koji Tsuda, Brooks Paige, Klaus-Robert Müller

S4, Fri Dec 08, 08:00 AM

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
[1] Behler, J., Lorenz, S., Reuter, K. (2007). Representing molecule-surface interactions with symmetry-adapted neural networks. J. Chem. Phys., 127(1), 07B603.
[2] Behler, J., Parrinello, M. (2007). Generalized neural-network representation of high-dimensional potential-energy surfaces. Phys. Rev. Lett., 98(14), 146401.
[3] Kang, B., Ceder, G. (2009). Battery materials for ultrafast charging and discharging. Nature, 458(7235), 190.

[4] Bartók, A. P., Payne, M. C., Kondor, R., Csányi, G. (2010). Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. Phys. Rev. Lett., 104(13), 136403.
[5] Behler, J. (2011). Atom-centered symmetry functions for constructing high-dimensional neural network potentials. J. Chem. Phys, 134(7), 074106.
[6] Behler, J. (2011). Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations. Phys. Chem. Chem. Phys., 13(40), 17930-17955.
[7] Rupp, M., Tkatchenko, A., Müller, K.-R., von Lilienfeld, O. A. (2012). Fast and accurate modeling of molecular atomization energies with machine learning. Phys. Rev. Lett., 108(5), 058301.
[8] Snyder, J. C., Rupp, M., Hansen, K., Müller, K.-R., Burke, K. (2012). Finding density functionals with machine learning. Phys. Rev. Lett., 108(25), 253002.
[9] Montavon, G., Rupp, M., Gobre, V., Vazquez-Mayagoitia, A., Hansen, K., Tkatchenko, A., Müller, K.-R., von Lilienfeld, O. A. (2013). Machine learning of molecular electronic properties in chemical compound space. New J. Phys., 15(9), 095003.
[10] Hansen, K., Montavon, G., Biegler, F., Fazli, S., Rupp, M., Scheffler, M., Tkatchenko, A., Müller, K.-R. (2013). Assessment and validation of machine learning methods for predicting molecular atomization energies. J. Chem. Theory Comput., 9(8), 3404-3419.
[11] Bartók, A. P., Kondor, R., Csányi, G. (2013). On representing chemical environments. Phys. Rev. B, 87(18), 184115.
[12] Schütt K. T., Glawe, H., Brockherde F., Sanna A., Müller K.-R., Gross E. K. U. (2014). How to represent crystal structures for machine learning: towards fast prediction of electronic properties. Phys. Rev. B., 89(20), 205118.
[13] Ramsundar, B., Kearnes, S., Riley, P., Webster, D., Konerding, D., Pande, V. (2015). Massively multitask networks for drug discovery. arXiv preprint arXiv:1502.02072.
[14] Rupp, M., Ramakrishnan, R., & von Lilienfeld, O. A. (2015). Machine learning for quantum mechanical properties of atoms in molecules. J. Phys. Chem. Lett., 6(16), 3309-3313.
[15] V. Botu, R. Ramprasad (2015). Learning scheme to predict atomic forces and accelerate materials simulations., Phys. Rev. B, 92(9), 094306.
[16] Hansen, K., Biegler, F., Ramakrishnan, R., Pronobis, W., von Lilienfeld, O. A., Müller, K.-R., Tkatchenko, A. (2015). Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space. J. Phys. Chem. Lett, 6(12), 2326-2331.
[17] Alipanahi, B., Delong, A., Weirauch, M. T., Frey, B. J. (2015). Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning. ■Nat. Biotechnol., 33(8), 831-838.
[18] Duvenaud, D. K., Maclaurin, D., Aguilera-Iparraguirre, J., Gomez-Bombarelli, R., Hirzel, T., Aspuru-Guzik, A., Adams, R. P. (2015). Convolutional networks on graphs for learning molecular fingerprints. NIPS, 2224-2232.
[19] Faber F. A., Lindmaa A., von Lilienfeld, O. A., Armiento, R. (2016). Machine learning energies of 2 million elpasolite (A B C 2 D 6) crystals. Phys. Rev. Lett., 117(13), 135502.
[20] Gomez-Bombarelli, R., Duvenaud, D., Hernandez-Lobato, J. M., Aguilera-Iparraguirre, J., Hirzel, T. D., Adams, R. P., Aspuru-Guzik, A. (2016). Automatic chemical design using a data-driven continuous representation of molecules. arXiv preprint arXiv:1610.02415.
[21] Wei, J. N., Duvenaud, D, Aspuru-Guzik, A. (2016). Neural networks for the prediction of organic chemistry reactions. ACS Cent. Sci., 2(10), 725-732.
[22] Sadowski, P., Fooshee, D., Subrahmanya, N., Baldi, P. (2016). Synergies between quantum mechanics and machine learning in reaction prediction. J. Chem. Inf. Model., 56(11), 2125-2128.
[23] Lee, A. A., Brenner, M. P., Colwell L. J. (2016). Predicting protein-ligand affinity with a random matrix framework. Proc. Natl. Acad. Sci., 113(48), 13564-13569.
[24] Behler, J. (2016). Perspective: Machine learning potentials for atomistic simulations. J. Chem. Phys., 145(17), 170901.
[25] De, S., Bartók, A. P., Csányi, G., Ceriotti, M. (2016). Comparing molecules and solids across structural and alchemical space. Phys. Chem. Chem. Phys., 18(20), 13754-13769.
[26] Schütt, K. T., Arbabzadah, F., Chmiela, S., Müller, K.-R., Tkatchenko, A. (2017). Quantum-chemical insights from deep tensor neural networks. Nat. Commun., 8, 13890.
[27] Segler, M. H., Waller, M. P. (2017). Neural■symbolic machine learning for retrosynthesis and reaction prediction. ■Chem. Eur. J., 23(25), 5966-5971.
[28] Kusner, M. J., Paige, B., Hernández-Lobato, J. M. (2017). Grammar variational autoencoder. arXiv preprint arXiv:1703.01925.
[29] Coley, C. W., Barzilay, R., Jaakkola, T. S., Green, W. H., Jensen K. F. (2017). Prediction of organic reaction outcomes using machine learning. ACS Cent. Sci., 3(5), 434-443.
[30] Altae-Tran, H., Ramsundar, B., Pappu, A. S., Pande, V. (2017). Low data drug discovery with one-shot learning. ACS Cent. Sci., 3(4), 283-293.
[31] Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., Dahl, G. E. (2017). Neural message passing for quantum chemistry. arXiv preprint arXiv:1704.01212.
[32] Chmiela, S., Tkatchenko, A., Sauceda, H. E., Poltavsky, Igor, Schütt, K. T., Müller, K.-R. (2017). Machine learning of accurate energy-conserving molecular force fields. Sci. Adv., 3(5), e1603015.
[33] Ju, S., Shiga T., Feng L., Hou Z., Tsuda, K., Shiomi J. (2017). Designing nanostructures for phonon transport via bayesian optimization. Phys. Rev. X, 7(2), 021024.
[34] Ramakrishnan, R, von Lilienfeld, A. (2017). Machine learning, quantum chemistry, and chemical space. Reviews in Computational Chemistry, 225-256.
[35] Hernandez-Lobato, J. M., Requeima, J., Pyzer-Knapp, E. O., Aspuru-Guzik, A. (2017). Parallel and distributed Thompson sampling for large-scale accelerated exploration of chemical space. arXiv preprint arXiv:1706.01825.
[36] Smith, J., Isayev, O., Roitberg, A. E. (2017). ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. Chem. Sci., 8(4), 3192-3203.
[37] Brockherde, F., Li, L., Burke, K., Müller, K.-R. By-passing the Kohn-Sham equations with machine learning. Nat. Commun., in press.
[38] Schütt, K. T., Kindermans, P. J., Sauceda, H. E., Chmiela, S., Tkatchenko, A., Müller, K. R. (2017). MolecuLeNet: A continuous-filter convolutional neural network for modeling quantum interactions. NIPS (accepted).

Schedule

08:00 AM	**Opening remarks**	*Müller*
08:20 AM	**Machine Learning for Molecular Materials Design**	*Aspuru-Guzik*
08:45 AM	**TBA4**	*DiStasio Jr.*
09:00 AM	**New Density Functionals Created by Machine Learning**	*burke*

09:35 AM	**Poster spotlights**	*Strubell, Goh, Tsubaki, Gaudin, Schwaller, Ragoza, Gomez-Bombarelli*
10:45 AM	**Quantum Machine Learning**	*von Lilienfeld*
11:05 AM	**Machine Learning in Organic Synthesis Planning And Execution**	*Jensen*
11:25 AM	**Neural-network Quantum States**	*Carleo*
11:40 AM	**Quantitative Attribution: Do Neural Network Models Learn the Correct Chemistry?**	*Colwell*
01:55 PM	**TBA11**	*Smola*
02:00 PM	**ChemTS: An Efficient Python Library for De Novo Molecular Generation**	*Tsuda*
02:15 PM	**Symmetry Matters: Learning Scalars and Tensors in Materials and Molecules**	*Ceriotti*
02:35 PM	**Towards Exact Molecular Dynamics Simulations with Machine-Learned Force Fields**	*Chmiela*
03:25 PM	**N-body Neural Networks: A General Compositional Architecture For Representing Multiscale Physical Systems**	*Kondor*
03:45 PM	**Distilling Expensive Simulations with Neural Networks**	*Vinyals*
04:05 PM	**Automatic Chemical Design Using a Data-driven Continuous Representation of Molecules**	*Duvenaud*
04:20 PM	**Planning Chemical Syntheses with Neural Networks and Monte Carlo Tree Search**	*Segler*
05:50 PM	**Closing remarks**	*Hernández-Lobato*

Workshop on Worm's Neural Information Processing (WNIP)

Ramin Hasani, Manuel Zimmer, Stephen Larson, Radu Grosu

S5, 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 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/

Schedule

09:00 AM	**Openning Remarks**	*Hasani*
09:15 AM	**Capturing the continuous complexity of natural behavior**	*Stephens*
09:45 AM	**Neuronal analysis of value-based decision making in C. elegans**	*Lockery*
10:15 AM	**3 spotlight presentations**	*Buchanan, Lechner, Li*
10:30 AM	**Poster Session 1**	*Fuchs, Lung, Lechner, Li, Gordus, Venkatachalam, Chaudhary, H■la, Rolnick, Linderman, Mena, Paninski, Cohen*
11:00 AM	**Mechanisms and Functions of Neuronal Population Dynamics in C. elegans**	*Zimmer*
11:30 AM	**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**	*Cohen*
12:00 PM	**Multi-neuronal imaging of C. elegans courtship and mating**	*Venkatachalam*
12:15 PM	**Using Network Control Principles to Probe the Structure and Function of Neuronal Connectomes**	*Schafer*
12:30 PM	**Lunch Break**	
02:00 PM	**Biological Neurons Are Different From Neural Networks: Simulating C.elegans in an open science project**	*Larson*
02:30 PM	**Parking with a worm's brain**	*Grosu*
03:00 PM	**Break / Poster Session 2**	
03:30 PM	**Evolving Neural Circuits for Behavior: C. elegans Locomotion**	*Izquierdo*
04:00 PM	**Panel Discussion**	

Abstracts (7):

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 ($A > B > C \Rightarrow A > C$). 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 chemosensory 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 importantproperties to understand its dynamic state. While the action potential or discrete “spike” featureof mammalian neurons has been emphasized as an information bearing signal, biologicalevidence exists that even without action potentials, neurons process information and give rise todifferent 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, andmany others.
For thirty years, the C. elegans nervous system has remained the only adult animal that has hadits 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 a1simulation framework, known as c302, that enables us to assemble the known connectivity andother biological data of the C. elegans nervous system into a HodgkinHuxleybased simulationthat 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 simplecrawling and swimming behavior. With the goal of producing the same simple sinusoids in themuscle cells, we have used c302 to select a subnetwork from the full C. elegans nervoussystem and used machine learning techniques to fit dynamic parameters that are underspecifiedby the data. Our preliminary results still leave many important biological features out, butinitially demonstrate that it is possible to make motor neurons produce sinusoidal activitypatterns in the muscles as used in the physical simulation.

In this talk I will discuss these initial results and discuss future directions for a betterunderstanding 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 are 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.

Machine Learning for the Developing World

Maria De-Arteaga, William Herlands

S7, 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 the developing world.

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.

Schedule

08:45 AM	**Introductory remarks**	*Dubrawski*
09:00 AM	**Skyler Speakman (IBM Research Africa): Three Population Covariate Shift for Mobile Phone-based Credit Scoring**	*Speakman*

09:30 AM	**Contributed talk: Unique Entity Estimation with Application to the Syrian Conflict**	*Chen*
09:40 AM	**Contributed talk: Field Test Evaluation of Predictive Models for Wildlife Poaching Activity in Uganda**	*Gholami*
09:50 AM	**Contributed talk: Household poverty classification in data-scarce environments: a machine learning approach**	*Kshirsagar*
10:00 AM	**Ernest Mwebaze (UN Global Pulse): ML4D: what works and how it works - case studies from the developing world**	*Mwebaze*
11:00 AM	**Emma Brunskill (Stanford)**	*Brunskill*
11:30 AM	**P. Anandan (Wadhwani Institute of AI)**	
11:40 AM	**Posters**	*Bhattacharya, Lam, Vidyapu, Shankar, Anders, Wilder, Khan, Li, Saquib, Kshirsagar, Perez, Zhang, Gholami, Abebe*
12:30 PM	**Lunch**	
02:00 PM	**Jen Ziemke (International Crisis Mappers)**	*Ziemke*
02:30 PM	**Caitlin Augustin (DataKind): Data for Social Good**	*Augustin*
03:00 PM	**Coffee break**	
03:30 PM	**Stefano Ermon (Stanford): Measuring Progress Towards Sustainable Development Goals with Machine Learning**	*Ermon*
04:00 PM	**Panel discussion**	

Abstracts (5):

Abstract 2: **Skyler Speakman (IBM Research Africa): Three Population Covariate Shift for Mobile Phone-based Credit Scoring in Machine Learning for the Developing World**, *Speakman* 09:00 AM

Mobile money platforms are gaining traction across developing markets as a convenient way of sending and receiving money over mobile phones. Recent joint collaborations between banks and mobile-network operators leverage a customer's past mobile phone transactions in order to create a credit score for the individual. These scores allow access to low-value, short-term, un-collateralized loans. In this talk we will look at the problem of launching a mobile-phone based credit scoring system in a new market without either labeled examples of repayment or the marginal distribution of features of borrowers in the new market. The latter assumption rules out traditional transfer learning approaches such as a direct covariate shift. We apply a Three Population Covariate Shift method to account for the differences in the original and new markets. The three populations are: a) Original Market Members, b) Original Market Borrowers who self-selected into a loan product, and c) New Market Members. The goal of applying a generalized covariate shift to these three populations is to understand the repayment behavior of a fourth: d) New Market Borrowers who will self-select into a loan product when it becomes available.

Abstract 6: **Ernest Mwebaze (UN Global Pulse): ML4D: what works and how it works - case studies from the developing world in Machine Learning for the Developing World**, *Mwebaze* 10:00 AM

Present advances in Artificial Intelligence and Machine Learning offer unique opportunities for solving impactful developing world problems. At the AI and Data Science research lab at Makerere University and UN PulseLab Kampala in Uganda, we have 8 years of trying to marry good computational techniques with good developing world problems. In this talk I will give some examples of some of the projects we are working on or have worked on. These will include automating disease diagnosis in crops and humans, crowd-sourcing surveillance, traffic monitoring and using public radio data to infer humanitarian crises. I will talk about the relative strengths of the different types of data that can be reliably collected in the developing world and some deployment options that (seem to) work.

Abstract 12: **Caitlin Augustin (DataKind): Data for Social Good in Machine Learning for the Developing World**, *Augustin* 02:30 PM

We are living inside a data revolution that is transforming the way we understand and interact with each other and the world - and it has only just begun. Every field is now having its “data moment,” giving mission-driven organizations brand new opportunities to harness data to advance their work. In fact, the same algorithms that companies use to boost profits can help these organizations boost their impact. From poverty alleviation to healthcare access to improved education, machine learning has the potential to move the needle on seemingly insurmountable issues, but only if there is close collaboration between data scientists and subject matter experts. Since DataKind was founded in 2011, its volunteers have delivered over $25 million in pro bono services to social change organizations worldwide - helping organizations deliver vaccines more effectively to creating chatbots that connect people to critical services during a natural disaster to helping at risk students reach graduation, using satellite imagery to estimate poverty and identify crop diseases, and more. This talk will focus on the ways that DataKind engages with nonprofits across industries and economies, with particular emphasis on techniques, tools, and approaches that can provide guidance to ML in the developing world. We’ll dive in on the exciting potential of big data to tackle big social issues and how data scientists can apply their skills for the greater good.

Abstract 14: **Stefano Ermon (Stanford): Measuring Progress Towards Sustainable Development Goals with Machine Learning in Machine Learning for the Developing World**, *Ermon* 03:30 PM

Recent technological developments are creating new spatio-temporal data streams that contain a wealth of information relevant to sustainable development goals. Modern AI techniques have the potential to yield accurate, inexpensive, and highly scalable models to inform research and policy. As a first example, I will present a machine learning method we developed to predict and map poverty in developing countries. Our method can reliably predict economic well-being using only high-resolution satellite imagery. Because images are passively collected in every corner of the world, our method can provide timely and accurate measurements in a very scalable end economic way, and could revolutionize efforts towards global poverty eradication. As a second example, I will present some ongoing work on monitoring food security outcomes.

Abstract 15: **Panel discussion in Machine Learning for the Developing World**, 04:00 PM

This panel discussion brings together core machine learning researchers and developing world application domain experts in a conversation regarding challenges, opportunities and future directions of ML4D.

Advances in Approximate Bayesian Inference

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

Seaside Ballroom, 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

08:30 AM	**Introduction**	*Zhang, Ruiz, Tran, McInerney, Mandt*
08:35 AM	**Invited talk: Iain Murray (TBA)**	*Murray*
09:00 AM	**Contributed talk: Learning Implicit Generative Models Using Differentiable Graph Tests**	*Djolonga*
09:15 AM	**Invited talk: Gradient Estimators for Implicit Models**	*Li*
09:40 AM	**Industry talk: Variational Autoencoders for Recommendation**	*Liang*
10:00 AM	**Poster Spotlights**	*Locatello, Pakman, Tang, Rainforth, Borsos, Järvenpää, Nalisnick, Abbati, LU, Huggins, Singh, Luo*
10:30 AM	**Coffee Break and Poster Session 1**	
11:25 AM	**Industry talk: Cedric Archambeau (TBA)**	*Archambeau*
11:45 AM	**Contributed talk: Variational Inference based on Robust Divergences**	*Futami*
12:00 PM	**Lunch Break**	
01:00 PM	**Poster Session**	*Horii, Jeong, Schwedes, He, Calderhead, Erdil, Altosaar, Muchmore, Khanna, Gemp, Zhang, Zhou, Cremer, DeYoreo, Terenin, McVeigh, Singh, Yang, Bodin, Evans, Chai, Zhe, Ling, ADAM, Maaløe, Miller, Pakman, Djolonga, Ge*
02:05 PM	**Contributed talk: Adversarial Sequential Monte Carlo**	*Kempinska*
02:20 PM	**Contributed talk: Scalable Logit Gaussian Process Classification**	*Wenzel*
02:35 PM	**Invited talk: Variational Inference in Deep Gaussian Processes**	*Damianou*
03:00 PM	**Coffee Break and Poster Session 2**	
03:30 PM	**Contributed talk: Taylor Residual Estimators via Automatic Differentiation**	*Miller*
03:45 PM	**Invited talk: Differential privacy and Bayesian learning**	*Honkela*
04:10 PM	**Contributed talk: Frequentist Consistency of Variational Bayes**	*Wang*
04:25 PM	**Panel: On the Foundations and Future of Approximate Inference**	*Blei, Ghahramani, Heller, Salimans, Welling, Hoffman*

Dec. 9, 2017

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

Benjamin Guedj, Pascal Germain, Francis Bach

101 A, 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 Willamson (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
[1] J. Shawe-Taylor and R. Williamson. A PAC analysis of a Bayes estimator. In Proceedings of COLT, 1997.
[2] D. A. McAllester. Some PAC-Bayesian theorems. In Proceedings of COLT, 1998.
[3] D. A. McAllester. PAC-Bayesian model averaging. In Proceedings of COLT, 1999.
[4] O. Catoni. Statistical Learning Theory and Stochastic Optimization. Saint-Flour Summer School on Probability Theory 2001 (Jean Picard ed.), Lecture Notes in Mathematics. Springer, 2004.
[5] O. Catoni. PAC-Bayesian supervised classification: the thermodynamics of statistical learning. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 56. Institute of Mathematical Statistics, 2007.
[6] P. D. Grünwald. The Minimum Description Length Principle. The MIT Press, 2007.

Schedule

08:30 AM	**Overture**	*Guedj, Bach, Germain*
08:30 AM	**François Laviolette - A Tutorial on PAC-Bayesian Theory**	*Laviolette*
09:30 AM	**Peter Grünwald - A Tight Excess Risk Bound via a Unified PAC-Bayesian-Rademacher-Shtarkov-MDL Complexity**	*Grünwald*
11:00 AM	**Jean-Michel Marin - Some recent advances on Approximate Bayesian Computation techniques**	*Marin*
11:45 AM	**Contributed talk 1 - A PAC-Bayesian Approach to Spectrally-Normalized Margin Bounds for Neural Networks**	*Neyshabur*
02:00 PM	**Olivier Catoni - Dimension-free PAC-Bayesian Bounds**	*Catoni*
02:40 PM	**Contributed talk 2 - Dimension free PAC-Bayesian bounds for the estimation of the mean of a random vector**	*Catoni*
03:30 PM	**Yevgeny Seldin - A Strongly Quasiconvex PAC-Bayesian Bound**	*Seldin*
04:15 PM	**John Shawe-Taylor - Distribution Dependent Priors for Stable Learning**	*Shawe-Taylor*
05:00 PM	**Daniel Roy - Deep Neural Networks: From Flat Minima to Numerically Nonvacuous Generalization Bounds via PAC-Bayes**	*Roy*
05:30 PM	**Neil Lawrence, Francis Bach and François Laviolette**	*Lawrence, Bach, Laviolette*
06:25 PM	**Concluding remarks**	*Bach, Guedj, Germain*

Abstracts (5):

Abstract 3: **Peter Grünwald - A Tight Excess Risk Bound via a Unified PAC-Bayesian-Rademacher-Shtarkov-MDL Complexity in (Almost) 50 shades of Bayesian Learning: PAC-Bayesian trends and insights**, *Grünwald* 09:30 AM

Over the last 15 years, machine learning theorists have bounded the performance of empirical risk minimization by (localized) Rademacher complexity; Bayesians with frequentist sympathies have studied Bayesian consistency and rate of convergence theorems, sometimes under misspecification, and PAC-Bayesians have studied convergence properties of generalized Bayesian and Gibbs posteriors. We show that, amazingly, most such bounds readily follow from essentially a single result that bounds excess risk in terms of a novel complexity COMP(η,w). which depends on a learning rate η and a luckiness function w, the latter generalizing the concept of a 'prior'. Depending on the choice of w, COMP(η,w) specializes to PAC-Bayesian (KL(posterior||prior) complexity, MDL (normalized maximum likelihood) complexity and Rademacher complexity, and the bounds obtained are optimized for generalized Bayes, ERM, penalized ERM (such as Lasso) or other methods. Tuning η leads to optimal excess risk convergence rates, even for very large (polynomial entropy) classes which have always been problematic for the PAC-Bayesian approach; the optimal η depends on 'fast rate' properties of the domain, such as central, Bernstein and Tsybakov conditions.

Joint work with Nishant Mehta, University of Victoria. See https://arxiv.org/abs/1710.07732

Abstract 4: **Jean-Michel Marin - Some recent advances on Approximate Bayesian Computation techniques in (Almost) 50 shades of Bayesian Learning: PAC-Bayesian trends and insights**, *Marin* 11:00 AM

In an increasing number of application domains, the statistical model is so complex that the point-wise computation of the likelihood is intractable. That is typically the case when the underlying probability distribution involves numerous latent variables. Approximate Bayesian Computation (ABC) is a widely used technique to bypass that difficulty. I will review some recent developments on ABC techniques, emphazing the fact that modern machine learning approaches are useful in this field. Although intrinsically very different of PAC-Bayesian strategies - the choice of a generative model is essential within the ABC paradigm - I will highlight some links between these two methodologies.

Abstract 6: **Olivier Catoni - Dimension-free PAC-Bayesian Bounds in (Almost) 50 shades of Bayesian Learning: PAC-Bayesian trends and insights**, *Catoni* 02:00 PM

PAC-Bayesian inequalities have already proved to be a great tool to obtain dimension free generalization bounds, such as margin bounds for Support Vector Machines. In this talk, we will play with PAC-Bayesian inequalities and influence functions to present new robust estimators for the mean of random vectors and random matrices, as well as for linear least squares regression. A common theme of the presentation will be to establish dimension free bounds and to work under mild polynomial moment assumptions regarding the tail of the sample distribution.

Joint work with Ilaria Giulini.

Abstract 8: **Yevgeny Seldin - A Strongly Quasiconvex PAC-Bayesian Bound in (Almost) 50 shades of Bayesian Learning: PAC-Bayesian trends and insights**, *Seldin* 03:30 PM

We propose a new PAC-Bayesian bound and a way of constructing a hypothesis space, so that the bound is convex in the posterior distribution and also convex in a trade-off parameter between empirical performance of the posterior distribution and its complexity. The complexity is measured by the Kullback-Leibler divergence to a prior. We derive an alternating procedure for minimizing the bound. We show that the bound can be rewritten as a one-dimensional function of the trade-off parameter and provide sufficient conditions under which the function has a single global minimum. When the conditions are satisfied the alternating minimization is guaranteed to converge to the global minimum of the bound. We provide experimental results demonstrating that rigorous minimization of the bound is competitive with cross-validation in tuning the trade-off between complexity and empirical performance. In all our experiments the trade-off turned to be quasiconvex even when the sufficient conditions were violated.

Joint work with Niklas Thiemann, Christian Igel, and Olivier Wintenberger.

Abstract 10: **Daniel Roy - Deep Neural Networks: From Flat Minima to Numerically Nonvacuous Generalization Bounds via PAC-Bayes in (Almost) 50 shades of Bayesian Learning: PAC-Bayesian trends and insights**, *Roy* 05:00 PM

One of the defining properties of deep learning is that models are chosen to have many more parameters than available training data. In light of this capacity for overfitting, it is remarkable that simple algorithms like SGD reliably return solutions with low test error. One roadblock to explaining these phenomena in terms of implicit regularization, structural properties of the solution, and/or easiness of the data is that many learning bounds are quantitatively vacuous when applied to networks learned by SGD in this "deep learning" regime. Logically, in order to explain generalization, we need nonvacuous bounds.

I will discuss recent work using PAC-Bayesian bounds and optimization to arrive at nonvacuous generalization bounds for neural networks with millions of parameters trained on only tens of thousands of examples. We connect our findings to recent and old work on flat minima and MDL-based explanations of generalization, as well as to variational inference for deep learning. Time permitting, I'll discuss new work interpreting Entropy-SGD as a PAC-Bayesian method.

Joint work with Gintare Karolina Dziugaite, based on https://arxiv.org/abs/1703.11008

Deep Learning at Supercomputer Scale

Erich Elsen, Danijar Hafner, Zak Stone, Brennan Saeta

101 B, 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.

Schedule

08:10 AM	**Generalization Gap**	*Keskar*
08:30 AM	**Closing the Generalization Gap**	*Hubara*
08:50 AM	**Don't Decay the Learning Rate, Increase the Batch Size**	*Smith*
09:10 AM	**ImageNet In 1 Hour**	*Goyal*
09:30 AM	**Training with TPUs**	*Ying*
09:50 AM	**Coffee Break**	
10:10 AM	**KFAC and Natural Gradients**	*Johnson, Duckworth*
10:30 AM	**Neumann Optimizer**	*Krishnan*
10:50 AM	**Evolutionary Strategies**	*Salimans*
11:15 AM	**Future Hardware Directions**	*Diamos, Dean, Knowles, James, Gray*
01:30 PM	**Learning Device Placement**	*Mirhoseini*
01:50 PM	**Scaling and Sparsity**	*Diamos*
02:10 PM	**Small World Network Architectures**	*Gray*
02:30 PM	**Scalable RL and AlphaGo**	*Lillicrap*
03:20 PM	**Scaling Deep Learning to 15 PetaFlops**	*Kurth*
03:40 PM	**Scalable Silicon Compute**	*Knowles*
04:00 PM	**Practical Scaling Techniques**	*Kapasi*
04:20 PM	**Designing for Supercompute-Scale Deep Learning**	*James*
05:00 PM	**Adaptive Memory Networks**	*Li*
05:00 PM	**Supercomputers for Deep Learning**	*Sukumar*

Machine Learning on the Phone and other Consumer Devices

Hrishi Aradhye, Joaquin Quinonero Candela, Rohit Prasad

102 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 open 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 in on-device machine learning, and in 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 ground-breaking experiences on real-world products.

[1] https://developer.apple.com/machine-learning/
[2] https://caffe2.ai/
[3] https://www.tensorflow.org/mobile/
[4]
https://opensource.googleblog.com/2017/06/mobilenets-open-source-models-for.html

Schedule

08:05 AM	**Qualcomm presentation on ML-optimized mobile hardware**	*Teague*
08:30 AM	**fpgaConvNet: A Toolflow for Mapping Diverse Convolutional Neural Networks on Embedded FPGAs**	*Venieris*
08:45 AM	**High performance ultra-low-precision convolutions on mobile devices**	*Tulloch, Jia*
09:00 AM	**Caffe2: Lessons from Running Deep Learning on the World’s Smart Phones**	*Jia*
09:30 AM	**CoreML: High-Performance On-Device Inference**	*Kapoor*
10:00 AM	**Data center to the edge: a journey with TensorFlow**	*Monga*
11:00 AM	**On-Device ML Frameworks**	*Gehlhaar, Jia, Monga*
11:45 AM	**Poster Spotlight 1**	
12:05 PM	**Lunch**	
12:05 PM	**Poster Session 1**	
01:30 PM	**Federated learning for model training on decentralized data**	*Ramage*
02:00 PM	**Personalized and Private Peer-to-Peer Machine Learning**	*Bellet, Guerraoui, Tommasi*
02:15 PM	**SquishedNets: Squishing SqueezeNet further for edge device scenarios via deep evolutionary synthesis**	*Li*
02:30 PM	**A Cascade Architecture for Keyword Spotting on Mobile Devices**	*Alvarez, Thornton, Ghodrat*
02:45 PM	**Multiple-Instance, Cascaded Classification for Keyword Spotting in Narrow-Band Audio**	*Abdulkader, Nassar, Mahmoud, Galvez*
03:00 PM	**Coffee Break**	
03:30 PM	**Machine Learning for Alexa**	*Mandal*
04:00 PM	**Now Playing: Continuous low-power music recognition**	*Ritter, Guo, Kumar, Odell, Velimirovi■, Roblek, Lyon*
04:15 PM	**Learning On-Device Conversational Models**	*Ravi, Rudick, Fan*
04:30 PM	**Google Lens**	*Adam*
05:00 PM	**Poster Spotlight 2**	
05:20 PM	**Poster Session 2**	*Shafiq, Nevado Vilchez, Yamada, Dasgupta, Geyer, Nabi, Rodrigues, Manino, Serb, Carreira-Perpinan, Lim, Low, Pandey, White, Pidlypenskyi, Wang, Kaeser-Chen, Zhu, Gupta, Leroux*

Synergies in Geometric Data Analysis (TWO DAYS)

Marina Meila, Frederic Chazal, yuchaz Chen

102 C, Sat Dec 09, 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 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 expliecitly 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. Trough the tutorials, we are reaching out to the wider NIPS audience, to the many potential users of 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.

Medical Imaging meets NIPS

Ben Glocker, Ender Konukoglu, Hervé Lombaert, Kanwal Bhatia

103 A+B, 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)

Schedule

08:45 AM	**Opening**	*Glocker, Konukoglu, Lombaert, Bhatia*
09:00 AM	**Shaping the future through innovations: Artificial intelligence for healthcare (Siemens)**	*Comaniciu*
09:30 AM	**Role of AI and Deep Learning in Radiology (IBM)**	*Syeda-Mahmood*
10:00 AM	**Pixel Perfectionism - Machine learning and Adaptive Radiation Therapy (University of Cambridge & Microsoft Research)**	*Jena*
10:30 AM	**Coffee break - Morning**	
10:30 AM	**Poster session - Morning**	*Kwon, Kim, Konukoglu, Li, Guibas, Virdi, Kumar, Mardani, Wolterink, Gong, Antropova, Stelzer, Bidart, Weng, Rajchl, Górriz, Singh, Sandino, Chougrad, Hu, Godfried, Xiao, Tejeda Lemus, Harrod, WOO, Chen, Cheng, Gupta, Yee, Glocker, Lombaert, Ilse, Lisowska, Doyle, Makkie*
11:00 AM	**Deep learning for cardiovascular image analysis (University Medical Center Utrecht)**	*Isgum*
11:30 AM	**Deep learning for image reconstruction, segmentation and super-resolution in medical imaging (Imperial College London)**	*Rueckert*
12:00 PM	**Biomedical Imaging and Genetic (BIG) Data Analytics in Dementia and Oncology (Erasmus MC Rotterdam, Quantib)**	*Niessen*
12:30 PM	**Lunch break**	
01:30 PM	**Scikit-learn & nilearn: Democratisation of machine learning for brain imaging (INRIA)**	*Varoquaux*
02:00 PM	**The Multimodal Brain Tumor Segmentation Challenge (TU Munich)**	*Menze*
02:30 PM	**DeepLumen: Accurate Vessel Segmentation for Better Cardiovascular Care (HeartFlow)**	*Petersen*

03:00 PM	**Poster session - Afternoon**	*Kwon, Kim, Konukoglu, Li, Guibas, Virdi, Kumar, Mardani, Wolterink, Gong, Antropova, Stelzer, Bidart, Weng, Rajchl, Górriz, Singh, Sandino, Chougrad, Hu, Godfried, Xiao, Tejeda Lemus, Harrod, WOO, Chen, Cheng, Gupta, Yee, Glocker, Lombaert, Ilse, Lisowska, Doyle, Makkie*
03:00 PM	**Coffee break - Afternoon**	
03:30 PM	**NiftyNet: An open-source comunity-driven framework for neural networks in medical imaging (University College London)**	*Cardoso*
04:00 PM	**Winner of The Digital Mammography DREAM Challenge (Therapixel)**	*Nikulin*
04:30 PM	**Panel discussion**	
05:00 PM	**Closing**	*Glocker, Konukoglu, Lombaert, Bhatia*

Abstracts (1):

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

Opening remarks by the organizers

Workshop on Prioritising Online Content

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

103 C, 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**	*Vlachos*
10:30 AM	**Reducing controversy by connecting opposing views**	*Gionis, Garimella*
10:50 AM	**Leveraging the Crowd to Detect and Reduce the Spread of Fake News and Misinformation**	*Oh, Schölkopf*
11:10 AM	**A Framework for Automated Fact-Checking for Real-Time Validation of Emerging Claims on the Web**	*Hanselowski*
11:30 AM	**Equality of Opportunity in Rankings**	*Joachims, Singh*
11:35 AM	**The Unfair Externalities of Exploration**	*Slivkins, Wortman Vaughan*
11:40 AM	**Developing an Information Source Lexicon**	*Addawood, Mishra*
11:45 AM	**Mitigating the spread of fake news by identifying and disrupting echo chambers**	*Warren, Nguyen*
11:50 AM	**An Efficient Method to Impose Fairness in Linear Models**	*Pontil, Shawe-Taylor*
11:55 AM	**Poster session**	*Cesa-Bianchi*
12:10 PM	**Lunch Break**	
01:00 PM	**Philosophy and ethics of defining, identifying, and tackling fake news and inappropriate content**	*Watkins*
02:00 PM	**Political echo chambers in social media**	*Gionis*
03:00 PM	**Coffee break**	
03:30 PM	**Reality around fake news**	

Abstracts (1):

Abstract 13: **Political echo chambers in social media in Workshop on Prioritising Online Content**, *Gionis* 02:00 PM

Echo chambers describe situations where one is exposed only to opinions that agree with their own. In this talk we will discuss the phenomenon of political echo chambers in social media. We identify the different components in the phenomenon and characterize users based on their behavior with respect to content production and consumption. Among other findings, we observe that users who try to bridge the echo chambers have to pay a "price of bipartisanship." We then discuss ideas for combating echo chambers. We first present a model for learning ideological-leaning factors, of social-media users and media sources, in a joint latent space. The model space can be used to develop exploratory and interactive interfaces that can help users to diffuse their information filter bubble. Second we present an influence-based approach for balancing the information exposure of users in the social network.

Cognitively Informed Artificial Intelligence: Insights From Natural Intelligence

Mike Mozer, Brenden Lake, Angela J Yu

104 A, Sat Dec 09, 08:00 AM

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 premise of this workshop is that there are valuable data and models from cognitive science that can inform the development of intelligent adaptive machines, and can endow learning architectures with the strength and flexibility of the human cognitive architecture. The structures and mechanisms of the mind and brain can provide the sort of strong inductive bias needed for machine-learning systems to attain human-like performance. We conjecture that this inductive bias will become more important as researchers move from domain-specific tasks such as object and speech recognition toward tackling general intelligence and the human-like ability to dynamically reconfigure cognition in service of changing goals. For ML researchers, the workshop will provide access to a wealth of data and concepts situated in the context of contemporary ML. For cognitive scientists, the workshop will suggest research questions that are of critical interest to ML researchers.

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

08:30 AM	**Workshop overview**	*Mozer, Yu, Lake*
08:40 AM	**Cognitive AI**	*Lake*
09:05 AM	**Computational modeling of human face processing**	*Yu*
09:30 AM	**People infer object shape in a 3D, object-centered coordinate system**	*Jacobs*
09:55 AM	**Relational neural expectation maximization**	*van Steenkiste*
10:10 AM	**Contextual dependence of human preference for complex objects: A Bayesian statistical account**	*Ryali*
10:15 AM	**A biologically-inspired sparse, topographic recurrent neural network model for robust change detection**	*Sridharan*
10:20 AM	**Visual attention guided deep imitation learning**	*Zhang*
10:25 AM	**Human learning of video games**	*Tsividis*
10:30 AM	**COFFEE BREAK AND POSTER SESSION**	
11:00 AM	**Life history and learning: Extended human childhood as a way to resolve explore/exploit trade-offs and improve hypothesis search**	*Gopnik*
11:25 AM	**Meta-reinforcement learning in brains and machines**	*Botvinick*
11:50 AM	**Revealing human inductive biases and metacognitive processes with rational models**	*Griffiths*
12:15 PM	**Learning to select computations**	*Lieder, Callaway, Gul, Krueger*
02:00 PM	**From deep learning of disentangled representations to higher-level cognition**	*Bengio*
02:25 PM	**Access consciousness and the construction of actionable representations**	*Mozer*
02:50 PM	**Evaluating the capacity to reason about beliefs**	*Nematzadeh*
03:05 PM	**COFFEE BREAK AND POSTER SESSION II**	
03:30 PM	**Mapping the spatio-temporal dynamics of cognition in the human brain**	*Oliva*
03:55 PM	**Scale-invariant temporal memory in AI**	*Howard*
04:20 PM	**Scale-invariant temporal history (SITH): Optimal slicing of the past in an uncertain world**	*Spears, Jacques, Howard, Sederberg*
04:35 PM	**Efficient human-like semantic representations via the information bottleneck principle**	*Zaslavsky*
04:40 PM	**The mutation sampler: A sampling approach to causal representation**	*Davis*
04:45 PM	**Generating more human-like recommendations with a cognitive model of generalization**	*Bourgin*
04:50 PM	**POSTER: Curiosity-driven reinforcement learning with hoemostatic regulation**	*Magrans de Abril*
04:50 PM	**POSTER: Context-modulation of hippocampal dynamics and deep convolutional networks**	*Aimone*
04:50 PM	**POSTER: Cognitive modeling and the wisdom of the crowd**	*Lee*
04:50 PM	**POSTER: Concept acquisition through meta-learning**	*Grant*

04:50 PM	**POSTER: Pre-training attentional mechanisms**	*Lindsey*
04:50 PM	**POSTER: Using STDP for unsupervised, event-based online learning**	*Thiele*
04:50 PM	**POSTER: Learning to organize knowledge with N-gram machines**	*Yang*
04:50 PM	**POSTER: Sample-efficient reinforcement learning through transfer and architectural priors**	*Spector*
04:50 PM	**POSTER: Improving transfer using augmented feedback in progressive neural networks**	*Bablani, Chadha*
04:50 PM	**POSTER: Question asking as program generation**	*Rothe*
04:50 PM	**POSTER: Variational probability flow for biologically plausible training of deep neural networks**	*LIU, Lin*
04:50 PM	**POSTER: Power-law temporal discounting over a logarithmically compressed timeline for scale invariant reinforcement learning**	*Tiganj*
05:25 PM	**Object-oriented intelligence**	*Battaglia*
05:50 PM	**Representational primitives, in minds and machines**	*Marcus*

Machine Learning in Computational Biology

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

104 B, 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

08:55 AM	**Opening Remarks**
09:00 AM	**Christina Leslie - Decoding immune cell states and dysfunction**
09:40 AM	**Denoising scRNA-seq Data Using Deep Count Autoencoders (Gökcen Eraslan)**
09:55 AM	**Fine Mapping of Chromatin Interactions via Neural Nets (Artur Jaroszewicz)**
10:10 AM	**F-MoDISco: Learning High-Quality, Non-Redundant Transcription Factor Binding Motifs Using Deep Learning (Avanti Shrikumar)**
10:30 AM	**Coffee break + posters**
11:00 AM	**Spotlights**
12:00 PM	**Lunch+Posters**
01:30 PM	**Posters**
02:00 PM	**Eran Halperin - A new sparse PCA algorithm with guaranteed asymptotic properties and applications in methylation data**

02:40 PM	**Variational Bayes inference algorithm for causal multivariate mediation with linkage disequilibrium (Yongjin Park)**
02:55 PM	**Reference-Free Archaic Admixture Segmentation Using A Permutation-Equivariant Network (Jeffrey Chan)**
03:10 PM	**Robust and Scalable Models of Microbiome Dynamics for Bacteriotherapy Design (Travis Gibson)**
04:00 PM	**Ben Raphael - Inferring Tumor Evolution**
04:40 PM	**Drug Response Variational Autoencoder (Ladislav Rampášek)**
04:55 PM	**Variational auto-encoding of protein sequences (Sam Sinai)**
05:10 PM	**Antigen Identification for Cancer Immunotherapy by Deep Learning on Tumor HLA Peptides**
05:25 PM	**Sponsors + prizes**
06:00 PM	**Test push to whova**

The future of gradient-based machine learning software & techniques

Alex Wiltschko, Bart van Merriënboer, Pascal Lamblin

104 C, 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

09:00 AM	**Introduction and opening remarks**	*Wiltschko*
09:10 AM	**Beyond backprop: automatic differentiation in machine learning**	*Baydin*
09:50 AM	**Automatic differentiation in PyTorch**	*Paszke*
10:30 AM	**Morning Coffee Break**	
11:00 AM	**Optimal Smoothing for Pathwise Adjoints**	*Hüser*
11:40 AM	**Poster session**	*Naumann, Schwartz, Wei, Meissner, Druce, Lin, Pothen, Yang*
01:40 PM	**Algorithmic differentiation techniques in the deep learning context**	*Utke*
02:20 PM	**Some highlights on Source-to-Source Adjoint AD**	*Hascoet*
03:00 PM	**Afternoon Coffee Break**	
03:30 PM	**Divide-and-Conquer Checkpointing for Arbitrary Programs with No User Annotation**	*Siskind*
04:10 PM	**Automatic Differentiation of Parallelised Convolutional Neural Networks - Lessons from Adjoint PDE Solvers**	*Hueckelheim*
04:50 PM	**Panel discussion**	*Baydin, Paszke, Hüser, Utke, Hascoet, Siskind, Hueckelheim, Griewank*

2017 NIPS Workshop on Machine Learning for Intelligent Transportation Systems

Li Erran Li, Anca Dragan, Juan Carlos Niebles, Silvio Savarese

201 A, Sat Dec 09, 08:00 AM

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,
machine learning approaches for control and coordination of traffic leveraging V2V and V2X infrastructures,

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

08:45 AM	**Opening Remarks**	*Li*
09:00 AM	**Machine Learning for Self-Driving Cars, Raquel Urtasun, Uber ATG and University of Toronto**	*Urtasun*
09:30 AM	**Exhausting the Sim with Domain Randomization and Trying to Exhaust the Real World, Pieter Abbeel, UC Berkeley and Embodied Intelligence**	*Abbeel, Kahn*
10:00 AM	**Learning-based system identification and decision making for autonomous driving, Marin Kobilarov, Zoox**	*Kobilarov*
10:30 AM	**Poster and Coffee**	
11:00 AM	**Hesham M. Eraqi, Mohamed N. Moustafa, Jens Honer, End-to-End Deep Learning for Steering Autonomous Vehicles Considering Temporal Dependencies**	
11:10 AM	**Abhinav Jauhri (CMU), Carlee Joe-Wong, John Paul Shen, On the Real-time Vehicle Placement Problem**	*Jauhri, Shen*
11:20 AM	**Andrew Best (UNC), Sahil Narang, Lucas Pasqualin, Daniel Barber, Dinesh Manocha, AutonoVi-Sim: Autonomous Vehicle Simulation Platform with Weather, Sensing, and Traffic control**	*Best*
11:30 AM	**Nikita Japuria (MIT), Golnaz Habibi, Jonathan P. How, CASNSC: A context-based approach for accurate pedestrian motion prediction at intersections**	*Jaipuria*
11:40 AM	**6 Spotlight Talks (3 min each)**	*Siam, Prabhushankar, Parashar, Mukadam, yao, Senanayake*
12:00 PM	**Lunch**	
01:30 PM	**Adaptive Deep Learning for Perception, Action, and Explanation, Trevor Darrell (UC Berkeley)**	*Darrell*
02:00 PM	**The challenges of applying machine learning to autonomous vehicles, Andrej Karpathy, Telsa**	

02:30 PM	**Micro-Perception Approach to Intelligent Transport, Ramesh Sarukkai (Lyft)**	*Sarukkai*
03:00 PM	**Posters and Coffe**	
03:30 PM	**On Autonomous Driving: Challenges and Opportunities, Sertac Karaman, MIT**	*Karaman*
04:00 PM	**Generative Adversarial Imitation Learning, Stefano Ermon, Stanford**	*Ermon*
04:30 PM	**Deep Learning and photo-realistic Simulation for real-world Machine Intelligence in Autonomous Driving, Adrien Gaidon (TRI)**	*Gaidon*
05:00 PM	**Panel Discussion**	*Kahn, Sarukkai, Gaidon, Karaman*

Abstracts (11):

Abstract 2: **Machine Learning for Self-Driving Cars, Raquel Urtasun, Uber ATG and University of Toronto in 2017 NIPS Workshop on Machine Learning for Intelligent Transportation Systems**, *Urtasun* 09:00 AM

Bio: Raquel Urtasun is the Head of Uber ATG Toronto. She is also an Associate Professor in the Department of Computer Science at the University of Toronto, a Raquel Urtasun is the Head of Uber ATG Toronto. She is also an Associate Professor in the Department of Computer Science at the University of Toronto, a Canada Research Chair in Machine Learning and Computer Vision and a co-founder of the Vector Institute for AI. Prior to this, she was an Assistant Professor at the Toyota Technological Institute at Chicago (TTIC), an academic computer science institute affiliated with the University of Chicago. She was also a visiting professor at ETH Zurich during the spring semester of 2010. She received her Bachelors degree from Universidad Publica de Navarra in 2000, her Ph.D. degree from the Computer Science department at Ecole Polytechnique Federal de Lausanne (EPFL) in 2006 and did her postdoc at MIT and UC Berkeley. She is a world leading expert in machine perception for self-driving cars. Her research interests include machine learning, computer vision, robotics and remote sensing. Her lab was selected as an NVIDIA NVAIL lab. She is a recipient of an NSERC EWR Steacie Award, an NVIDIA Pioneers of AI Award, a Ministry of Education and Innovation Early Researcher Award, three Google Faculty Research Awards, an Amazon Faculty Research Award, a Connaught New Researcher Award and two Best Paper Runner up Prize awarded at the Conference on Computer Vision and Pattern Recognition (CVPR) in 2013 and 2017 respectively. She is also an Editor of the International Journal in Computer Vision (IJCV) and has served as Area Chair of multiple machine learning and vision conferences (i.e., NIPS, UAI, ICML, ICLR, CVPR, ECCV).

Abstract 3: **Exhausting the Sim with Domain Randomization and Trying to Exhaust the Real World, Pieter Abbeel, UC Berkeley and Embodied Intelligence in 2017 NIPS Workshop on Machine Learning for Intelligent Transportation Systems**, *Abbeel, Kahn* 09:30 AM

Abstract: Reinforcement learning and imitation learning have seen success in many domains, including autonomous helicopter flight, Atari, simulated locomotion, Go, robotic manipulation. However, sample complexity of these methods remains very high. In this talk I will present two ideas towards effective data collection, and initial findings indicating promise for both: (i) Domain Randomization, which relies on extensive variation (none of it necessarily realistic) in simulation, aiming at generalization to the real world per the real world (hopefully) being like just another random sample. (ii) Self-supervised Deep RL, which considers the problem of autonomous data collection. We evaluate our approach on a real-world RC car and show it can learn to navigate through a complex indoor environment with a few hours of fully autonomous, self-supervised training.

BIO: Pieter Abbeel (Professor at UC Berkeley [2008-], Co-Founder Embodied Intelligence [2017-], Co-Founder Gradescope [2014-], Research Scientist at OpenAI [2016-2017]) works in machine learning and robotics, in particular his research focuses on making robots learn from people (apprenticeship learning), how to make robots learn through their own trial and error (reinforcement learning), and how to speed up skill acquisition through learning-to-learn. His robots have learned advanced helicopter aerobatics, knot-tying, basic assembly, and organizing laundry. His group has pioneered deep reinforcement learning for robotics, including learning visuomotor skills and simulated locomotion. He has won various awards, including best paper awards at ICML, NIPS and ICRA, the Sloan Fellowship, the Air Force Office of Scientific Research Young Investigator Program (AFOSR-YIP) award, the Office of Naval Research Young Investigator Program (ONR-YIP) award, the DARPA Young Faculty Award (DARPA-YFA), the National Science Foundation Faculty Early Career Development Program Award (NSF-CAREER), the Presidential Early Career Award for Scientists and Engineers (PECASE), the CRA-E Undergraduate Research Faculty Mentoring Award, the MIT TR35, the IEEE Robotics and Automation Society (RAS) Early Career Award, IEEE Fellow, and the Dick Volz Best U.S. Ph.D. Thesis in Robotics and Automation Award.

Abstract 4: **Learning-based system identification and decision making for autonomous driving, Marin Kobilarov, Zoox in 2017 NIPS Workshop on Machine Learning for Intelligent Transportation Systems**, *Kobilarov* 10:00 AM

Abstract: The talk will include a brief overview of methods for planning and perception developed at Zoox, and focus on some recent results for learning-based system identification and decision making.

Marin Kobilarov is principal engineer for planning and control at Zoox and assistant professor in Mechanical Engineering at the Johns Hopkins University where he leads the Laboratory for Autonomous Systems, Control, and Optimization. His research focuses on planning and control of robotic systems, on approximation methods for optimization and statistical learning, and applications to autonomous vehicles. Until 2012, he was a postdoctoral fellow in Control and Dynamical Systems at the California Institute of Technology. He obtained a Ph.D. from the University of Southern California in Computer Science (2008) and a B.S. in Computer Science and Applied Mathematics from Trinity College, Hartford, CT (2003).

Abstract 10: **6 Spotlight Talks (3 min each) in 2017 NIPS Workshop on Machine Learning for Intelligent Transportation Systems**, *Siam, Prabhushankar, Parashar, Mukadam, yao, Senanayake* 11:40 AM

1. Mennatullah Siam, Heba Mahgoub, Mohamed Zahran, Senthil Yogamani, Martin Jagersand, Ahmad El-Sallab, Motion and appearance based Multi-Task Learning network for autonomous driving

2. Dogancan Temel, Gukyeong Kwon, Mohit Prabhushankar, Ghassan AlRegib, CURE-TSR: Challenging Unreal and Real Environments for Traffic Sign Recognition

3. Priyam Parashar, Akansel Cosgun, Alireza Nakhaei and Kikuo Fujimura, Modeling Preemptive Behaviors for Uncommon Hazardous Situations From Demonstrations

4. Mustafa Mukadam, Akansel Cosgun, Alireza Nakhaei, Kikuo Fujimura, Tactical Decision Making for Lane Changing with Deep Reinforcement Learning

5. Hengshuai Yao, Masoud S. Nosrati, Kasra Rezaee, Monte-Carlo Tree Search vs. Model-Predictive Controller: A Track-Following Example

6. Ransalu Senanayake, Thushan Ganegedara, Fabio Ramos, Deep occupancy maps: a continuous mapping technique for dynamic environments

Abstract 12: **Adaptive Deep Learning for Perception, Action, and Explanation, Trevor Darrell (UC Berkeley) in 2017 NIPS Workshop on Machine Learning for Intelligent Transportation Systems**, *Darrell* 01:30 PM

Learning of layered or "deep" representations has provided significant advances in computer vision in recent years, but has traditionally been limited to fully supervised settings with very large amounts of training data, where the model lacked interpretability. New results in adversarial adaptive representation learning show how such methods can also excel when learning across modalities and domains, and further can be trained or constrained to provide natural language explanations or multimodal visualizations to their users. I'll present recent long-term recurrent network models that learn cross-modal description and explanation, using implicit and explicit approaches, which can be applied to domains including fine-grained recognition and visuomotor policies.

Abstract 13: **The challenges of applying machine learning to autonomous vehicles, Andrej Karpathy, Telsa in 2017 NIPS Workshop on Machine Learning for Intelligent Transportation Systems**, 02:00 PM

Abstract:
I'll cover cover some of the discrepancies between machine learning problems in academia and those in the industry, especially in the context of autonomous vehicles. I will also cover Tesla's approach to massive fleet learning and some of the associated open research problems.

Bio:
Andrej is a Director of AI at Tesla, where he focuses on computer vision for the Autopilot. Previously he was a research scientist at OpenAI working on Reinforcement Learning and a PhD student at Stanford working on end-to-end learning of Convolutional/Recurrent neural network architectures for images and text.

Abstract 14: **Micro-Perception Approach to Intelligent Transport, Ramesh Sarukkai (Lyft) in 2017 NIPS Workshop on Machine Learning for Intelligent Transportation Systems**, *Sarukkai* 02:30 PM

Abstract: 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.

Bio: Dr. Ramesh Sarukkai currently heads up the Geo teams (Mapping, Localization & Perception) at Lyft. Prior to that he was a Director of Engineering at Facebook and Google/YouTube where he led a number of platform & products initiatives including applied machine learning teams, consumer/advertising video products and core payments/risk/developer platforms. He has given a number of talks/keynotes/panelist at major conferences/workshops such as W3C WWW Conferences, ACM Multimedia, and published/presented papers at leading journals/conferences on internet technologies, speech/audio, computer vision and machine learning, in addition to authoring a book on “Foundations of Web Technology” (Kluwer/Springer). He also holds a large number of patents in the aforementioned areas and graduated with a PhD in computer science from the University of Rochester.

Abstract 16: **On Autonomous Driving: Challenges and Opportunities, Sertac Karaman, MIT in 2017 NIPS Workshop on Machine Learning for Intelligent Transportation Systems**, *Karaman* 03:30 PM

Abstract: Fully-autonomous driving has long been an sought after. DARPA’s efforts dating back a decade has ignited the first spark, showcasing the possibilities. Then, the AI revolution pushed the boundaries. These lead to the creation of a rapidly-growing ecosystem around developing self-driving capabilities. In this talk, we briefly summarize our experience in the DARPA Urban Challenge as Team MIT, one of the six finishers of the race. We then highlight few of our recent research results at MIT, including end-to-end deep learning for parallel autonomy and sparse-to-dense depth estimation for autonomous driving. We conclude with a few questions that may be relevant in the near future.

Bio: Sertac Karaman is an Associate Professor of Aeronautics and Astronautics at the Massachusetts Institute of Technology (since Fall 2012). He has obtained B.S. degrees in mechanical engineering and in computer engineering from the Istanbul Technical University, Turkey, in 2007; an S.M. degree in mechanical engineering from MIT in 2009; and a Ph.D. degree in electrical engineering and computer science also from MIT in 2012. His research interests lie in the broad areas of robotics and control theory. In particular, he studies the applications of probability theory, stochastic processes, stochastic geometry, formal methods, and optimization for the design and analysis of high-performance cyber-physical systems. The application areas of his research include driverless cars, unmanned aerial vehicles, distributed aerial surveillance systems, air traffic control, certification and verification of control systems software, and many others. He is the recipient of an IEEE Robotics and Automation Society Early Career Award in 2017, an Office of Naval Research (ONR) Young Investigator Award in 2017, Army Research Office (ARO) Young Investigator Award in 2015, National Science Foundation Faculty Career Development (CAREER) Award in 2014, AIAA Wright Brothers Graduate Award in 2012, and an NVIDIA Fellowship in 2011.

Abstract 17: **Generative Adversarial Imitation Learning, Stefano Ermon, Stanford in 2017 NIPS Workshop on Machine Learning for Intelligent Transportation Systems**, *Ermon* 04:00 PM

Abstract: Consider learning a policy from example expert behavior, without interaction with the expert or access to a reward or cost signal. One approach is to recover the expert's cost function with inverse reinforcement learning, then compute an optimal policy for that cost function. This approach is indirect and can be slow. In this talk, I will discuss a new generative modeling framework for directly extracting a policy from data, drawing an analogy between imitation learning and generative adversarial networks. I will derive a model-free imitation learning algorithm that obtains significant performance gains over existing methods in imitating complex behaviors in large, high-dimensional environments. Our approach can also be used to infer the latent structure of human demonstrations in an unsupervised way. As an example, I will show a driving application where a model learned from demonstrations is able to both produce different driving styles and accurately anticipate human actions using raw visual inputs.

Bio:

Stefano Ermon is currently an Assistant Professor in the Department of Computer Science at Stanford University, where he is affiliated with the Artificial Intelligence Laboratory. He completed his PhD in computer science at Cornell in 2015. His research interests include techniques for scalable and accurate inference in graphical models, large-scale combinatorial optimization, and robust decision making under uncertainty, and is motivated by a range of applications, in particular ones in the emerging field of computational sustainability. Stefano's research has won several awards, including three Best Paper Awards, a World Bank Big Data Innovation Challenge, and was selected by Scientific American as one of the 10 World Changing Ideas in 2016. He is a recipient of the Sony Faculty Innovation Award and NSF CAREER Award.

Abstract 18: **Deep Learning and photo-realistic Simulation for real-world Machine Intelligence in Autonomous Driving, Adrien Gaidon (TRI) in 2017 NIPS Workshop on Machine Learning for Intelligent Transportation Systems**, *Gaidon* 04:30 PM

Photo-realistic simulation is rapidly gaining momentum for visual training and test data generation in autonomous driving and general robotic contexts. This is particularly the case for video analysis, where manual labeling of data is extremely difficult or even impossible. This scarcity of adequate labeled training data is widely accepted as a major bottleneck of deep learning algorithms for important video understanding tasks like segmentation, tracking, and action recognition. In this talk, I will describe our use of modern game engines to generate large scale, densely labeled, high-quality synthetic video data with little to no manual intervention. In contrast to approaches using existing video games to record limited data from human game sessions, we build upon the more powerful approach of "virtual world generation". Pioneering this approach, the recent Virtual KITTI [1] and SYNTHIA [2] datasets are among the largest fully-labelled datasets designed to boost perceptual tasks in the context of autonomous driving and video understanding (including semantic and instance segmentation, 2D and 3D object detection and tracking, optical flow estimation, depth estimation, and structure from motion). With our recent PHAV dataset [3], we push the limits of this approach further by providing stochastic simulations of human actions, camera paths, and environmental conditions. I will describe our work on these synthetic 4D environments to automatically generate potentially infinite amounts of varied and realistic data. I will also describe how to measure and mitigate the domain gap when learning deep neural networks for different perceptual tasks needed for self-driving. I will finally show some recent results on more interactive simulation for autonomous driving and adversarial learning to automatically improve the output of simulators.

Abstract 19: **Panel Discussion in 2017 NIPS Workshop on Machine Learning for Intelligent Transportation Systems**, *Kahn, Sarukkai, Gaidon, Karaman* 05:00 PM

Dmitry Chichkov - NVIDIA
Adrien Gaidon - TRI
Gregory Kahn - Berkeley
Sertac Karaman - MIT
Ramesh Sarukkai - Lyft

Aligned Artificial Intelligence

Dylan Hadfield-Menell, Jacob Steinhardt, David Duvenaud, David Krueger, Anca Dragan

201 B, 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. Misaligned 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 whose behavior will be aligned with the values and goals of society at large? As AI capabilities develop, it is crucial for the AI community to come to satisfying and trustworthy answers to these questions. This workshop will focus on three central challenges in value alignment: learning complex rewards that reflect human preferences (e.g. meaningful oversight, preference elicitation, inverse reinforcement learning, learning from demonstration or feedback), engineering reliable AI systems (e.g. robustness to distributional shift, model misspecification, or adversarial data, via methods such as adversarial training, KWIK-style learning, or transparency to human inspection), and dealing with bounded rationality and incomplete information in both AI systems and their users (e.g. acting on incomplete task specifications, learning from users who sometimes make mistakes). We also welcome submissions that do not directly fit these categories but generally deal with problems relating to value alignment in artificial intelligence.

Schedule

09:15 AM	**Opening Remarks**	*Hadfield-Menell*
09:30 AM	**Dynamic Safe Interruptibility for Decentralized Multi-Agent Reinforcement Learning**	*Hendrikx*

09:45 AM	**Minimax-Regret Querying on Side Effects in Factored Markov Decision Processes**	*Singh*
10:15 AM	**Robust Covariate Shift with Exact Loss Functions**	*Liu*
11:00 AM	**Adversarial Robustness for Aligned AI**	*Goodfellow*
11:30 AM	**Incomplete Contracting and AI Alignment**	*Hadfield*
01:15 PM	**Learning from Human Feedback**	*Christiano*
01:45 PM	**Finite Supervision Reinforcement Learning**	*Saunders, Langlois*
02:00 PM	**Safer Classification by Synthesis**	*Wang*
02:15 PM	**Aligned AI Poster Session**	*Askell, Muszynski, Wang, Yang, Nguyen, Low, Jaillet, Schumann, Liu, Eckersley, Wang, Saunders*
03:30 PM	**Machine Learning for Human Deliberative Judgment**	*Evans*
04:00 PM	**Learning Reward Functions**	*Leike*
04:30 PM	**Informal Technical Discussion: Open Problems in AI Alignment**	

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

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

202, 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.
For more information visit here
For more information https://www.mltrain.cc/

Schedule

09:00 AM	**Lessons learned from designing Edward**	*Tran*
09:45 AM	**Tips and tricks of coding papers on PyTorch**	*Chintala*
10:05 AM	**Differentiable Learning of Logical Rules for Knowledge Base Reasoning**	*Cohen, Yang*
10:45 AM	**Coding Reinforcement Learning Papers**	*Zhang*
11:15 AM	**A Linear-Time Kernel Goodness-of-Fit Test (NIPS best paper)**	*Jitkrittum*
11:35 AM	**Imagination-Augmented Agents for Deep Reinforcement Learning**	*Racanière*
11:55 AM	**Inductive Representation Learning on Large Graphs**	*Hamilton*
12:15 PM	**Probabilistic Programming with PYRO**	*Goodman*
12:35 PM	**Poster Session**	
02:00 PM	**Simple and Efficient Implementation of Neural Nets with Automatic Operation Batching**	*Neubig*
02:45 PM	**Learning Texture Manifolds with the Periodic Spatial GAN by Nikolay Jetchev , Zalando**	*Vollgraf*
03:05 PM	**MLPACK, A case study: implementing ID3 decision trees to be as fast as possible**	*Curtin*
03:40 PM	**Self-Normalizing Neural Networks**	*Unterthiner*
04:00 PM	**Best of Both Worlds: Transferring Knowledge from Discriminative Learning to a Generative Visual Dialog Mode**	*Lu*
04:25 PM	**Break**	
05:00 PM	**Spotlights**	

Abstracts (2):

Abstract 10: **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.

Abstract 16: **Spotlights in NIPS Highlights (MLTrain), Learn How to code a paper with state of the art frameworks**, 05:00 PM

Ben Athiwaratkun: Bayesian GAN in Pytorch,

Dhyani Dushyanta: A Convolutional Encoder Model for Neural Machine Translation,

Forough Arabshahi: Combining Symbolic Expressions and Black-box Function Evaluations in Neural Programs,

Jean Kossaifi: Tensor Regression Networks with TensorLy and MXNet,

Joseph Paul Cohen: ShortScience.org - Reproducing Intuition,

Kamyar Azizzadenesheli: Efficient Exploration through Bayesian Deep Q-Networks,

Ashish Khetan: Learning from noisy, single-labeled data,

Rose Yu: Long-Term Forecasting using Tensor-Train RNNs

Shayenne da Lu Moura: Melody Transcription System

Tschannen Michael: Fast Linear Algebra in Stacked Strassen Networks

Yang Shi: Multimodal Compact Bilinear Pooling for Visual Question Answering

Yu-Chia Chen: Improved Graph Laplacian via Geometric Consistency

Learning Disentangled Features: from Perception to Control

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

203, 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

08:30 AM	**Welcome: Josh Tenenbaum**	*Tenenbaum*
09:00 AM	**Stefano Soatto**	*Soatto*
09:30 AM	**Irina Higgins**	*Higgins*
10:00 AM	**Finale Doshi-Velez**	*Doshi-Velez*
10:30 AM	**Poster session + Coffee break**	
11:00 AM	**Doris Tsao**	*Tsao*
11:30 AM	**Spotlight talks**	
12:15 PM	**Lunch**	
02:00 PM	**Doina Precup**	*Precup*
02:30 PM	**Pushmeet Kohli**	*Kohli*
03:00 PM	**Poster session + Coffee break**	*Kågebäck, Melnyk, Karimi, Brunner, Banijamali, Donahue, Zhao, Parascandolo, Thomas, Kumar, Burgess, Nilsson, Larsson, Eastwood, Peychev*
03:30 PM	**Yoshua Bengio**	*Bengio*
04:00 PM	**Ahmed Elgammal**	*Elgammal*
04:30 PM	**Final Poster Break**	
05:00 PM	**Panel discussion**	

Abstracts (1):

Abstract 2: **Stefano Soatto in Learning Disentangled Features: from Perception to Control**, *Soatto* 09:00 AM

Emergence of Invariance and Disentangling in Deep Representations

BigNeuro 2017: Analyzing brain data from nano to macroscale

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

204, 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

08:40 AM	**Opening Remarks**	*Dyer, Gray Roncal*
09:00 AM	**Can brain data be used to reverse engineer the algorithms of human perception?**	*DiCarlo*
09:35 AM	**Backpropagation and deep learning in the brain**	*Lillicrap*
09:55 AM	**Algorithms, tools, and progress in connectomic reconstruction of neural circuits**	*Jain*
10:15 AM	**Multimodal deep learning for natural human neural recordings and video**	*Brunton*
10:35 AM	**Coffee Break**	
11:00 AM	**More Steps towards Biologically Plausible Backprop**	*Bengio*
11:40 AM	**Panel on "What neural systems can teach us about building better machine learning systems"**	*Lillicrap, DiCarlo, Rozell, Jain, Kutz, Gray Roncal, Brunton*
12:20 PM	**Lunch Break**	
01:40 PM	**Discovery of governing equations and biological principles from spatio-temporal time-series recordings**	*Kutz*
02:15 PM	**Large scale calcium imaging data analysis for the 99%**	*Pnevmatikakis*
02:35 PM	**Machine learning for cognitive mapping**	*Varoquaux*
02:55 PM	**Dealing with clinical heterogeneity in the discovery of new biomarkers of brain disorders**	*Dansereau*
03:15 PM	**Morphological error detection for connectomics**	*Rolnick*
03:25 PM	**Poster Spotlights**	
03:45 PM	**Poster Session**	
04:30 PM	**Deep nets meet real neurons: pattern selectivity of V4 through transfer learning and stability analysis**	*Yu*
05:05 PM	**Mapping brain structure and function with deep learning**	*Calhoun*
05:25 PM	**bossDB: A Petascale Database for Large-Scale Neuroscience Informing Machine Learning**	*Drenkow*
05:45 PM	**Machine vision and learning for extracting a mechanistic understanding of neural computation**	*Branson*
06:05 PM	**Closing Panel: Analyzing brain data from nano to macroscale**	*Gray Roncal, Dyer*

Hierarchical Reinforcement Learning

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

Grand Ballroom A, 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 horizons, and more generally a scarcity of useful supervision signals. Unfortunately, the most valuable control tasks are specified in terms of high-level instructions, implying sparse rewards when formulated as an RL problem. Internal spatio-temporal abstractions and memory structures can constrain the decision space, improving data efficiency in the face of scarcity, but are likewise challenging for a supervisor to teach.

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 synergy among the research community and increase its impact by promoting better understanding of the challenges and potential of HRL. This workshop further aims to bring together researchers studying both theoretical and practical aspects of HRL, for a joint presentation, discussion, and evaluation of some of the numerous novel approaches to HRL developed in recent years.

Schedule

09:00 AM	**Opening Remarks**	*Fox*
09:10 AM	**Deep Reinforcement Learning with Subgoals (David Silver)**	*Silver*
09:40 AM	**Landmark Options Via Reflection (LOVR) in Multi-task Lifelong Reinforcement Learning (Nicholas Denis)**	*Denis*
09:50 AM	**Crossmodal Attentive Skill Learner (Shayegan Omidshafiei)**	*Omidshafiei*
10:00 AM	**HRL with gradient-based subgoal generators, asymptotically optimal incremental problem solvers, various meta-learners, and PowerPlay (Jürgen Schmidhuber)**	*Schmidhuber*
11:00 AM	**Meta-Learning Shared Hierarchies (Pieter Abbeel)**	*Abbeel*
11:30 AM	**Best Paper Award and Talk — Learning with options that terminate off-policy (Anna Harutyunyan)**	*Harutyunyan*
11:55 AM	**Spotlights & Poster Session**	*Abel, Denis, Eckstein, Fruit, Goel, Gruenstein, Harutyunyan, Klissarov, Kong, Kumar, Kumar, Liu, McNamee, Omidshafiei, Pitis, Rauber, Roderick, Shu, Wang, Zhang*
12:30 PM	**Lunch Break**	
01:30 PM	**Hierarchical Imitation and Reinforcement Learning for Robotics (Jan Peters)**	*Peters*
02:00 PM	**Deep Abstract Q-Networks (Melrose Roderick)**	*Roderick*
02:10 PM	**Federated Control with Hierarchical Multi-Agent Deep Reinforcement Learning (Saurabh Kumar)**	*Kumar*
02:20 PM	**Effective Master-Slave Communication On A Multi-Agent Deep Reinforcement Learning System (Xiangyu Kong)**	*Kong*
02:30 PM	**Sample efficiency and off policy hierarchical RL (Emma Brunskill)**	*Brunskill*
03:00 PM	**Coffee Break**	
03:30 PM	**Applying variational information bottleneck in hierarchical domains (Matt Botvinick)**	*Botvinick*
04:00 PM	**Progress on Deep Reinforcement Learning with Temporal Abstraction (Doina Precup)**	*Precup*
04:30 PM	**Panel Discussion**	*Botvinick, Brunskill, Campos, Peters, Precup, Silver, Tenenbaum, Fox*
05:30 PM	**Poster Session**	*Abel, Denis, Eckstein, Fruit, Goel, Gruenstein, Harutyunyan, Klissarov, Kong, Kumar, Kumar, Liu, McNamee, Omidshafiei, Pitis, Rauber, Roderick, Shu, Wang, Zhang*

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

Grand Ballroom B, 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, 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 constraints
-Data augmentation and/or the use of simulated data
- Frameworks that can tackle both very few samples and settings with more data without
extensive intervention.
- Effective and practical techniques for incorporating domain knowledge
- Applications of machine learning for small data problems in medical images and industry

Schedule

08:30 AM	**Welcome & Opening Remarks**	
08:40 AM	**Invited Talk: "Tales from fMRI: Learning from limited labeled data"**	*Varoquaux*
09:10 AM	**Invited Talk: Learning from Limited Labeled Data (But a Lot of Unlabeled Data)**	*Mitchell*
09:40 AM	**Contributed Talk 1: "Smooth Neighbors on Teacher Graphs for Semi-supervised Learning"**	
09:55 AM	**1-minute Poster Spotlights (Session #1)**	
10:15 AM	**Poster Sessions**	*Forster, Inouye, Srivastava, De Cock, Sharma, Kozinski, Babkin, he, Cui, Rao, Raskar, Das, Zhao, Lanka*
11:00 AM	**Invited Talk: "Light Supervision of Structured Prediction Energy Networks"**	*McCallum*
11:30 AM	**Invited Talk: "Forcing Neural Link Predictors to Play by the Rules", Sebastian Riedel**	
12:00 PM	**Lunch**	
02:00 PM	**Panel: Limited Labeled Data in Medical Imaging**	*Rubin, Lungren*
02:30 PM	**1-minute Poster Spotlights (Session #2)**	
02:50 PM	**Poster Session / Coffee Break**	*Ren, Lundquist, Hickson, Dubey, Shinoda, Marasovic, Stretcu, Bekele, Raunak, dos Santos, Canas, Mager Hois, Hirzel*
03:30 PM	**Invited Talk: Sample and Computationally Efficient Active Learning Algorithms**	*Balcan*
04:00 PM	**Contributed Talk 2: "EZLearn: Exploiting Organic Supervision in Large-Scale Data Annotation"**	
04:15 PM	**Invited Talk: Overcoming Limited Data with GANs**	*Goodfellow*
04:45 PM	**Invited Talk: Sameer Singh, "That Doesn't Make Sense! A Case Study in Actively Annotating Model Explanations"**	

05:15 PM	**Contributed Talk 3: Local Affine Approximators of Deep Neural Nets for Improving Knowledge Transfer**	
05:30 PM	**Contributed Talk 4: Co-trained Ensemble Models for Weakly Supervised Cyberbullying Detection**	
05:45 PM	**Invited Talk: What's so Hard About Natural Language Understanding?**	*Ritter*
06:15 PM	**Closing Remarks & Awards**	

Deep Learning: Bridging Theory and Practice

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

Hall 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.

Schedule

08:35 AM	**Opening Remarks**	
08:45 AM	**Invited Talk #1 (Yoshua Bengio)**	
09:15 AM	**Invited Talk #2 (Ian Goodfellow)**	
09:45 AM	**Spotlights 1**	
10:00 AM	**Invited Talk #3 (Peter Bartlett)**	
10:30 AM	**Coffee**	
11:00 AM	**Invited Talk #4 (Doina Precup)**	
11:30 AM	**Spotlights 2**	
11:45 AM	**Poster Session 1 and Lunch**	*Dathathri, Rangamani, Sharma, RoyChowdhury, Advani, Guss, Yun, Hardy, Alberti, Sachan, Veit, Shinozaki, chin*
01:30 PM	**Invited Talk #5 (Percy Liang)**	
02:00 PM	**Contributed Talks 1,2,3,4**	
03:00 PM	**Poster Session 2**	
04:00 PM	**Invited Talk #6 (Sham Kakade)**	
04:30 PM	**Panel**	

Abstracts (9):

Abstract 2: **Invited Talk #1 (Yoshua Bengio) in Deep Learning: Bridging Theory and Practice**, 08:45 AM

Generalization, Memorization and SGD

Abstract 3: **Invited Talk #2 (Ian Goodfellow) in Deep Learning: Bridging Theory and Practice**, 09:15 AM

Bridging Theory and Practice of GANs

Abstract 4: **Spotlights 1 in Deep Learning: Bridging Theory and Practice**, 09:45 AM

1) Generalization in deep nets: the role of distance from initialization
2) Entropy-SG(L)D optimizes the prior of a (valid) PAC-Bayes bound
3) Large Batch Training of DNNs with Layer-wise Adaptive Rate Scaling

Abstract 5: **Invited Talk #3 (Peter Bartlett) in Deep Learning: Bridging Theory and Practice**, 10:00 AM

Generalization in Deep Networks

Abstract 7: **Invited Talk #4 (Doina Precup) in Deep Learning: Bridging Theory and Practice**, 11:00 AM

Experimental design in Deep Reinforcement Learning

Abstract 8: **Spotlights 2 in Deep Learning: Bridging Theory and Practice**, 11:30 AM

1) Measuring robustness of NNs via Minimal Adversarial Examples
2) A classification based perspective on GAN-distributions
3) Learning one hidden layer neural nets with landscape design

Abstract 10: **Invited Talk #5 (Percy Liang) in Deep Learning: Bridging Theory and Practice**, 01:30 PM

Fighting Black Boxes, Adversaries, and Bugs in Deep Learning

Abstract 11: **Contributed Talks 1,2,3,4 in Deep Learning: Bridging Theory and Practice**, 02:00 PM

1) Don't Decay the Learning Rate, Increase the Batch Size
2) Meta-Learning and Universality: Deep Representations and Gradient Descent Can Approximate Any Learning Algorithm
3) Hyperparameter Optimization: A Spectral Approach
4) Learning Implicit Generative Models with Method of Learned Moments

Abstract 13: **Invited Talk #6 (Sham Kakade) in Deep Learning: Bridging Theory and Practice**, 04:00 PM

Towards Bridging Theory and Practice in DeepRL

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:
[1] - Kingma, DP and Welling, M, ''Auto-encoding variational bayes'', 2013.
[2] - Rezende, D, Mohamed, S, and Wierstra, D, ''Stochastic backpropagation and approximate inference in deep generative models'', 2014.
[3] - Blundell, C, Cornebise, J, Kavukcuoglu, K, and Wierstra, D, ''Weight uncertainty in neural network'', 2015.
[4] - Hernandez-Lobato, JM and Adams, R, ''Probabilistic backpropagation for scalable learning of Bayesian neural networks'', 2015.
[5] - Gal, Y and Ghahramani, Z, ''Dropout as a Bayesian approximation: Representing model uncertainty in deep learning'', 2015.
[6] - Gal, Y and Ghahramani, G, ''Bayesian convolutional neural networks with Bernoulli approximate variational inference'', 2015.
[7] - Kingma, D, Salimans, T, and Welling, M. ''Variational dropout and the local reparameterization trick'', 2015.
[8] - Balan, AK, Rathod, V, Murphy, KP, and Welling, M, ''Bayesian dark knowledge'', 2015.
[9] - Louizos, C and Welling, M, "Structured and Efficient Variational Deep Learning with Matrix Gaussian Posteriors", 2016.
[10] - Lawrence, ND and Quinonero-Candela, J, "Local distance preservation in the GP-LVM through back constraints", 2006.
[11] - Tran, D, Ranganath, R, and Blei, DM, "Variational Gaussian Process", 2015.
[12] - Neal, R, ''Bayesian Learning for Neural Networks'', 1996.
[13] - MacKay, D, ''A practical Bayesian framework for backpropagation networks'', 1992.
[14] - Dayan, P, Hinton, G, Neal, R, and Zemel, S, ''The Helmholtz machine'', 1995.
[15] - Wilson, AG, Hu, Z, Salakhutdinov, R, and Xing, EP, "Deep Kernel Learning", 2016.
[16] - Saatchi, Y and Wilson, AG, "Bayesian GAN", 2017.
[17] - MacKay, D.J.C. "Bayesian Methods for Adaptive Models", PhD thesis, 1992.

Schedule

08:05 AM	**Deep Probabilistic Programming**	*Tran*
08:30 AM	**TBD 1**	
08:45 AM	**Automatic Model Selection in BNNs with Horseshoe Priors**	*Doshi-Velez*

09:10 AM	**Deep Bayes for Distributed Learning, Uncertainty Quantification and Compression**	*Welling*
09:40 AM	**Poster spotlights**	
09:55 AM	**Discussion over coffee and poster session 1**	
10:55 AM	**Stochastic Gradient Descent as Approximate Bayesian Inference**	
11:20 AM	**TBD 2**	
11:35 AM	**TBD 2.5**	*Kalchbrenner*
12:00 PM	**Lunch**	
01:35 PM	**Deep Kernel Learning**	*Salakhutdinov*
02:00 PM	**TBD 3**	
02:15 PM	**Bayes by Backprop**	*Fortunato*
02:40 PM	**Discussion over coffee and poster session 2**	
03:35 PM	**How do the Deep Learning layers converge to the Information Bottleneck limit by Stochastic Gradient Descent?**	*Tishby*
04:00 PM	**Panel Session**	*Lawrence, Doshi-Velez, Ghahramani, LeCun, Welling, Teh, Winther*
05:00 PM	**Poster session**	*Zheng, Rudner, Tegho, McClure, Tang, D'CRUZ, Gamboa Higuera, Seelamantula, Arias Figueroa, Berlin, Voisin, Amini, DOAN, Hu, Botev, Sünderhauf, ZHANG, Lambert*

Workshop on Meta-Learning

Roberto Calandra, Frank Hutter, Hugo Larochelle, Sergey Levine

Hyatt Beacon Ballroom D+E+F+H, 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 improving current 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.

Schedule

08:30 AM	**Introduction and opening remarks**	*Calandra*
08:40 AM	**Learning to optimize with reinforcement learning**	*Malik*
09:10 AM	**Informing the Use of Hyperparameter Optimization Through Metalearning**	*Giraud-Carrier*
09:40 AM	**Poster Spotlight**	
10:00 AM	**Poster session (and Coffee Break)**	*Andreas, Li, Vercellino, Miconi, Zhang, Franceschi, Xiong, Ahmed, Itti, Klinger, Rohaninejad*
11:00 AM	**Invited talk: Jane Wang**	

11:30 AM	**Model-Agnostic Meta-Learning: Universality, Inductive Bias, and Weak Supervision**	*Finn*
01:30 PM	**Learn to learn high-dimensional models from few examples**	*Tenenbaum*
02:00 PM	**Multiple Adaptive Bayesian Linear Regression for Scalable Bayesian Optimization with Warm Start**	
02:15 PM	**Learning to Model the Tail**	
02:30 PM	**Poster session (and Coffee Break)**	
03:30 PM	**Meta Unsupervised Learning**	*Vinyals*
04:00 PM	**Panel Discussion**	

Abstracts (2):

Abstract 7: **Model-Agnostic Meta-Learning: Universality, Inductive Bias, and Weak Supervision in Workshop on Meta-Learning**, *Finn* 11:30 AM

Meta-learning holds the promise of enabling machine learning systems to replace manual engineering of hyperparameters and architectures, effectively reuse data across tasks, and quickly adapt to unexpected scenarios. In this talk, I will present a unified view of the meta-learning problem, discussing how a variety of approaches attempt to solve the problem, and when we might prefer some approaches over others. Further, I will discuss interesting theoretical and empirical properties of the model-agnostic meta-learning algorithm. Finally, I will conclude by showing new results on learning to learn from weak supervision with applications in imitation learning on a real robot and human-like concept acquisition.

Abstract 12: **Meta Unsupervised Learning in Workshop on Meta-Learning**, *Vinyals* 03:30 PM

In this talk I'll cover some recent work on few shot learning which we did at DeepMind. I'll describe how the work in MANN and Matching Networks influenced our most recent work on few shot learning for distributions, "Few-shot Autoregressive Density Estimation: Towards Learning to Learn Distributions".

Interpreting, Explaining and Visualizing Deep Learning - Now what ?

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

Hyatt Hotel, Regency Ballroom A+B+C, 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, Goebel*
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*
03:00 PM	**Coffee Break (afternoon)**	
03:30 PM	**Applications 2**	*Greydanus*
03:45 PM	**Invited Talk 6**	*Caruana*
04:15 PM	**Invited Talk 7**	*Darrell*
04:45 PM	**Closing Remarks**	*Hansen*

Optimal Transport and Machine Learning

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

Hyatt Hotel, Seaview Ballroom, 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

08:00 AM	**Structured Optimal Transport (with T. Jaakkola, S. Jegelka)**	*Alvarez Melis*
08:20 AM	**Approximate Bayesian computation with the Wasserstein distance**	*Jacob*
09:00 AM	**Gradient flow in the Wasserstein metric**	*Craig*
09:40 AM	**Approximate inference with Wasserstein gradient flows (with T. Poggio)**	*Frogner*
10:00 AM	**6 x 3 minutes spotlights**	*Flamary, Chen, Rujeerapaiboon, Adler, Lee, Roberts*
11:00 AM	**Optimal planar transport in near-linear time**	*Andoni*
11:40 AM	**Laplacian operator and Brownian motions on the Wasserstein space**	*Gangbo*
01:40 PM	**Geometrical Insights for Unsupervised Learning**	*Bottou*
02:20 PM	**Improving GANs Using Optimal Transport (with H. Zhang, A. Radford, D. Metaxas)**	*Salimans*
02:40 PM	**Overrelaxed Sinkhorn-Knopp Algorithm for Regularized Optimal Transport (with L. Chizat, C. Dossal, N. Papadakis)**	*THIBAULT*
03:30 PM	**Domain adaptation with optimal transport : from mapping to learning with joint distribution**	*Flamary*
04:10 PM	**Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance**	*Bach*
04:50 PM	**7 x 3 minutes spotlights**	*Cazelles, Genevay, Mena, Brauer, Fischer, Petzka, Seguy, Rolet, SONODA*
05:10 PM	**short Q&A session with plenary speakers**	
05:30 PM	**Closing session**	

Abstracts (9):

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, Rujeerapaiboon, Adler, Lee, Roberts* 10:00 AM

1. Nicolas Courty, Rémi Flamary and Mélanie Ducoffe. Learning Wasserstein Embeddings
2. Yongxin Chen, Tryphon Georgiou and Allen Tannenbaum. Optimal transport for Gaussian mixture models
3. Napat Rujeerapaiboon, Kilian Schindler, Daniel Kuhn and Wolfram Wiesemann. Size Matters: Cardinality-Constrained Clustering and Outlier Detection via Conic Optimization
4. Jonas Adler, Axel Ringh, Ozan Öktem and Johan Karlsson. Learning to solve inverse problems using Wasserstein loss
5. John Lee, Adam Charles, Nicholas Bertrand and Christopher Rozell. An Optimal Transport Tracking Regularizer
6. Lucas Roberts, Leo Razoumov, Lin Su and Yuyang Wang. Gini-regularized Optimal Transport with an Application in Spatio-Temporal Forecasting

Abstract 6: **Optimal planar transport in near-linear time in Optimal Transport and Machine Learning**, *Andoni* 11:00 AM

We show how to compute the Earth Mover Distance between two planar sets of size N in N^{1+o(1)} time. The algorithm is based on a generic framework that decomposes the natural Linear Programming formulation for the transport problem into a tree of smaller LPs, and recomposes it in a divide-and-conquer fashion. The main enabling idea is use *sketching* -- a generalization of the dimension reduction method -- in order to reduce the size of the "partial computation" so that the conquer step is more efficient. We will conclude with some open questions in the area.

This is joint work with Aleksandar Nikolov, Krzysztof Onak, and Grigory Yaroslavtsev.

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 $\mathbb{R}^d$ with Δ_w, a Laplacian operator.
A Brownian motion is shown to be consistent with the Laplacian operator. The smoothing
effect of the heat equation is established for a class of functions. Special perturbations of
the Laplacian operator, denoted $\Delta_{w,\epsilon}$, 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 maximum mean discrepancy. These insights allow us to give a fresh viewpoint on reported experimental results and to risk a couple 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 developent 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 ptf=(X,f(X)) 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 TUIA, 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 fromnindependent samples from μ approaches μ 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 ratesof 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

1. Jérémie Bigot, Elsa Cazelles and Nicolas Papadakis. Central limit theorems for Sinkhorn divergence between probability distributions on finite spaces.
2. Aude Genevay. Learning Generative Models with Sinkhorn Divergences
3. Gonzalo Mena, David Belanger, Gonzalo Muñoz and Jasper Snoek. Sinkhorn Networks: Using Optimal Transport Techniques to Learn Permutations
4. Christoph Brauer, Christian Clason, Dirk Lorenz and Benedikt Wirth. A Sinkhorn-Newton method for entropic optimal transport
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
7. Sho Sonoda and Noboru Murata. Transportation analysis of denoising autoencoders: a novel method for analyzing deep neural networks

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

Hyatt Hotel, Shoreline, Sat Dec 09, 08:00 AM

For more please visit
http://www.katherinelgorman.com/nips-2017-collaboration-and-communication
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 confounding, and for those who have never worked with someone from a different field, it 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 collaboaration 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 will 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

09:00 AM	**DUE TO LOW ATTENDANCE THIS WORKSHOP IS CANCELED**

Machine Learning Challenges as a Research Tool

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

S1, 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 (coopetitions), 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

08:00 AM	**Introduction - Isabelle Guyon and Evelyne Viegas**	*Guyon*
08:10 AM	**Baázs Kégl, RAMP platform**	*Kégl*
08:40 AM	**Automatic evaluation of chatbots - Varvara Logacheva**	
09:10 AM	**TrackML - David Rousseau**	

09:40 AM	**Data Science Bowl - Andre Farris**	*Farris*
10:10 AM	**CrowdAI - Mohanty Sarada**	
10:40 AM	**Break, poster viewing 1**	
11:00 AM	**Ben Hamner, Kaggle platform**	*Hamner*
11:30 AM	**Incentivizing productivity in data science - Jacob Abernethy**	
12:20 PM	**Lunch Break, poster viewing 2**	
12:20 PM	**Establishing Uniform Image Segmentation - Maximilian Chen and Michael C. Darling**	*Chen, Darling*
01:30 PM	**Welcome - Isabelle Guyon and Evelyne Viegas**	
01:40 PM	**Project Malmo, Minecraft - Katja Hofmann**	
02:10 PM	**Project Alloy - Laura Seaman and Israel Ridgley**	*Ridgley, Seaman*
02:40 PM	**Education and public service - Jonathan C. Stroud**	
03:10 PM	**Break, poster viewing 3**	
03:30 PM	**AutoDL (Google challenge) - Olivier Bousquet**	
04:00 PM	**Scoring rule markets - Rafael Frongillo and Bo Waggoner**	
04:30 PM	**ENCODE-DREAM challenge - Akshay Balsubramani**	
05:00 PM	**Codalab platform, Xavier Baro**	
05:30 PM	**New opportunities for challenges, looking to the future - Sergio Escalera**	
06:20 PM	**Wrap up -Evelyne Viegas and Isabelle Guyon**	

Emergent Communication Workshop

Jakob Foerster, Igor Mordatch, Angeliki Lazaridou, Kyunghyun Cho, Douwe Kiela, Pieter Abbeel

S4, 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/

Schedule

08:50 AM	**Opening Remarks**	
09:00 AM	**"Language in context"**	*Goodman*
09:30 AM	**"Communication via Physical Action"**	*Dragan*
10:00 AM	**Invited Talk 3**	*Kohli*
10:30 AM	**Coffee Break**	
11:00 AM	**Morning panel discussion**	*Schmidhuber, Goodman, Dragan, Kohli, Batra*
12:15 PM	**Poster Session and Lunch**	
02:15 PM	**Contributed Talks 1**	*Resnick, Wen, Zheng, Bhutani, Choi*
03:00 PM	**Coffee Break + Poster Presentation**	
03:30 PM	**Contributed Talks 2**	*Raileanu, Kottur, Grouchy*
04:00 PM	**"From Democritus to Signaling Networks"**	*Skyrms*
04:30 PM	**"Language Emergence as Boundedly Optimal Control"**	*Singh*
05:00 PM	**Afternoon Panel discussion**	*Skyrms, Singh, Andreas*
06:15 PM	**Closing Remarks**	

Abstracts (4):

Abstract 6: **Morning panel discussion in Emergent Communication Workshop**, *Schmidhuber, Goodman, Dragan, Kohli, Batra* 11:00 AM

All speakers from the morning session and further invited panelists.

Abstract 8: **Contributed Talks 1 in Emergent Communication Workshop**, *Resnick, Wen, Zheng, Bhutani, Choi* 02:15 PM

- Vehicle Communication Strategies for Simulated Highway Driving. Cinjon Resnick*, NYU; Ilia Kulikov, NYU; Kyunghyun Cho, New York University; Jason Weston, FAIR

- Multiagent Bidirectionally-Coordinated Nets: Emergence of Human-level Coordination in Learning to Play StarCraft Combat Games. Ying Wen*, UCL; Jun Wang, UCL

- A two dimensional decomposition approach for matrix completion through gossip. Mukul Bhutani*, Amazon; Bamdev Mishra, Amazon.com

- Multi-Agent Compositional Communication Learning from Raw Visual Input. Edward Choi*, Georgia Institute of Technology; Angeliki Lazaridou, DeepMind; Nando de Freitas, DeepMind

- Analyzing Language Learned by an Active Question Answering Agent. Jannis Bulian*, Google; Christian Buck, Google; Massimiliano Ciaramita, Google; Wojciech Gajewski, Google; Andrea Gesmundo, Google; Neil Houlsby, Google; Wei Wang, Google

- MACE: Structured Exploration via Deep Hierarchical Coordination. Stephan Zheng*, Caltech; Yisong Yue, Caltech

- Multi-agent Communication by Bi-directional Recurrent Neural Networks. Ying Wen*, UCL; Yaodong Yang, University College London; Jun Wang, UCL

Abstract 10: **Contributed Talks 2 in Emergent Communication Workshop**, *Raileanu, Kottur, Grouchy* 03:30 PM

- Modeling Opponent Hidden States in Deep Reinforcement Learning. Roberta Raileanu*, NYU; Arthur Szlam, FAIR; Rob Fergus, FAIR

- Emergent Communication through Negotiatio.n. Kris Cao*, University of Cambridge; Angeliki Lazaridou, DeepMind; Stephen Clark, University of Cambridge/Deepmind; Joel Z Leibo, DeepMind; Marc Lanctot, Deepmind; Karl Tuyls, Deepmind

-Natural Language Does Not Emerge 'Naturally' in Multi-Agent Dialog. Satwik Kottur*, Carnegie Mellon University; Satwik Kottur; José M.F. Moura; Stefan Lee; Dhruv Batra (EMNLP 2017)

- Emergence of Language with Multi-agent Games: Learning to Communicate with Sequences of Symbols. Serhii Havrylov*, Institute for Language, Cognition and Computation, the University of Edinburgh; Ivan Titov, Institute for Language, Cognition and Computation, the University of Edinburgh (NIPS 2017)

- An unbroken evolutionary pathway to symbolic communication. Paul Grouchy*, University of Toronto; Paul Grouchy, Gabriele M. T. D'Eleuterio; Morten H. Christiansen; Hod Lipson (Nature Scientific Reports)

Abstract 13: **Afternoon Panel discussion in Emergent Communication Workshop**, *Skyrms, Singh, Andreas* 05:00 PM

Brian Skyrms, Satinder Singh, Jacob Andreas

Bayesian optimization for science and engineering

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

S7, 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

09:00 AM	**Introduction**	
09:10 AM	**Invited talk: Towards Safe Bayesian Optimization**	*Krause*
09:40 AM	**Invited talk: Learning to learn without gradient descent by gradient descent.**	*Chen*
10:10 AM	**Poster spotlights 1**	
11:00 AM	**Invited talk: Scaling Bayesian Optimization in High Dimensions**	*Jegelka*
11:30 AM	**Poster spotlights 2**	
02:00 PM	**Invited talk: Neuroadaptive Bayesian Optimization - Implications for Cognitive Sciences**	*Lorenz*
02:30 PM	**Poster session**	
04:00 PM	**Invited talk: Knowledge Gradient Methods for Bayesian Optimization**	*Frazier*
04:30 PM	**Invited talk: Quantifying and reducing uncertainties on sets under Gaussian Process priors**	*Ginsbourger*
04:55 PM	**Panel**	

Teaching Machines, Robots, and Humans

Maya Cakmak, Anna Rafferty, Adish Singla, Jerry Zhu, Sandra Zilles

Seaside Ballroom, Sat Dec 09, 08:00 AM

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]

* there are strong theoretical connections between the Teaching-dimension (the sample complexity of teaching) and the VC-dimension (the sample complexity of learning from randomly chosen examples). [Doliwa et al. 2014]

* machine teaching problem formulation has been recently studied in the context of diverse applications including personalized educational systems, cyber-security problems, robotics, program synthesis, human-in-the-loop systems, and crowdsourcing. [Jha et al. 2016; Zhu 2015; Mei & Zhu 2015; Ba & Caruana 2014; Patil et al. 2014; Singla et al. 2014; Cakmak & Thomaz 2014]

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

09:00 AM	**Overview of Machine Teaching**

10:00 AM	**"Reinforcement Learning with People" - Emma Brunskill (Stanford)**
11:00 AM	**"Active Classification with Comparison Queries" - Shay Moran (Technion)**
11:30 AM	**Discussion session (Topic: Open questions and new research directions)**
02:00 PM	**"Iterative Machine Teaching" - Le Song (Georgia Tech)**
02:30 PM	**Spotlight presentations for posters (14 papers)**
03:30 PM	**Poster presentations (14 papers)**
04:30 PM	**"Machine Teaching: A New Paradigm for Building Machine Learning Systems" - Patrice Simard (Microsoft Research)**
05:00 PM	**"Improving Language Learning and Assessment with Data" - Burr Settles (Duolingo)**
05:30 PM	**Discussion session (Topic: Novel applications and industry insights)**