WORKSHOP PROGRAM

Thu Aug 10th - Fri Aug 11th

WORKSHOP CHAIRS: ANIMA ANANDKUMAR
                      FEI SHA
Workshop organizers make last-minute changes to their schedule. Download this document again to get the latest changes, or use the ICML mobile application.

Schedule Highlights

**Aug. 10, 2017**

C4.1. Lifelong Learning: A Reinforcement Learning Approach
Chandar, Ravindran, Mankowitz, Mannor, Zahavy

C4.10. ICML Workshop on Machine Learning for Autonomous Vehicles 2017
Li, Urtasun, Gray, Savarese

C4.11. Learning to Generate Natural Language
Miao, Ling, Wen, Cao, Gerz, Blunsom, Dyer

C4.3. Workshop on Visualization for Deep Learning
Jiang, Canny, Chau, Fan, Zhu

C4.4. Workshop on Computational Biology
Pe'er, Leslie, Azizi, Prabhakaran, Kshirsagar, Carr

C4.5. Principled Approaches to Deep Learning
Pronobis, Gens, Kakade, Domingos

C4.6. Video Games and Machine Learning
Synnaeve, Togelius, Schaul, Vinyals, Usunier

C4.7. ML on a budget: IoT, Mobile and other tiny-ML applications
Varma, Saligrama, Jain

C4.8. Workshop on Human Interpretability in Machine Learning (WHI)
Varshney, Weller, Kim, Malioutov

Vanschoren, Garnett

Parkside 1, Implicit Generative Models
Ranganath, Goodfellow, Tran, Blei, Lakshminarayanan, Mohamed

**Aug. 11, 2017**

C4.1. Time Series Workshop
Kuznetsov, Liu, Yang, Yu

C4.10. Reproducibility in Machine Learning Research
Ke, Goyal, Lamb, Pineau, Bengio, Bengio

C4.11. Interactive Machine Learning and Semantic Information Retrieval
Glowacka, Buntine, Myllymaki

C4.3. Machine Learning in Speech and Language Processing
Livescu, Sainath, Lu, Ragni

C4.4. Private and Secure Machine Learning
Honkela, Shimizu, Kaski

C4.5. Deep Structured Prediction
Augenstein, Chang, Chechik, Huang, Torres Martins, Meshi, Schwing, Miao

Cortes, Chaudhuri, DeSalvo, Zhang, Zhang

C4.7. Reliable Machine Learning in the Wild
Hadfield-Menell, Steinhardt, Weller, Milli

C4.8. Human in the Loop Machine Learning
Nock, Ong

C4.9. Machine Learning for Music Discovery
Schmidt, Nieto, Gouyon, Lanckriet

Parkside 1, Reinforcement Learning Workshop
Precup, Ravindran, Bacon
Lifelong Learning: A Reinforcement Learning Approach

Sarath Chandar, Balaraman Ravindran, Daniel J. Mankowitz, Shie Mannor, Tom Zahavy

C4.1, Thu Aug 10, 08:30 AM

One of the most challenging and open problems in Artificial Intelligence (AI) is that of Lifelong Learning:

"Lifelong Learning is the continued learning of tasks, from one or more domains, over the course of a lifetime, by a lifelong learning system. A lifelong learning system efficiently and effectively (1) retains the knowledge it has learned; (2) selectively transfers knowledge to learn new tasks; and (3) ensures the effective and efficient interaction between (1) and (2)."

Lifelong learning is still in its infancy. Many issues currently exist such as learning general representations, catastrophic forgetting, efficient knowledge retention mechanisms and hierarchical abstractions. Much work has been done in the Reinforcement Learning (RL) community to tackle different elements of lifelong learning. Active research topics include hierarchical abstractions, transfer learning, multi-task learning and curriculum learning. With the emergence of powerful function approximators such as in Deep Learning, we feel that now is a perfect time to provide a forum to discuss ways to move forward and provide a truly general lifelong learning framework, using RL-based algorithms, with more rigour than ever before. This workshop will endeavour to promote interaction between researchers working on the different elements of lifelong learning to try and find a synergy between the various techniques.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30 AM</td>
<td>Introduction and Overview</td>
</tr>
<tr>
<td>08:40 AM</td>
<td>Marc G. Bellemare: The role of density models in reinforcement learning Bellemare</td>
</tr>
<tr>
<td>09:20 AM</td>
<td>Poster Spotlights</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Poster Session + break - I</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Joelle Pineau: A few modest insights from my lifelong learning Pineau</td>
</tr>
<tr>
<td>11:10 AM</td>
<td>Andrei Rusu: Sequential Learning in Complex Environments Rusu</td>
</tr>
<tr>
<td>12:00 PM</td>
<td>Lunch</td>
</tr>
<tr>
<td>02:00 PM</td>
<td>Contributed Talks</td>
</tr>
<tr>
<td>02:45 PM</td>
<td>Poster Session + break - II</td>
</tr>
</tbody>
</table>

ICML Workshop on Machine Learning for Autonomous Vehicles 2017

Li Erran Li, Raquel Urtasun, Andrew Gray, Silvio Savarese

C4.10, Thu Aug 10, 08:30 AM

Although dramatic progress has been made in the field of autonomous driving, there are many major challenges in achieving full-autonomy. For example, how to make perception accurate and robust to accomplish safe autonomous driving? How to reliably track cars, pedestrians, and cyclists? 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 to achieve near-zero fatality?

These complex challenges associated with autonomy in physical world naturally suggest that we take a machine learning approach. Deep learning and computer vision have found many real-world applications such as face tagging. However, perception for autonomous driving has a unique set of requirements such as safety and explainability. Autonomous vehicles need to choose actions, e.g. steering commands which will affect the subsequent inputs (driving scenes) encountered. This setting is well-suited to apply reinforcement learning to determine the best actions to take. Many autonomous driving tasks such as perception and tracking requires large data sets of labeled examples to learn rich and high-performance visual representation. However, the progress is hampered by the sheer expenses of human labelling needed. Naturally we would like to employ unsupervised learning, transfer learning leveraging simulators, and techniques can learn efficiently. The goal of this workshop is to bring together researchers and practitioners from in the field of autonomous driving 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.

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 200 participants from leading university researchers, auto-companies and ride-sharing companies.
Schedule

08:20 AM
Opening Remarks: Drew Gray and Li Erran Li (Uber ATG)

08:30 AM
Carl Wellington, Uber ATG

09:00 AM
Efficient deep neural networks for perception in autonomous driving (Jose M. Alvarez, TRI)

09:30 AM
Visual 3D Scene Understanding and Prediction for ADAS (Manmohan Chandraker, NEC Labs)

10:00 AM
Coffee

10:30 AM
2 x 15 Contributed Talks on Datasets and Occupancy Maps

11:00 AM
Beyond Hand Labeling: Simulation and Self-Supervision for Self-Driving Cars (Matt Johnson, University of Michigan)

11:30 AM
Learning Affordance for Autonomous Driving (JianXiong Xiao, AutoX)

12:00 PM
Lunch

02:00 PM
Deep Reinforcement Learning for Real-World Mobility (Sergey Levine, UC Berkeley)

02:30 PM
2 x 15 Contributed Talks on Reinforcement Learning

03:00 PM
Coffee and Posters

03:30 PM
Min Sun, National Tsing Hua University: Assessing Risk and Adapting Changes on the Road

04:00 PM
Are we over-engineering autonomous vehicles? (Amar Shah, University of Cambridge)

04:30 PM
2 x 5 min Lightening Talks

04:40 PM
Panel Discussion (Jose M. Alvarez, Manmohan Chandraker, Matt Johnson, Min Sun, Carl Wellington)

05:25 PM
Closing Remarks: Li Erran Li and Drew Gray (Uber ATG)

Abstracts (10):

Abstract 3: Efficient deep neural networks for perception in autonomous driving (Jose M. Alvarez, TRI)

Abstract
Convolutional neural networks have achieved impressive success in many tasks in computer vision such as image classification, object detection/reognition or semantic segmentation. While these networks have proven effective in all these applications, they come at a high memory and computational cost, thus not feasible for applications where power and computational resources are limited. In addition, the process to train the network reduces productivity as it not only requires large computer servers but also takes a significant amount of time (several weeks) with the additional cost of engineering the architecture. In this talk, I first introduce our efficient architecture based on filter-compositions and then, a novel approach to jointly learn the architecture and explicitly account for compression during the training process. Our results show that we can learn much more compact models and significantly reduce training and inference time.

Bio:
Dr. Jose Alvarez is a senior research scientist at Toyota Research Institute. His main research interests are in developing robust and efficient deep learning algorithms for perception with focus on autonomous vehicles. Previously, he was a researcher at Data61 / CSIRO (formerly NICTA), a Postdoctoral researcher at the Courant Institute of Mathematical Science, New York University, and visiting scholar at University of Amsterdam and Group Research Electronics at Volkswagen. Dr. Alvarez graduated in 2012 and he was awarded the best Ph.D. Thesis award. Dr. Alvarez serves as associate editor for IEEE Trans. on Intelligent Transportation Systems.

Abstract 4: Visual 3D Scene Understanding and Prediction for ADAS

Abstract:
Modern advanced driver assistance systems (ADAS) rely on a range of sensors including radar, ultrasound, LIDAR and cameras. Active sensors have found applications in detecting traffic participants (TPs) such as cars or pedestrians and scene elements (SEs) such as roads. However, camera-based systems have the potential to achieve or augment these capabilities at a much lower cost, while allowing new ones such as determination of TP and SE semantics as well as their interactions in complex traffic scenes.

In this talk, we present several technical advances for vision-based ADAS. A common theme is to overcome the challenges posed by lack of large-scale annotations in deep learning frameworks. We introduce approaches to correspondence estimation that are trained on purely synthetic data but adapt well to real data at test-time. We introduce object detectors that are light enough for ADAS, trained with knowledge distillation to retain accuracies of deeper architectures. Our semantic segmentation methods are trained on weak supervision that requires only a tenth of conventional annotation time. We propose methods for 3D reconstruction that use deep supervision to recover fine TP part locations while relying on purely synthetic 3D CAD models. We develop deep
learning frameworks for multi-target tracking, as well as occlusion-reasoning in TP localization and SE layout estimation. Finally, we present a framework for TP behavior prediction in complex traffic scenes that accounts for TP-TP and TP-SE interactions. Our approach allows prediction of diverse multimodal outcomes and aims to account for long-term strategic behaviors in complex scenes.

Bio:
Manmohan Chandraker is an assistant professor at the CSE department of the University of California, San Diego and leads the computer vision research effort at NEC Labs America in Cupertino. He received a B.Tech. in Electrical Engineering at the Indian Institute of Technology, Bombay and a PhD in Computer Science at the University of California, San Diego. His personal research interests are 3D scene understanding and reconstruction, with applications to autonomous driving and human-computer interfaces. His works have received the Marr Prize Honorable Mention for Best Paper at ICCV 2007, the 2009 CSE Dissertation Award for Best Thesis at UCSD, a PAMI special issue on best papers of CVPR 2011 and the Best Paper Award at CVPR 2014.

Abstract 6: 2 x 15 Contributed Talks on Datasets and Occupancy Maps in ICML Workshop on Machine Learning for Autonomous Vehicles 2017, 10:30 AM

Jonathan Binas, Daniel Neil, Shih-Chii Liu, Tobi Delbruck, DDD17: End-To-End DAVIS Driving Dataset

Ransalu Senanayake and Fabio Ramos, Bayesian Hilbert Maps for Continuous Occupancy Mapping in Dynamic Environments

Abstract 7: Beyond Hand Labeling: Simulation and Self-Supervision for Self-Driving Cars (Matt Johnson, University of Michigan) in ICML Workshop on Machine Learning for Autonomous Vehicles 2017, 11:00 AM

Self-driving cars now deliver vast amounts of sensor data from large unstructured environments. In attempting to process and interpret this data there are many unique challenges in bridging the gap between prerecorded data sets and the field. This talk will present recent work addressing the application of deep learning techniques to robotic perception. We focus on solutions to several pervasive problems when attempting to deploy such techniques on fielded robotic systems. The themes of the talk revolve around alternatives to gathering and curating data sets for training. Are there ways of avoiding the labor-intensive human labeling required for supervised learning? These questions give rise to several lines of research based around self-supervision, adversarial learning, and simulation. We will show how these approaches applied to self-driving car problems have great potential to change the way we train, test, and validate machine learning-based systems.

Bio:
Matthew Johnson-Roberson is Assistant Professor of Engineering in the Department of Naval Architecture & Marine Engineering and the Department of Electrical Engineering and Computer Science at the University of Michigan. He received a PhD from the University of Sydney in 2010. He has held prior postdoctoral appointments with the Centre for Autonomous Systems - CAS at KTH Royal Institute of Technology in Stockholm and the Australian Centre for Field Robotics at the University of Sydney. He is a recipient of the NSF CAREER award (2015). He has worked in robotic perception since the first DARPA grand challenge and his group focuses on enabling robots to better see and understand their environment.

Abstract 8: Learning Affordance for Autonomous Driving (JianXiong Xiao, AutoX) in ICML Workshop on Machine Learning for Autonomous Vehicles 2017, 11:30 AM

Today, there are two major paradigms for vision-based autonomous driving systems: mediated perception approaches that parse an entire scene to make a driving decision, and behavior reflex approaches that directly map an input image to a driving action by a regressor. In this paper, we propose a third paradigm: a direct perception based approach to estimate the affordance for driving. We propose to map an input image to a small number of key perception indicators that directly relate to the affordance of a road/traffic state for driving.

Our representation provides a set of compact yet complete descriptions of the scene to enable a simple controller to drive autonomously. Falling in between the two extremes of mediated perception and behavior reflex, we argue that our direct perception representation provides the right level of abstraction. We evaluate our approach in a virtual racing game as well as real world driving and show that our model can work well to drive a car in a very diverse set of virtual and realistic environments.

Jianxiong Xiao (a.k.a., Professor X) is the Founder and CEO of AutoX Inc., a high-tech startup working on A.I. software solution for self-driving vehicles. AutoX’s mission is to democratize autonomy and make autonomous driving universally accessible to everyone. Its innovative camera-first self-driving solution amounts to only a tiny fraction of the cost of traditional LiDar-based approaches. Dr. Xiao has over ten years of research and engineering experience in Computer Vision, Autonomous Driving, and Robotics. In particular, he is a pioneer in the fields of 3D Deep Learning, RGB-D Recognition and Mapping, Big Data, Large-scale Crowdsourcing, and Deep Learning for Robotics. Jianxiong received a BEng, and MPhil in Computer Science from the Hong Kong University of Science and Technology in 2009. He received his Ph.D. from the Computer Science and Artificial Intelligence Laboratory (CSAIL) at the Massachusetts Institute of Technology (MIT) in 2013. And he was an Assistant Professor at Princeton University and the founding director of the Princeton Computer Vision and Robotics Labs from 2013 to 2016. His work has received the Best Student Paper Award at the European Conference on Computer Vision (ECCV) in 2012 and the Google Research Best Papers Award for 2012, and has appeared in the popular press. He was awarded the Google U.S./Canada Fellowship in Computer Vision in 2012, the MIT CSW Best Research Award in 2011, NSF/Intel VEC Research Award in 2016, and two Google Faculty Awards in 2014 and in 2015 respectively. He co-lead the MIT+Princeton joint team to participate in the Amazon Picking Challenge in 2016, and won the 3rd and 4th place worldwide. More information can be found at: http://www.jianxiongxiao.com.

Abstract 10: Deep Reinforcement Learning for Real-World Mobility (Sergey Levine, UC Berkeley) in ICML Workshop on Machine Learning for Autonomous Vehicles 2017, 02:00 PM

Abstract:
Deep reinforcement learning algorithms can acquire complex tasks using their own autonomously collected experience. However, applications of deep reinforcement learning to real world tasks have been limited by a number of challenging obstacles: (1) reinforcement learning algorithms tend to require large amounts of experience, often much larger than
equivalent supervised learning methods; (2) reinforcement learning often requires risky exploration, which can be extremely dangerous in safety-critical applications such as robotic flight or driving; (3) reinforcement learning methods suffer from problems with stability and convergence. In this talk, I will discuss some of our recent work on making it feasible to use reinforcement learning to train robotic systems that perform well in the real world. Specifically, I will cover methods for transferring skills from simulation to the real world, safe uncertainty-aware exploration methods, and meta-learning algorithms that can dramatically accelerate reinforcement learning.

Bio:

Sergey Levine received a BS and MS in Computer Science from Stanford University in 2009, and a Ph.D. in Computer Science from Stanford University in 2014. He joined the faculty of the Department of Electrical Engineering and Computer Sciences at UC Berkeley in fall 2016. His work focuses on machine learning for decision making and control, with an emphasis on deep learning and reinforcement learning algorithms. Applications of his work include autonomous robots and vehicles, as well as computer vision and graphics. His research includes developing algorithms for end-to-end training of deep neural network policies that combine perception and control, scalable algorithms for inverse reinforcement learning, deep reinforcement learning algorithms, and more.

Abstract 11: 2 x 15 Contributed Talks on Reinforcement Learning in ICML Workshop on Machine Learning for Autonomous Vehicles 2017, 02:30 PM

David Isele, Akansel Cosgun, To Go or Not to Go: A Case for Q-Learning at Unsignalized Intersections

Tomoki Nishi, Prashant Doshi, Danil Prokhorov, Freeway Merging in Congested Traffic based on Multipolicy Decision Making with Passive Actor Critic

Abstract 13: Min Sun, National Tsing Hua University: Assessing Risk and Adapting Changes on the Road in ICML Workshop on Machine Learning for Autonomous Vehicles 2017, 03:30 PM

It is critical for a self-driving car in the wild to assess risk and adapt to changes on the road. In this talk, we will first go over our proposed accident anticipation method which is tested on a large dataset consisting of real-world accident videos. Then, we will present our latest ICCV paper about how to adapt a semantic segmentation model across 4 cities in three continents.

Bio: Min Sun is an assistant professor at National Tsing Hua University in Taiwan. Before that, he was a postdoctoral researcher at Washington University in Seattle and he graduated from the University of Michigan with a Ph.D. degree in EE: System. He also won the best paper award of 3dRR in 2007 and best paper award of CVGIP in 2015 and 2016.

Abstract 14: Are we over-engineering autonomous vehicles? (Amar Shah, University of Cambridge) in ICML Workshop on Machine Learning for Autonomous Vehicles 2017, 04:00 PM

Research, media and corporate interest in autonomous road vehicles has exploded. Large industrial labs are hiring large teams to engineer solutions to the problem. As machine learners, I conjecture that it will be big data and cutting edge end-to-end reinforcement learning techniques that will have the best chance to achieve autonomy over the next decade and not the vastly hand engineered, rule based approaches of current teams. A crucial component of self driving car systems which is overlooked is model uncertainty. I claim that model uncertainty is as important as model accuracy, and encourage the community to pursue research in this direction. Good decisions require good predictions and well calibrated uncertainty estimates around those predictions.

Abstract 15: 2 x 5 min Lightening Talks in ICML Workshop on Machine Learning for Autonomous Vehicles 2017, 04:30 PM

1. Kangwook Lee, Hoon Kim, Changho Suh, Crash To Not Crash: Playing Video Games To Predict Vehicle Collisions


Ahmad El Sallab, Mahmoud Saeed, Omar Abdel Tawab, Mohammed Abdou, Meta learning Framework for Automated Driving

Learning to Generate Natural Language

Yishu Miao, Wang Ling, Tsung-Hsien Wen, Kris Cao, Daniela Gerz, Phil Blunsom, Chris Dyer

C4.11, Thu Aug 10, 08:30 AM

Research on natural language generation is rapidly growing due to the increasing demand for human-machine communication in natural language. This workshop aims to promote the discussion, exchange, and dissemination of ideas on the topic of text generation, touching several important aspects in this modality: learning schemes and evaluation, model design and structures, advanced decoding strategies, and natural language generation applications. This workshop aims to be a venue for the exchange of ideas regarding data-driven machine learning approaches for text generation, including mainstream tasks such as dialogue generation, instruction generation, and summarization; and for establishing new directions and ideas with potential for impact in the fields of machine learning, deep learning, and NLP.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30 AM</td>
<td>Tim Baldwin: Learning to Label Documents</td>
</tr>
<tr>
<td>09:15 AM</td>
<td>Dani Yogatama</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Coffee Break &amp; Poster session 1</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Andre Martins: Beyond Softmax: Sparsemax, Constrained Softmax,</td>
</tr>
<tr>
<td></td>
<td>Differentiable Easy-First</td>
</tr>
<tr>
<td>11:15 AM</td>
<td>Spotlight Paper Presentation</td>
</tr>
<tr>
<td>12:00 PM</td>
<td>Lunch Break &amp; Poster session 2</td>
</tr>
</tbody>
</table>
In the first part of the talk, I will propose sparsemax, a new activation function similar to the traditional softmax, but able to output sparse probabilities. After deriving its properties, I will show how its Jacobian can be efficiently computed, enabling its use in a network trained with backpropagation. Then, I will propose a new smooth and convex loss function which is the sparsemax analogue of the logistic loss, revealing an unexpected connection with the Huber classification loss. I will show promising empirical results in multi-label classification problems and in attention-based neural networks for natural language inference.

In the second part, I will introduce constrained softmax, another activation function that allows imposing upper bound constraints on attention probabilities. Based on this activation, I will introduce a novel neural end-to-end differentiable easy-first decoder that learns to solve sequence tagging tasks in a flexible order. The decoder iteratively updates a "sketch" of the predictions over the sequence. The proposed models compare favourably to BILSTM taggers on three sequence tagging tasks.

This is joint work with Ramon Astudillo and Julia Kreutzer.

Abstract 5: Spotlight Paper Presentation in Learning to Generate Natural Language, 11:15 AM

Workshop Paper Presentation

Abstract 6: Lunch Break & Poster session 2 in Learning to Generate Natural Language, 12:00 PM

Lunch Break & Poster session

Abstract 7: Joelle Pineau: Discriminative and Generative Models for Building Dialogue Systems in Learning to Generate Natural Language, 02:00 PM

Discriminative and Generative Models for Building Dialogue Systems

Abstract 8: Mark Johnson: Generation in Image Captioning in Learning to Generate Natural Language, 02:45 PM

This talk describes generation in an automatic image captioning system. We start by describing a novel captioning evaluation method that uses syntactic parsing to quantify the similarity of the predicate argument structure of the generated and reference captions, and show that this correlates better with human judgements than existing n-gram overlap evaluation measures. Then we describe a novel decoding algorithm for caption generation that uses a set of beams structured as a finite state automaton to guarantee that the generated caption satisfies certain constraints. We show how this decoder can be used to force the generated captions to include words identified by an external image labelling system, enabling the caption generator to produce vocabulary not seen in its training data.

This is joint work with Peter Anderson, Stephen Gould and Basura Ferndando.

Abstract 9: Coffee Break & Poster session 3 in Learning to Generate Natural Language, 03:30 PM

Coffee Break & Poster session 3

Abstract 10: Trevor Cohn: Sequence-to-sequence transduction for language in Learning to Generate Natural Language, 04:00 PM

Neural sequence-to-sequence models are revolutionising machine translation, as well as dozens of other transduction applications in language processing. This talk will cover two illustrative applications: the generation of derivational morphological forms of words, and neural machine translation, in both text- and speech- to text translation. Finally, I will turn to decoding, one of the big open problems for sequence-to-sequence models, for which I will describe a novel inference method based on continuous relaxation. This improves output quality over existing methods, and enables inference in richer model architectures with more global interactions.

Abstract 11: Panel Discussion in Learning to Generate Natural Language, 04:45 PM

This is joint work with Tim Baldwin, Phil Blunsom, Trevor Cohn, Mark Johnson, Wang Ling, Andre Martins
Deep networks have had profound impact across machine learning research and in many application areas. DNNs are complex to design and train. They are non-linear systems that almost always have many local optima and are often sensitive to training parameter settings and initial state. Systematic optimization of structure and hyperparameters is possible e.g. with Bayesian optimization, but hampered by the expense of training each design on realistic datasets. Exploration is still ongoing for best design principles. We argue that visualization can play an essential role in understanding DNNs and in developing new design principles. With rich tools for visual exploration of networks during training and inference, one should be able to form closer ties between theory and practice: validating expected behaviors, and exposing the unexpected which can lead to new insights. With the rise of generative modeling and reinforcement learning, more interesting directions like understanding and visualization of generative models, visual explanation for driving policy could be explored as well.

As the second edition of this workshop, we are proposing changes based on the lessons we learned last year. We would like to organize a few domain specific tutorials, and panel discussions. We do think machine learning researchers need a lot of tutorials and advice from the visualization/HCI community and vice versa. Many audience in our workshop last year also suggested that more discussion can greatly help us better define such interdisciplinary area.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30 AM</td>
<td>Opening remark</td>
</tr>
<tr>
<td>08:40 AM</td>
<td>Becoming friends with every pixel, Phillip Isola (UC Berkeley)</td>
</tr>
<tr>
<td>09:20 AM</td>
<td>SmoothGrad: removing noise by adding noise</td>
</tr>
<tr>
<td></td>
<td>Daniel Smilkov, Nikhil Thorat, Been Kim, Fernanda Viegas, Martin M Wattenberg</td>
</tr>
<tr>
<td>09:40 AM</td>
<td>Towards Visual Explanations for Convolutional Neural Networks via Input Resampling, Benjamin J Lengerich, Sandeep Konam, Eric Xing, Stephanie Rosenthal, Manuela Veloso</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Coffee Breaks and Poster session 1</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Coffee Breaks and Poster session</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Understanding Generative Models in Google Brain Magenta, Cinjon Resnick (Google)</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Deep saliency: What is learnt by a deep network about saliency?</td>
</tr>
<tr>
<td></td>
<td>Sen He, Nicolas Pugeault</td>
</tr>
<tr>
<td>11:15 AM</td>
<td>Self-supervised attention for Deep Learning explanations,</td>
</tr>
<tr>
<td></td>
<td>Nathan Hodas, (PNL)</td>
</tr>
<tr>
<td>11:45 AM</td>
<td>Skip-Frame Embeddings for Feature Adaptation and Visualization, Zain Shah</td>
</tr>
<tr>
<td>12:00 PM</td>
<td>Lunch Breaks</td>
</tr>
<tr>
<td>02:00 PM</td>
<td>Quantifying the Interpretability of Deep Visual Representations,</td>
</tr>
<tr>
<td></td>
<td>Bolei Zhou (MIT)</td>
</tr>
<tr>
<td>02:40 PM</td>
<td>Visualizing Feature Maps in Deep Neural Networks using DeepResolve - A Genomics Case Study, Ge Liu, David Gifford</td>
</tr>
<tr>
<td>03:00 PM</td>
<td>Coffee Breaks and Poster session 2</td>
</tr>
<tr>
<td>03:30 PM</td>
<td>Visual Explanations from Deep Networks, Dhruv Batra</td>
</tr>
<tr>
<td></td>
<td>(Georgia Tech and Facebook AI Research)</td>
</tr>
<tr>
<td>04:00 PM</td>
<td>Evolutionary Visual Analysis of Deep Neural Networks, Wen Zhong, Cong Xie, Yuan Zhong, Yang Wang, Wei Xu, Shenghui Cheng, Klaus Mueller</td>
</tr>
<tr>
<td>04:20 PM</td>
<td>ActiVis: Visual Exploration of Industry-Scale Deep Neural Network Models, Pierre Andrews (Facebook)</td>
</tr>
<tr>
<td>04:50 PM</td>
<td>Brainstorming on deep learning visualization techniques and tools</td>
</tr>
<tr>
<td>05:30 PM</td>
<td>Closing remark</td>
</tr>
</tbody>
</table>

Abstracts (7):

Abstract 3: SmoothGrad: removing noise by adding noise Daniel Smilkov, Nikhil Thorat, Been Kim, Fernanda Viegas, Martin M Wattenberg in Workshop on Visualization for Deep Learning, 09:20 AM
Explaining the output of a deep network remains a challenge. In the case of an image classifier, one type of explanation is to identify pixels that strongly influence the final decision. A starting point for this strategy is the gradient of the class score function with respect to the input image. This gradient can be interpreted as a sensitivity map, and there are several techniques that elaborate on this basic idea. This paper makes two contributions: it introduces SMOOTHGRAD, a simple method that can help visually sharpen gradient-based sensitivity maps, and it discusses lessons in the visualization of these maps. We publish the code for our experiments and a website with our results.

Abstract 4: Towards Visual Explanations for Convolutional Neural Networks via Input Resampling, Benjamin J Lengerich, Sandeep Konam, Eric Xing, Stephanie Rosenthal, Manuela Veloso in Workshop on Visualization for Deep Learning, 09:40 AM

The predictive power of neural networks often costs model interpretability. Several techniques have been developed for explaining model outputs in terms of input features; however, it is difficult to translate such interpretations into actionable insight. Here, we propose a framework to analyze predictions in terms of the model's internal features by inspecting information flow through the network. Given a trained network and a test image, we select neurons by two metrics, both measured over a set of images created by perturbations to the input image: (1) magnitude of the correlation between the neuron activation and the network output and (2) precision of the neuron activation. We show that the former metric selects neurons that exert large influence over the network output while the latter metric selects neurons that activate on generalizable features. By comparing the sets of neurons selected by these two metrics, our framework offers a way to investigate the internal attention mechanisms of convolutional neural networks.

Abstract 8: Deep saliency: What is learnt by a deep network about saliency? Sen He, Nicolas Pugeault in Workshop on Visualization for Deep Learning, 11:00 AM

Deep convolutional neural networks have achieved impressive performance on a broad range of problems, beating prior art on established benchmarks, but it often remains unclear what are the representations learnt by those systems and how they achieve such performance. This article examines the specific problem of saliency detection, where benchmarks are currently dominated by CNN-based approaches, and investigates the properties of the learnt representation by visualizing the artificial neurons' receptive fields. We demonstrate that fine tuning a pre-trained network on the saliency detection task lead to a profound transformation of the network’s deeper layers. Moreover we argue that this transformation leads to the emergence of receptive fields conceptually similar to the centre-surround filters hypothesized by early research on visual saliency.

Abstract 10: Skip-Frame Embeddings for Feature Adaptation and Visualization, Zain Shah in Workshop on Visualization for Deep Learning, 11:45 AM

We present an unsupervised method for visualizing the generalization and adaptation capabilities of pre-trained features on video. Like the skip-grams method for unsupervised learning of word vector representations, we exploit temporal continuity in the target media, namely that neighboring frames are qualitatively similar. By enforcing this continuity in the adapted feature space we can adapt features to a new target task, like house price prediction, without supervision. The domain-specific embeddings can be easily visualized for qualitative introspection and evaluation.

Abstract 13: Visualizing Feature Maps in Deep Neural Networks using DeepResolve - A Genomics Case Study, Ge Liu, David Gifford in Workshop on Visualization for Deep Learning, 02:40 PM

Although many powerful visualization tools have been developed to interpret neural network decisions in input space, methods to interpret feature map space remain limited. Most existing tools examine a network’s response to a specific input sample and thus are locally faithful to that sample. We introduce DeepResolve, a gradient ascent based method that visualizes intermediate layer feature maps in an input independent manner. We examine DeepResolve’s capability to 1) discover network linear and non-linear combinatorial logic and summarize overall knowledge of a class, 2) reveal key features for a target class, 3) assess a network’s activeness in pattern learning and network’s vulnerability in feature space, and 4) analyze multi-task class similarity at high resolution. We demonstrate the value of DeepResolve on synthetic and experimental genomic datasets, and DeepResolve reveals biologically interesting observations from the experimental data.

Abstract 16: Evolutionary Visual Analysis of Deep Neural Networks, Wen Zhong, Cong Xie, Yuan Zhong, Yang Wang, Wei Xu, Shenghui Cheng, Klaus Mueller in Workshop on Visualization for Deep Learning, 04:00 PM

Recently, deep learning visualization gained a lot of attentions for understanding deep neural networks. However, there is a missing focus on the visualization of deep model training process. To bridge the gap, in this paper, we firstly define a
discriminability metric to evaluate neuron evolution
and a density metric to investigate output
feature maps. Based on these metrics, a level-of-detail
visual analytics framework is proposed to
locally and globally inspect the evolution of deep
neural networks. Finally, we demonstrate the effectiveness
of our system with two real world
case studies.

Abstract 17: ActiVis: Visual Exploration of Industry-Scale Deep
Neural Network Models, Pierre Andrews (Facebook) in Workshop on
Visualization for Deep Learning, 04:20 PM

While deep learning models have achieved state-of-the-art accuracies
for many prediction tasks, understanding these models remains a
challenge. Despite the recent interest in developing visual tools to help
users interpret deep learning models, the complexity and wide variety of
models deployed in industry, and the large-scale datasets that they used,
pose unique design challenges that are inadequately addressed by
existing work. Through participatory design sessions with over 15
researchers and engineers at Facebook, we have developed, deployed,
and iteratively improved ActiVis, an interactive visualization system for
interpreting large-scale deep learning models and results. By tightly
integrating multiple coordinated views, such as a computation graph
overview of the model architecture, and a neuron activation view for
pattern discovery and comparison, users can explore complex deep
neural network models at both the instance- and subset-level. ActiVis
has been deployed on Facebook's machine learning platform. We
present case studies with Facebook researchers and engineers, and
usage scenarios of how ActiVis may work with different models.

Workshop on Computational Biology

Dana Pe’er, Christina Leslie, Elham Azizi, Sandhya Prabhakaran,
Meghana Kshirsagar, Ambrose Carr

C4.4, Thu Aug 10, 08:30 AM

The workshop will showcase recent research in the field of
Computational Biology. There has been significant development in
genomic sequencing techniques as well as imaging technologies that not
only generate huge amounts of data but provide unprecedented levels of
resolution, that of a single cell and even subcellular resolution. This
availability of high dimensional data, at multiple spatial and temporal
resolutions and capturing several perspectives of biological phenomena
has made machine learning methods increasingly relevant for
computational analysis of the data. Conversely, biological data has also
exposed unique challenges and problems that call for the development
of new machine learning methods. This workshop aims at bringing in
researchers working at the intersection of Machine Learning and Biology
to present recent advances and open questions in computational biology
to the ICML community.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:45 AM</td>
<td>Opening Remarks</td>
</tr>
</tbody>
</table>
| 08:50 AM| Stability and Aggregation of
          Experimental Results              |
| 09:30 AM| Spotlight Presentations              |

Abstracts (1):

Abstract 2: Stability and Aggregation of Experimental Results in
Workshop on Computational Biology, 08:50 AM

Spearman’s correlation measures the association between ranked lists.
Given a set of ranked lists, we study two tasks: aggregating the set of
ranks into one single ranked list, and computing the agreement of the
lists as we traverse it. Applications include the analysis of the stability of
feature selection and integration of various sources of information. This
is illustrated with two examples respectively: We study the stability of
identifying variations in GWAS by considering replication studies. In
another study, we aggregate genomic distance, 3D associations, and
literature information to find promising disease associated variations. It
turns out that these problems can be tackled by considering a
multivariate Spearman’s correlation.

Principled Approaches to Deep Learning

Andrzej Pronobis, Robert Gens, Sham M. Kakade, Pedro Domingos

C4.5, Thu Aug 10, 08:30 AM

The recent advancements in deep learning have revolutionized the field
of machine learning, enabling unparalleled performance and many new
real-world applications. Yet, the developments that led to this success
have often been driven by empirical studies, and little is known about the
theory behind some of the most successful approaches. While
theoretically well-founded deep learning architectures had been proposed in the past, they came at a price of increased complexity and reduced tractability. Recently, we have witnessed considerable interest in principled deep learning. This led to a better theoretical understanding of existing architectures as well as development of more mature deep models with solid theoretical foundations. In this workshop, we intend to review the state of those developments and provide a platform for the exchange of ideas between the theoreticians and the practitioners of the growing deep learning community. Through a series of invited talks by the experts in the field, contributed presentations, and an interactive panel discussion, the workshop will cover recent theoretical developments, provide an overview of promising and mature architectures, highlight their challenges and unique benefits, and present the most exciting recent results.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30 AM</td>
<td>Welcome and Opening Remarks</td>
</tr>
<tr>
<td>08:45 AM</td>
<td>Invited Talk 1 - Sanjeev Arora</td>
</tr>
<tr>
<td>09:15 AM</td>
<td>Contributed Presentation 1 - Towards a Deeper Understanding of Training Quantized Neural Networks</td>
</tr>
<tr>
<td>09:30 AM</td>
<td>Invited Talk 2 - Surya Ganguli</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Coffee Break and Poster Session</td>
</tr>
<tr>
<td>10:45 AM</td>
<td>Invited Talk 3 - Ruslan Salakhutdinov</td>
</tr>
<tr>
<td>11:15 AM</td>
<td>Invited Talk 4 - Pedro Domingos</td>
</tr>
<tr>
<td>11:45 AM</td>
<td>Contributed Presentation 2 - LibSPN: A Library for Learning and Inference with Sum-Product Networks and TensorFlow</td>
</tr>
<tr>
<td>12:00 PM</td>
<td>Lunch</td>
</tr>
<tr>
<td>01:30 PM</td>
<td>Invited Talk 5 - Tomaso Poggio</td>
</tr>
<tr>
<td>02:00 PM</td>
<td>Contributed Presentation 3 - Emergence of invariance and disentangling in deep representations</td>
</tr>
<tr>
<td>02:15 PM</td>
<td>Invited Talk 6 - Nathan Srebro</td>
</tr>
<tr>
<td>02:45 PM</td>
<td>Contributed Presentation 4 - The Shattered Gradients Problem: If resnets are the answer, then what is the question?</td>
</tr>
<tr>
<td>03:00 PM</td>
<td>Coffee Break 2 and Poster Session</td>
</tr>
<tr>
<td>03:45 PM</td>
<td>Contributed Presentation 5 - Towards Deep Learning Models Resistant to Adversarial Attacks</td>
</tr>
<tr>
<td>04:00 PM</td>
<td>Panel Discussion</td>
</tr>
<tr>
<td>05:20 PM</td>
<td>Closing Remarks and Awards</td>
</tr>
</tbody>
</table>

Abstracts (10):

Abstract 2: Invited Talk 1 - Sanjeev Arora in Principled Approaches to Deep Learning, 08:45 AM

Do GANs Actually Learn the Distribution? Some Theory and Empirics

The Generative Adversarial Nets or GANs framework (Goodfellow et al’14) for learning distributions differs from older ideas such as autoencoders and deep Boltzmann machines in that it scores the generated distribution using a discriminator net, instead of a perplexity-like calculation. It appears to work well in practice, e.g., the generated images look better than older techniques. But how well do these nets learn the target distribution?

Our paper 1 (ICML’17) shows GAN training may not have good generalization properties; e.g., training may appear successful but the trained distribution may be far from target distribution in standard metrics. We show theoretically that this can happen even though the 2-person game between discriminator and generator is in near-equilibrium, where the generator appears to have “won” (with respect to natural training objectives).

Paper2 (arxiv June 26) empirically tests the whether this lack of generalization occurs in real-life training. The paper introduces a new quantitative test for diversity of a distribution based upon the famous birthday paradox. This test reveals that distributions learnt by some leading GANs techniques have fairly small support (i.e., suffer from mode collapse), which implies that they are far from the target distribution.

Paper1: “Equilibrium and Generalization in GANs” by Arora, Ge, Liang, Ma, Zhang. (ICML 2017)


Abstract 3: Contributed Presentation 1 - Towards a Deeper Understanding of Training Quantized Neural Networks in Principled Approaches to Deep Learning, 09:15 AM

Towards a Deeper Understanding of Training Quantized Neural Networks

Hao Li, Soham De, Zheng Xu, Christoph Studer, Hanan Samet, Tom Goldstein

Training neural networks with coarsely quantized weights is a key step
towards learning on embedded platforms that have limited computing resources, memory capacity, and power consumption. Numerous recent publications have studied methods for training quantized networks, but these studies have been purely experimental. In this work, we investigate the theory of training quantized neural networks. We analyze the convergence properties of commonly used quantized training methods. We also show that training algorithms that exploit high-precision representations have an important annealing property that purely quantized training methods lack, which explains many of the observed empirical differences between these types of algorithms.

Abstract 4: Invited Talk 2 - Surya Ganguli in Principled Approaches to Deep Learning, 09:30 AM

On the Beneficial Role of Dynamic Criticality and Chaos in Deep Learning

What does a generic deep function “look like” and how can we understand and exploit such knowledge to obtain practical benefits in deep learning? By combining Riemannian geometry with dynamic mean field theory, we show that generic nonlinear deep networks exhibit an order to chaos phase transition as synaptic weights vary from small to large. In the chaotic phase, deep networks acquire very high expressive power: measures of functional curvature and the ability to disentangle classification boundaries both grow exponentially with depth, but not with width. Moreover, we apply tools from free probability theory to study the propagation of error gradients through generic deep networks. We find, at the phase transition boundary between order and chaos, that not only the norms of gradients, but also angles between pairs of gradients are preserved even in infinitely deep sigmoidal networks with orthogonal weights. In contrast, ReLu networks do not enjoy such isometric propagation of gradients. In turn, this isometric propagation at the edge of chaos leads to training benefits, where very deep sigmoidal networks outperform ReLu networks, thereby pointing to a potential path to resurrecting saturating nonlinearities in deep learning.

Abstract 6: Invited Talk 3 - Ruslan Salakhutdinov in Principled Approaches to Deep Learning, 10:45 AM

Neural Map: Structured Memory for Deep Reinforcement Learning

A critical component to enabling intelligent reasoning in partially observable environments is memory. Despite this importance, Deep Reinforcement Learning (DRL) agents have so far used relatively simple memory architectures, with the main methods to overcome partial observability being either a temporal convolution over the past k frames or an LSTM layer. In this talk, we will introduce a memory system with an adaptable write operator that is customized to the sorts of 3D environments that DRL agents typically interact with. This architecture, called the Neural Map, uses a spatially structured 2D memory image to learn to store arbitrary information about the environment over long time lags. We demonstrate empirically that the Neural Map surpasses previous DRL memories on a set of challenging 2D and 3D maze environments and show that it is capable of generalizing to environments that were not seen during training.

Joint work with Emilio Parisotto

Abstract 7: Invited Talk 4 - Pedro Domingos in Principled Approaches to Deep Learning, 11:15 AM

The Sum-Product Theorem: A Foundation for Learning Tractable Deep Models

Inference in expressive probabilistic models is generally intractable, which makes them difficult to learn and limits their applicability. Sum-product networks are a class of deep models where, surprisingly, inference remains tractable even when an arbitrary number of hidden layers are present. In this talk, I generalize this result to a much broader set of learning problems: all those where inference consists of summing a function over a semiring. This includes satisfiability, constraint satisfaction, optimization, integration, and others. In any semiring, for summation to be tractable it suffices that the factors of every product have disjoint scopes. This unifies and extends many previous results in the literature. Enforcing this condition at learning time thus ensures that the learned models are tractable. I illustrate the power and generality of this approach by applying it to a new type of structured prediction problem: learning a nonconvex function that can be globally optimized in polynomial time. I show empirically that this greatly outperforms the standard approach of learning without regard to the cost of optimization.

(Join work with Abram Friesen)

Abstract 8: Contributed Presentation 2 - LibSPN: A Library for Learning and Inference with Sum-Product Networks and TensorFlow in Principled Approaches to Deep Learning, 11:45 AM

LibSPN: A Library for Learning and Inference with Sum-Product Networks and TensorFlow

Andrzej Pronobis, Avinash Ranganath, Rajesh Rao

Sum-Product Networks (SPNs) are a probabilistic deep architecture with solid theoretical foundations, which demonstrated state-of-the-art performance in several domains. Yet, surprisingly, there are no mature, general-purpose SPN implementations that would serve as a platform for the community of machine learning researchers centered around SPNs. Here, we present a new general-purpose Python library called LibSPN, which aims to become such a platform. The library is designed to make it straightforward and effortless to apply various SPN architectures to large-scale datasets and problems. The library achieves scalability and efficiency, thanks to a tight coupling with TensorFlow, a framework already used by a large community of researchers and developers in multiple domains. We describe the design and benefits of LibSPN, give several use-case examples, and demonstrate the applicability of the library to real-world problems on the example of spatial understanding in mobile robotics.

Abstract 11: Contributed Presentation 3 - Emergence of invariance and disentangling in deep representations in Principled Approaches to Deep Learning, 02:00 PM

Emergence of invariance and disentangling in deep representations

Alessandro Achille, Stefano Soatto

We show that invariance in a deep neural network is equivalent to the information minimality of the representation it computes, and that stacking layers and injecting noise during training naturally bias the network towards learning invariant representations. Then, we show that overfitting is related to the quantity of information stored in the weights, and derive a sharp bound between this information and the minimality and Total Correlation of the layers. This allows us to conclude that
Implicit and explicit regularization of the loss function not only help limit overfitting, but also foster invariance and disentangling of the learned representation. We also shed light on the properties of deep networks in relation to the geometry of the loss function.

Abstract 12: Invited Talk 6 - Nathan Srebro in Principled Approaches to Deep Learning, 02:15 PM

Geometry, Optimization and Generalization in Multilayer Networks

What is it that enables learning with multi-layer networks? What causes the network to generalize well despite the model class having extremely high capacity? In this talk, I will explore these questions through experimentation, analogy to matrix factorization (including some new results on the energy landscape and implicit regularization in matrix factorization), and study of alternate geometries and optimization approaches.

Abstract 13: Contributed Presentation 4 - The Shattered Gradients Problem: If resnets are the answer, then what is the question? in Principled Approaches to Deep Learning, 02:45 PM

The Shattered Gradients Problem: If resnets are the answer, then what is the question?

David Balduzzi, Brian McWilliams, Marcus Frean, John Lewis, Lennox Leary, Kurt Wan Duo Ma

A long-standing obstacle to progress in deep learning is the problem of vanishing and exploding gradients. Although, the problem has largely been overcome via carefully constructed initializations and batch normalization, architectures incorporating skip-connections such as highway and resnets perform much better than standard feedforward architectures despite well-chosen initialization and batch normalization. In this paper, we identify the shattered gradients problem. Specifically, we show that the correlation between gradients in standard feedforward networks decays exponentially with depth resulting in gradients that resemble white noise whereas, in contrast, the gradients in architectures with skip-connections are far more resistant to shattering, decaying sublinearly. Detailed empirical evidence is presented in support of the analysis, on both fully-connected networks and convnets. Finally, we present a new "looks linear" (LL) initialization that prevents shattering, with preliminary experiments showing the new initialization allows to train very deep networks without the addition of skip-connections.

Abstract 15: Contributed Presentation 5 - Towards Deep Learning Models Resistant to Adversarial Attacks in Principled Approaches to Deep Learning, 03:45 PM

Towards Deep Learning Models Resistant to Adversarial Attacks

Aleksander Mikdry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, Adrian Vladu

Recent work has demonstrated that neural networks are vulnerable to adversarial examples, i.e., inputs that are almost indistinguishable from natural data and yet classified incorrectly by the network. To address this problem, we study the adversarial robustness of neural networks through the lens of robust optimization. This approach provides a broad and unifying view on much of the prior work on this topic. Its principled nature also enables us to identify general methods for both training and attacking neural networks that are reliable and, in a certain sense, universal. These methods let us train networks with significantly improved resistance to a wide range of adversarial attacks. This suggests that adversarially resistant deep learning models might be within our reach after all.

Video Games and Machine Learning

Gabriel Synnaeve, Julian Togelius, Tom Schaul, Oriol Vinyals, Nicolas Usunier

C4.6, Thu Aug 10, 08:30 AM

Good benchmarks are necessary for developing artificial intelligence. Recently, there has been a growing movement for the use of video games as machine learning benchmarks [1,2,3], and also an interest in the applications of machine learning from the video games community. While games have been used for AI research for a long time, only recently have we seen modern machine learning methods applied to video games.

This workshop focuses on complex games which provide interesting and hard challenges for machine learning. Going beyond simple toy problems of the past, and games which can easily be solved with search, we focus on games where learning is likely to be necessary to play well. This includes strategy games such as StarCraft [4,5], open-world games such as Minecraft [6,7,8], first-person shooters such as Doom [9,10], as well as hard and unsolved 2D games such as Ms. Pac-Man and Montezuma’s Revenge [11,12,13]. While we see most of the challenges in game-playing, there are also interesting machine learning challenges in modeling and content generation [14]. This workshop aims at bringing together all researchers from ICML who want to use video games as a benchmark. We will have talks by invited speakers from machine learning, from the game AI community, and from the video games industry.

[5] StarCraft AI Competition @ AIIDE 2016
throughput.

example, modern passenger screening systems impose constraints on search engine is not within a few tens of milliseconds. In another users are known to abandon the service is the response time of the Additionally, search engines have time constraints at prediction-time as budgeted to enable business models such as online advertising.

For instance, in search engines CPU cost during prediction-time must be computational cost, time, network-throughput and power-consumption.

In these applications, budget constraints arise as a result of limits on prediction under budget constraints is a critical problem that arise in several settings like medical diagnosis, search engines and surveillance.

Motivation

Prediction under budget constraints is a critical problem that arise in several settings like medical diagnosis, search engines and surveillance. In these applications, budget constraints arise as a result of limits on computational cost, time, network-throughput and power-consumption. For instance, in search engines CPU cost during prediction-time must be budgeted to enable business models such as online advertising. Additionally, search engines have time constraints at prediction-time as users are known to abandon the service is the response time of the search engine is not within a few tens of milliseconds. In another example, modern passenger screening systems impose constraints on throughput.

An extreme version of these problems appear in the Internet of Things (IoT) setting where one requires prediction on tiny IoT devices which might have at most 2KB of RAM and no floating point computation unit. IoT is considered to be the next multi-billion industry with “smart” devices being designed for production-line, cars, retail stores, and even for toothbrush and spoons. Given that IoT based solutions seem destined to significantly permeate our day-to-day lives, ML based predictions on the device become critical due to several reasons like privacy, battery, latency etc.

Learning under resource constraints departs from the traditional machine learning setting and introduces new exciting challenges. For instance, features are accompanied by costs (e.g. extraction time in search engines or true monetary values in medical diagnosis) and their amortized sum is constrained at test-time. Also, different search strategies in prediction can have widely varying computational costs (e.g., binary search, linear search, dynamic programming). In other settings, a system must maintain a throughput constraint to keep pace with arriving traffic.

In IoT setting, the model itself has to be deployed on a 2-16KB RAM, posing an extremely challenging constraint on the algorithm.

The common aspect of all of these settings is that we must seeks trade-offs between prediction accuracy and prediction cost. Studying this tradeoff is an inherent challenge that needs to be investigated in a principled fashion in order to invent practically relevant machine learning algorithms. This problems lies at the intersection of ML, statistics, stochastic control and information theory. We aim to draw researchers working on foundational, algorithmic and application problems within these areas. We plan on organizing a demo session which would showcase ML algorithms running live on various resource-constrained device, demonstrating their effectiveness on challenging real-world tasks. In addition, we plan to invite Ofer Dekel from Microsoft Research to present a new platform for deploying ML on tiny devices which should provide a easy way to deploy and compare various ML techniques on realistic devices and further spur multiple research directions in this area.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:45 AM</td>
<td>Introduction</td>
</tr>
<tr>
<td>08:50 AM</td>
<td>Small Deep-Neural-Networks: Their Advantages, and Their Design by Forrest Llandola (DeepScale)</td>
</tr>
<tr>
<td>09:25 AM</td>
<td>An Adaptive Approximation for Prediction Under a Budget by Venkatesh Saligrama (Boston University)</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>On-Device Machine Intelligence with Neural Projections by Sujith Ravi (Google)</td>
</tr>
<tr>
<td>11:05 AM</td>
<td>Core ML: High-performance on-device machine learning by Bill March (Apple)</td>
</tr>
</tbody>
</table>
SqueezeNet was created using a few basic techniques including kernel model parameters. The popular DNN AlexNet but with a 50x reduction in the number of operations for the object classification problem that achieves the same accuracy as devices. The first result of our efforts was SqueezeNet, a DNN targeted to identify smaller DNN models that can be deployed on embedded devices. We benefitted from the advantages of small DNNs, we set out in 2015 to create DNNs that met the requirements of embedded systems and autonomous car (e.g. self-driving vehicles). This required re-thinking existing machine learning algorithms and coming up either required or strongly recommended. These reasons include:  
- Small models require less bandwidth communication when sending updated models from the cloud to the client (e.g. smartphone or autonomous car) 
- Small models train faster 
- Small models require fewer memory transfers during inference, and 
- Off-chip memory transfers require 100x more power than arithmetic operations.

To create DNNs that met the requirements of embedded systems and benefitted from the advantages of small DNNs, we set out in 2015 to identify smaller DNN models that can be deployed on embedded devices. The first result of our efforts was SqueezeNet, a DNN targeted for the object classification problem that achieves the same accuracy as the popular DNN AlexNet but with a 50x reduction in the number of model parameters. SqueezeNet was created using a few basic techniques including kernel reduction, channel reduction, and delayed pooling. Over the last year, many other researchers have pursued the same goals of small, fast, energy-efficient DNNs for computer-vision problems ranging from object classification to style-transfer. In this talk we review these developments and report our progress in developing a systematic approach to the design of small DNNs.

Abstract 3: An Adaptive Approximation for Prediction Under a Budget by Venkatesh Saligrama (Boston University) in ML on a budget: IoT, Mobile and other tiny-ML applications, 09:25 AM

We propose a novel adaptive approximation approach for test-time resource-constrained prediction for classification and sequential-decision making problems. Given an input instance at test-time, a gating function identifies a prediction model or policy for the input among a collection of models or policies. Our objective is to minimize overall average cost without sacrificing accuracy. We present a novel bottom-up method based on adaptively approximating a high-accuracy model in regions where low-cost models are capable of making highly accurate predictions. We pose an empirical loss minimization problem with cost constraints to jointly train gating and prediction models. On a number of benchmark datasets our method outperforms state-of-the-art achieving higher accuracy for the same cost.

Abstract 4: On-Device Machine Intelligence with Neural Projections by Sujith Ravi (Google) in ML on a budget: IoT, Mobile and other tiny-ML applications, 10:30 AM

Deep neural networks and other machine learning models have been transformative for building intelligent systems capable of visual recognition, speech and language understanding. While recent advances have led to progress for machine intelligence applications running on the cloud, it is often infeasible to use typical machine learning models on devices like mobile phones or smart watches due to computation and memory constraints — model sizes are huge and cannot fit into the limited memory available on such devices. While these devices could make use of models running on high-performance data centers with CPUs or GPUs, this is not feasible for many applications and scenarios where inference needs to be performed directly “on” device. This requires re-thinking existing machine learning algorithms and coming up with new models that are directly optimized for on-device machine intelligence rather than doing post-hoc model compression.

In this talk, I will introduce a novel “projection-based” machine learning system for training compact neural networks. The approach uses a joint optimization framework to simultaneously train a “full” deep network like feed-forward or recursive neural network and a lightweight “projection” network. Unlike the full deep network, the projection network uses random projection operations that are efficient to compute and operates in bit space yielding a low memory footprint. The system is trained end-to-end using backpropagation. We show that the approach is flexible and easily extensible to other machine learning paradigms, for example, we learn graph-based projection models using label propagation. The trained “projection” models are directly used for inference and achieve significant model size reductions and efficiency on several visual and language tasks while providing competitive performance. We have used the novel networks to power machine intelligence applications on devices such as mobile phones and smart watches, for example a fully on-device Smart Reply model that runs on Android smart watches.
Abstract 5: Core ML: High-performance on-device machine learning by Bill March (Apple) in ML on a budget: IoT, Mobile and other tiny-ML applications, 11:05 AM

Considering the limited computing power available on mobile devices, application developers have typically been constrained to either small, simple models or expensive network access to remote servers. This year, Apple introduced Core ML, a new framework for on-device inference.

Core ML combines an open format for encoding a wide-range of models with simple programming interfaces and highly-optimized, on-device evaluation methods. The combination of these factors makes Core ML a powerful tool to bridge the gap between cutting edge ML research and large scale impact on mobile device users. While on-device inference is typically regarded to be limited by power and computing constraints, we show that optimized methods can achieve excellent performance. We will show this first with a demo of Core ML in action, showing that efficient evaluation of state-of-the-art deep neural networks on a mobile device is possible with an extremely simple programming interface. We then discuss some of the optimizations underlying this performance in detail, including graph optimizations and automatic hardware selection algorithms. We then discuss Core ML’s open-source tools and model format, and highlight several ways in which we hope to work together with the wider machine learning community.

Abstract 6: Building Amazon Alexa’s embedded wake word detector by Shiv Naga Prasad (Amazon) in ML on a budget: IoT, Mobile and other tiny-ML applications, 11:40 AM

Alexa is a conversational AI agent that is accessible through several consumer devices such as Echo, Dot, Tap, Echo Show, etc. A key feature is that users can talk to Alexa eyes free and hands free, by invoking a wake up phrase, “Alexa”. Detecting the wake word on the device is one of the grand challenges in far field speech, given the CPU and memory constraints, background noise in household environment, and variation in user’s speech characteristics. We will provide an overview of the technical challenges in this area, and some of the research being conducted at Alexa in efficient ML on edge platforms.

Abstract 8: The Edge of Machine Learning: Resource-efficient ML in 2 KB RAM for the Internet of Things by Manik Varma (Microsoft) in ML on a budget: IoT, Mobile and other tiny-ML applications, 02:00 PM

We propose an alternative paradigm for the Internet of Things (IoT) where machine learning algorithms run locally on severely resource-constrained edge and endpoint devices without necessarily needing cloud connectivity. This enables many scenarios beyond the pale of the traditional paradigm including low-latency brain implants, precision agriculture on disconnected farms, privacy-preserving smart spectacles, etc.

Towards this end, we develop novel tree and kNN based algorithm, called Bonsai and ProtoNN, for efficient prediction on IoT devices -- such as those based on the Arduino Uno board having an 8 bit ATmega328P microcontroller operating at 16 MHz with no native floating point support, 2 KB RAM and 32 KB read-only flash memory. Bonsai and ProtoNN maintain prediction accuracy while minimizing model size and prediction costs by: (a) developing novel compressed yet expressive models; (b) sparsely projecting all data into a low-dimensional space in which the models are learnt; and (c) jointly learning all model and projection parameters. Experimental results on multiple benchmark datasets demonstrate that Bonsai and ProtoNN can make predictions in milliseconds even on slow microcontrollers, can fit in KB of memory, have lower battery consumption than all other algorithms while achieving prediction accuracies that can be as much as 30% higher than state-of-the-art methods for resource-efficient machine learning. Bonsai and ProtoNN are also shown to generalize to other resource constrained settings beyond IoT by generating significantly better search results as compared to Bing’s L3 ranker when the model size is restricted to 300 bytes.

Abstract 11: Trading-Off Cost of Deployment Versus Accuracy in Learning Predictive Models by Suchi Saria (JHU) in ML on a budget: IoT, Mobile and other tiny-ML applications, 03:30 PM

Predictive models are finding an increasing number of applications in many industries. As a result, a practical means for trading-off the cost of deploying a model versus its effectiveness is needed. Our work is motivated by risk prediction problems in healthcare. Cost-structures in domains such as healthcare are quite complex, posing a significant challenge to existing approaches. We propose a novel framework for designing cost-sensitive structured regularizers that is suitable for problems with complex cost dependencies. We draw upon a surprising connection to boolean circuits. In particular, we represent the problem costs as a multi-layer boolean circuit, and then use properties of boolean circuits to define an extended feature vector and a group regularizer that exactly captures the underlying cost structure. The resulting regularizer may then be combined with a fidelity function to perform model prediction, for example. For the challenging real-world application of risk prediction for sepsis in intensive care units, the use of our regularizer leads to models that are in harmony with the underlying cost structure and thus provide an excellent prediction accuracy versus cost tradeoff.

Abstract 12: Resource Efficient Driving Policy by Shaked Sammah (Mobileye) in ML on a budget: IoT, Mobile and other tiny-ML applications, 04:05 PM

When attacking the problem of Autonomous Driving, one must take into account strict computational constraints, posed by the desired low cost of sensors and processors, and by the required real-time performance. Specifically, when considering Driving Policy, many of the current state-of-the-art solutions for planning in large state spaces (applied to different problems), are ruled out. We discuss approaches which allow feasible planning, through different representations of the state space, along with the use of both supervised and reinforcement learning algorithms.

Workshop on Human Interpretability in Machine Learning (WHI)

Kush Varshney, Adrian Weller, Been Kim, Dmitry Malioutov

C4.8, Thu Aug 10, 08:30 AM

This workshop will bring together researchers who study the interpretability of predictive models, develop interpretable machine learning algorithms, and develop methodology to interpret black-box machine learning models (e.g., post-hoc interpretations). This is a very exciting time to study interpretable machine learning, as the advances in large-scale optimization and Bayesian inference that have enabled the
rise of black-box machine learning are now also starting to be exploited
to develop principled approaches to large-scale interpretable machine
learning. Participants in the workshop will exchange ideas on these and
allied topics, including:

- Quantifying and axiomatizing interpretability
- Psychology of human concept learning
- Rule learning, Symbolic regression and case-based reasoning
- Generalized additive models, sparsity and interpretability
- Visual analytics
- Interpretable unsupervised models (clustering, topic models, e.t.c)
- Interpretation of black-box models (including deep neural networks)
- Causality of predictive models
- Verifying, diagnosing and debugging machine learning systems
- Interpretability in reinforcement learning.

Doctors, judges, business executives, and many other people are faced
with making critical decisions that can have profound consequences. For
example, doctors decide which treatment to administer to patients,
judges decide on prison sentences for convicts, and business executives
decide to enter new markets and acquire other companies. Such
decisions are increasingly being supported by predictive models learned
by algorithms from historical data.

The latest trend in machine learning is to use highly nonlinear complex
systems such as deep neural networks, kernel methods, and large
ensembles of diverse classifiers. While such approaches often produce
impressive, state-of-the art prediction accuracies, their black-box nature
offers little comfort to decision makers. Therefore, in order for predictions
to be adopted, trusted, and safely used by decision makers in
mission-critical applications, it is imperative to develop machine learning
methods that produce interpretable models with excellent predictive
accuracy. It is in this way that machine learning methods can have
impact on consequential real-world applications.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Name</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30 AM</td>
<td>A. Dhurandhar, V. Iyengar, R. Luss, and K. Shanmugam</td>
<td>“A Formal Framework to Characterize Interpretability of Procedures”</td>
</tr>
<tr>
<td>08:45 AM</td>
<td>A. Henelius, K. Puolamäki, and A. Ukkonen</td>
<td>“Interpreting Classifiers through Attribute Interactions in Datasets”</td>
</tr>
<tr>
<td>09:00 AM</td>
<td>S. Lundberg and S.-I. Lee</td>
<td>“Consistent Feature Attribution for Tree Ensembles”</td>
</tr>
<tr>
<td>09:15 AM</td>
<td>Invited Talk: D. Sontag</td>
<td></td>
</tr>
<tr>
<td>10:00 AM</td>
<td>S. Penkov and S. Ramamoorthy</td>
<td>“Program Induction to Interpret Transition Systems”</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>C. Rosenbaum, T. Gao, and T. Klinger</td>
<td>“e-QRAQ: A Multi-turn Reasoning Dataset and Simulator with Explanations”</td>
</tr>
<tr>
<td>11:15 AM</td>
<td>Invited Talk: T. Jebara</td>
<td>“Interpretability Through Causality”</td>
</tr>
<tr>
<td>02:00 PM</td>
<td>W. Tansey, J. Thomason, and J. G. Scott</td>
<td>“Interpretable Low-Dimensional Regression via Data-Adaptive Smoothing”</td>
</tr>
<tr>
<td>02:15 PM</td>
<td>Invited Talk: P. W. Koh</td>
<td></td>
</tr>
<tr>
<td>03:30 PM</td>
<td>I. Valera, M. F. Pradier, and Z. Ghahramani</td>
<td>“General Latent Feature Modeling for Data Exploration Tasks”</td>
</tr>
<tr>
<td>03:45 PM</td>
<td>A. Weller</td>
<td>“Challenges for Transparency”</td>
</tr>
<tr>
<td>04:00 PM</td>
<td>ICML WHI 2017 Awards Ceremony</td>
<td></td>
</tr>
<tr>
<td>04:05 PM</td>
<td>Panel Discussion: Human Interpretability in Machine Learning</td>
<td></td>
</tr>
</tbody>
</table>

Abstracts (12):

Abstract 1: A. Dhurandhar, V. Iyengar, R. Luss, and K. Shanmugam,
“A Formal Framework to Characterize Interpretability of Procedures” in Workshop on Human Interpretability in Machine Learning (WHI), Shanmugam 08:30 AM

We provide a novel notion of what it means to be interpretable, looking
past the usual association with human understanding. Our key insight is
that interpretability is not an absolute concept and so we define it relative
to a target model, which may or may not be a human. We define a
framework that allows for comparing interpretable procedures by linking it
to important practical aspects such as accuracy and robustness. We
characterize many of the current state-of-the-art interpretable methods in
our framework portraying its general applicability.

Abstract 2: A. Henelius, K. Puolamäki, and A. Ukkonen, “Interpreting Classifiers through Attribute Interactions in Datasets” in Workshop on Human Interpretability in Machine Learning (WHI), Henelius 08:45 AM

In this work we present the novel ASTRID method for investigating which
attribute interactions classifiers exploit when making predictions.
Attribute interactions in classification tasks mean that two or more
attributes together provide stronger evidence for a particular class label.
Knowledge of such interactions makes models more interpretable by revealing associations between attributes. This has applications, e.g., in pharmacovigilance to identify interactions between drugs or in bioinformatics to investigate associations between single nucleotide polymorphisms. We also show how the found attribute partitioning is related to a factorisation of the data generating distribution and empirically demonstrate the utility of the proposed method.

Abstract 3: S. Lundberg and S.-I. Lee, “Consistent Feature Attribution for Tree Ensembles” in Workshop on Human Interpretability in Machine Learning (WHI), Hiranuma 09:00 AM

It is critical in many applications to understand what features are important for a model, and why individual predictions were made. For tree ensemble methods these questions are usually answered by attributing importance values to input features, either globally or for a single prediction. Here we show that current feature attribution methods are inconsistent, which means changing the model to rely more on a given feature can actually decrease the importance assigned to that feature. To address this problem we develop fast exact solutions for SHAP (SHapley Additive exPlanation) values, which were recently shown to be the unique additive feature attribution method based on conditional expectations that is both consistent and locally accurate. We integrate these improvements into the latest version of XGBoost, demonstrate the inconsistencies of current methods, and show how using SHAP values results in significantly improved supervised clustering performance. Feature importance values are a key part of understanding widely used models such as gradient boosting trees and random forests. We believe our work improves on the state-of-the-art in important ways, and may impact any current user of tree ensemble methods.

Abstract 5: S. Penkov and S. Ramamoorthy, “Program Induction to Interpret Transition Systems” in Workshop on Human Interpretability in Machine Learning (WHI), Penkov 10:30 AM

Explaining and reasoning about processes which underlie observed black-box phenomena enables the discovery of causal mechanisms, derivation of suitable abstract representations and the formulation of more robust predictions. We propose to learn high level functional programs in order to represent abstract models which capture the invariant structure in the observed data. We introduce the \(n\)-machine (program-induction machine) -- an architecture able to induce interpretable LISP-like programs from observed data traces. We propose an optimisation procedure for program learning based on backpropagation, gradient descent and \(A^*\) search. We apply the proposed method to two problems: system identification of dynamical systems, and explaining the behaviour of a DQN agent. Our results show that the \(n\)-machine can efficiently induce interpretable programs from individual data traces.


Active learning has long been a topic of study in machine learning. However, as increasingly complex and opaque models have become standard practice, the process of active learning, too, has become more opaque. There has been little investigation into interpreting what specific trends and patterns an active learning strategy may be exploring. This work expands on the Local Interpretable Model-agnostic Explanations framework (LIME) to provide explanations for active learning recommendations. We demonstrate how LIME can be used to generate locally faithful explanations for an active learning strategy, and how these explanations can be used to understand how different models and datasets explore a problem space over time. In order to quantify the per-subgroup differences in how an active learning strategy queries spatial regions, we introduce a notion of uncertainty bias (based on disparate impact) to measure the discrepancy in the confidence for a model’s predictions between one subgroup and another. Using the uncertainty bias measure, we show that our query explanations accurately reflect the subgroup focus of the active learning queries, allowing for an interpretable explanation of what is being learned as points with similar sources of uncertainty have their uncertainty bias resolved. We demonstrate that this technique can be applied to track uncertainty bias over user-defined clusters or automatically generated clusters based on the source of uncertainty.


In this paper we present a new dataset and user simulator e-QRAQ (explainable Query, Reason, and Answer Question) which tests an Agent’s ability to read an ambiguous text, ask questions until it can answer a challenge question; and explain the reasoning behind its questions and answer. The User simulator provides the Agent with a short, ambiguous story and a challenge question about the story. The story is ambiguous because some of the entities have been replaced by variables. At each turn the Agent may ask for the value of a variable or try to answer the challenge question. In response the User simulator provides a natural language explanation of why the Agent’s query or answer was useful in narrowing down the set of possible answers, or not. To demonstrate one potential application of the e-QRAQ dataset, we train a new neural architecture based on End-to-End Memory Networks to successfully generate both predictions and partial explanations of its current understanding of the problem. We observe a strong correlation between the quality of the prediction and explanation.


While interpretability often involves finding more parsimonious or sparser models to facilitate human understanding, Netflix also seeks to achieve human interpretability by pursuing causal learning. Predictive models can be impressively accurate in a passive setting but might disappoint a human user who expects the recovered relationships to be causal. More importantly, a predictive model’s outcomes may no longer be accurate if the input variables are perturbed through an active intervention. I will briefly discuss applications at Netflix across messaging, marketing and originals promotion which leverage causal modeling in order to achieve models that can be actionable as well as interpretable. In particular, techniques such as two stage least squares (2SLS), instrumental variables (IV), extensions to generalized linear models (GLMs), and other causal methods will be summarized. These causal models can surprisingly recover more interpretable and simpler models than their purely predictive counterparts. Furthermore, sparsity can potentially emerge when causal models ignore spurious relationships that might otherwise be recovered in a purely predictive objective function. In general, causal models achieve better results algorithmically in active intervention settings and enjoy broader adoption from human stakeholders.
Abstract 9: W. Tansey, J. Thomason, and J. G. Scott, "Interpretable Low-Dimensional Regression via Data-Adaptive Smoothing" in Workshop on Human Interpretability in Machine Learning (WHI), Tansey 02:00 PM

We consider the problem of estimating a regression function in the common situation where the number of features is small, where interpretability of the model is a high priority, and where simple linear or additive models fail to provide adequate performance. To address this problem, we present Maximum Variance Total Variation denoising (MVTV), an approach that is conceptually related both to CART and to the more recent CRISP algorithm, a state-of-the-art alternative method for interpretable nonlinear regression. MVTV divides the feature space into blocks of constant value and fits the value of all blocks jointly via a convex optimization routine. Our method is fully data-adaptive, in that it incorporates highly robust routines for tuning all hyperparameters automatically. We compare our approach against CART and CRISP via both a complexity-accuracy tradeoff metric and a human study, demonstrating that that MVTV is a more powerful and interpretable method.

Abstract 11: I. Valera, M. F. Pradier, and Z. Ghahramani, "General Latent Feature Modeling for Data Exploration Tasks" in Workshop on Human Interpretability in Machine Learning (WHI), 03:30 PM

This paper introduces a general Bayesian non-parametric latent feature model suitable to perform automatic exploratory analysis of heterogeneous datasets, where the attributes describing each object can be either discrete, continuous or mixed variables. The proposed model presents several important properties. First, it accounts for heterogeneous data while can be inferred in linear time with respect to the number of objects and attributes. Second, its Bayesian nonparametric nature allows us to automatically infer the model complexity from the data, i.e., the number of features necessary to capture the latent structure in the data. Third, the latent features in the model are binary-valued variables, easing the interpretability of the obtained latent features in data exploration tasks.

Abstract 12: A. Weller, "Challenges for Transparency" in Workshop on Human Interpretability in Machine Learning (WHI), Weller 03:45 PM

Transparency is often deemed critical to enable effective real-world deployment of intelligent systems. Yet the motivations for and benefits of different types of transparency can vary significantly depending on context, and objective measurement criteria are difficult to identify. We provide a brief survey, suggesting challenges and related concerns. We highlight and review settings where transparency may cause harm, discussing connections across privacy, multi-agent game theory, economics, fairness and trust.

Abstract 13: ICML WHI 2017 Awards Ceremony in Workshop on Human Interpretability in Machine Learning (WHI), 04:00 PM

Join us in recognizing the best papers of the workshop.

Abstract 14: Panel Discussion: Human Interpretability in Machine Learning in Workshop on Human Interpretability in Machine Learning (WHI), 04:05 PM

Panelists: Tony Jebara, Bernhard Schölkopf, Been Kim, Kush Varshney
Moderator: Adrian Weller

---

**Automatic Machine Learning (AutoML 2017)**

**Joaquin Vanschoren, Roman Garnett**

**C4.9, Thu Aug 10, 08:30 AM**

Machine learning has achieved considerable successes in recent years and an ever-growing number of disciplines rely on it. However, this success crucially relies on human machine learning experts, who select appropriate features, workflows, machine learning paradigms, algorithms, and their hyperparameters. As the complexity of these tasks is often beyond non-experts, the rapid growth of machine learning applications has created a demand for off-the-shelf machine learning methods that can be used easily and without expert knowledge. We call the resulting research area that targets progressive automation of machine learning AutoML.

**Schedule**

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30 AM</td>
<td>Welcome</td>
</tr>
<tr>
<td>08:40 AM</td>
<td>Generalizing from Few Examples with Meta-Learning, Hugo Larochelle (Google)</td>
</tr>
<tr>
<td>09:40 AM</td>
<td>Neural Block Sampling. Tongzhou Wang, Yi Wu, Dave Moore and Stuart Russell.</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Rob DeLine (Microsoft)</td>
</tr>
<tr>
<td>11:10 AM</td>
<td>Spotlight session 1</td>
</tr>
<tr>
<td>11:30 AM</td>
<td>Poster session 1</td>
</tr>
<tr>
<td>12:00 PM</td>
<td>Lunch</td>
</tr>
<tr>
<td>02:00 PM</td>
<td>Himabindu Lakkaraju (Stanford)</td>
</tr>
<tr>
<td>02:40 PM</td>
<td>YellowFin and the Art of Momentum Tuning. Jian Zhang, Ioannis Mitliagkas and Christopher Re.</td>
</tr>
<tr>
<td>03:00 PM</td>
<td>Coffee</td>
</tr>
<tr>
<td>03:30 PM</td>
<td>Spotlight session 2</td>
</tr>
<tr>
<td>03:50 PM</td>
<td>Poster session 2</td>
</tr>
<tr>
<td>04:30 PM</td>
<td>The future of AutoML</td>
</tr>
</tbody>
</table>

Abstracts (4): 

Abstract 2: Generalizing from Few Examples with Meta-Learning, Hugo Larochelle (Google) in Automatic Machine Learning (AutoML...
A lot of the recent progress on many AI tasks was enable in part by the availability of large quantities of labeled data. Yet, humans are able to learn concepts from as little as a handful of examples. Meta-learning is a very promising framework for addressing the problem of generalizing from small amounts of data, known as few-shot learning. In meta-learning, our model is itself a learning algorithm: it takes as input a training set and outputs a classifier. For few-shot learning, it is (meta-)trained directly to produce classifiers with good generalization performance for problems with very little labeled data. In this talk, I’ll review recent research that has made exciting progress on this topic.

Abstract 3: **Thompson Sampling for Asynchronous Parallel Bayesian Optimisation.** Kirthevasan Kandasamy. in Automatic Machine Learning (AutoML 2017), 09:20 AM

We design and analyse variations of Thompson sampling (TS) for Bayesian optimisation (BO) in settings where function evaluations are expensive, but can be performed in parallel. Our theoretical analysis shows that a direct application of the sequential Thompson sampling algorithm in either synchronous or asynchronous parallel settings yields a surprisingly powerful result: making $\$n\$ evaluations distributed among $\$M\$ workers is essentially equivalent to performing $\$n\$ evaluations in sequence. Further, by modeling the time taken to complete a function evaluation, we show that, under a time constraint, asynchronously parallel TS achieves asymptotically lower regret than both the synchronous and sequential versions. These results are complemented by an experimental analysis, showing that synchronous TS outperforms a suite of existing parallel BO algorithms in simulations and in a hyper-parameter tuning application. In addition, the proposed procedure is conceptually and computationally much simpler than existing work for parallel BO.

Abstract 4: **Neural Block Sampling.** Tongzhou Wang, Yi Wu, Dave Moore and Stuart Russell. in Automatic Machine Learning (AutoML 2017), 09:40 AM

Efficient Monte Carlo inference often requires manual construction of model-specific proposals. We propose an approach to automated proposal construction by training neural networks to provide fast approximations to block Gibbs conditionals. The learned proposal generalizes to occurrences of common structural motifs both within a given model and across models, allowing for the construction of a library of learned inference primitives that can accelerate inference on unseen models with no model-specific training required.

Abstract 10: **YellowFin and the Art of Momentum Tuning.** Jian Zhang, Ioannis Mitliagkas and Christopher Re. in Automatic Machine Learning (AutoML 2017), 02:40 PM

Hyperparameter tuning is one of the big costs of deep learning. State-of-the-art optimizers, such as Adagrad, RMSProp and Adam, make things easier by adaptively tuning an individual learning rate for each variable. This level of fine adaptation is understood to yield a more powerful method. However, our experiments suggest that simple momentum SGD is typically just as good or better. Motivated by these results, we revisit momentum SGD and analyze its robustness in learning rate misspecification and objective curvature variation. Based on these insights, we design YellowFin, an automatic tuner for a single momentum and a single learning rate in SGD. We empirically show YellowFin converges in fewer iterations than Adam on large ResNet and LSTM models, a speedup of up to 2.8x. We also describe closed-loop YellowFin, an extension that uses a novel momentum-sensing component along with a negative-feedback loop mechanism to compensate for the dynamics of certain settings, like asynchronous parallelization. We show that closed-loop YellowFin is up to 2.7x faster than Adam under

**Implicit Generative Models**

*Rajesh Ranganath, Ian Goodfellow, Dustin Tran, David Blei, Balaji Lakshminarayanan, Shakir Mohamed*

**Parkside 1, Thu Aug 10, 08:30 AM**

Probabilistic models are a central implement in machine learning practice. They form the basis for models that generate realistic data, uncover hidden structure, and make predictions. Traditionally, probabilistic models in machine learning have focused on prescribed models. Prescribed models specify a joint density over observed and hidden variables that can be easily evaluated. The requirement of a tractable density simplifies their learning but limits their flexibility --- several real world phenomena are better described by simulators that do not admit a tractable density. Probabilistic models defined only via the simulations they produce are called implicit models.

Arguably starting with generative adversarial networks, research on implicit models in machine learning has exploded in recent years. This workshop’s aim is to foster a discussion around the recent developments and future directions of implicit models.

Implicit models have many applications. They are used in ecology where models simulate animal populations over time; they are used in phylogeny, where simulations produce hypothetical ancestry trees; they are used in physics to generate particle simulations for high energy processes. Recently, implicit models have been used to improve the state-of-the-art in image and content generation. Part of the workshop’s focus is to discuss the commonalities among applications of implicit models.

Of particular interest at this workshop is to unite fields that work on implicit models. For example:

+ Generative adversarial networks (a NIPS 2016 workshop) are implicit models with an adversarial training scheme.

+ Recent advances in variational inference (a NIPS 2015 and 2016 workshop) have leveraged implicit models for more accurate approximations.

+ Approximate Bayesian computation (a NIPS 2015 workshop) focuses on posterior inference for models with implicit likelihoods.

+ Learning implicit models is deeply connected to two sample testing and density ratio estimation.

We hope to bring together these different views on implicit models, identifying their core challenges and combining their innovations.

We invite submission of 4 page papers for posters, contributed talks, and travel awards. Topics of interests are: implicit models, approximate
Bayesian computation, generative adversarial networks, learning and inference for implicit models, implicit variational approximations, evaluation of implicit models and two sample testing. We encourage both theoretical and applied submissions.

**Schedule**

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30 AM</td>
<td>Introduction</td>
</tr>
<tr>
<td>08:40 AM</td>
<td>Kerrie Mengerson: Probabilistic Modelling in the Real World</td>
</tr>
<tr>
<td>09:10 AM</td>
<td>Yingzhen Li: Approximate Inference with Amortised MCMC</td>
</tr>
<tr>
<td>09:20 AM</td>
<td>Zenna Tavares: Adversarial Inversion for Amortized Inference</td>
</tr>
<tr>
<td>09:30 AM</td>
<td>Spotlight Session 1</td>
</tr>
<tr>
<td>09:40 AM</td>
<td>Stefano Ermon: Generative Adversarial Imitation Learning</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Jun-Yan Zhu: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>Spotlight session 2</td>
</tr>
<tr>
<td>12:30 PM</td>
<td>Lunch</td>
</tr>
<tr>
<td>02:00 PM</td>
<td>Qiang Liu: Wild Variational Inference with Expressive Variational Families</td>
</tr>
<tr>
<td>02:30 PM</td>
<td>Sanjeev Arora: Do GANs actually learn the distribution? Some theory and experiments.</td>
</tr>
<tr>
<td>03:00 PM</td>
<td>Break</td>
</tr>
<tr>
<td>03:30 PM</td>
<td>Dougal Sutherland: Evaluating and Training Implicit Generative Models with Two-Sample Tests</td>
</tr>
<tr>
<td>04:00 PM</td>
<td>Lara Mescheder: The Numerics of GANs</td>
</tr>
<tr>
<td>04:10 PM</td>
<td>Eric Nalisnick: The Amortised bootstrap</td>
</tr>
<tr>
<td>04:20 PM</td>
<td>Philemon Brakel: Maximizing Independence with GANs for Non-linear ICA</td>
</tr>
<tr>
<td>04:30 PM</td>
<td>Panel</td>
</tr>
</tbody>
</table>
Time Series Workshop

Vitaly Kuznetsov, Yan Liu, Scott Yang, Rose Yu

C4.1, Fri Aug 11, 08:30 AM

Time series data is ubiquitous. In domains as diverse as finance, entertainment, transportation and health-care, we observe a fundamental shift away from parsimonious, infrequent measurement to nearly continuous monitoring and recording. Rapid advances in diverse sensing technologies, ranging from remote sensors to wearables and social sensing, are generating a rapid growth in the size and complexity of time series archives. Thus, although time series analysis has been studied extensively, its importance only continues to grow. Furthermore, modern time series data pose significant challenges to existing techniques both in terms of the structure (e.g., irregular sampling in hospital records and spatiotemporal structure in climate data) and size. These challenges are compounded by the fact that standard i.i.d. assumptions used in other areas of machine learning are not appropriate for time series and new theory, models and algorithms are needed to process and analyse this data.

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.

Our workshop will build on the success of past two time series workshops that were held at NIPS and KDD (also co-organized by the proposers). The workshop will attract a broader audience from ICML community. In particular, when we have the KDD workshop on time series in 2015 held in Sydney, it attracts many local researchers in Australia who work on time series research or related applications. We expect the proposed workshop will be a hit given its large interest in the ICML community as well as the local interest in Sydney.

Schedule

02:00 PM  Structured Black Box
           Variational Inference for Latent Time Series Models

02:25 PM  Mehryar Mohri

03:15 PM  Afternoon Coffee Break

03:45 PM  Online Variational Bayesian Inference: Algorithms for Sparse Gaussian Processes and Theoretical Bounds

04:15 PM  Vijay M. Janakiraman

05:00 PM  Closing Remarks

Reproducibility in Machine Learning Research

Nan Ke, Anirudh Goyal, Alex Lamb, Joelle Pineau, Samy Bengio, Yoshua Bengio

C4.10, Fri Aug 11, 08:30 AM

This workshop focuses on issues of reproducibility and replication of results in the Machine Learning community. Papers from the Machine Learning community are supposed to be a valuable asset. They can help to inform and inspire future research. They can be a useful educational tool for students. They can give guidance to applied researchers in industry. Perhaps most importantly, they can help us to answer the most fundamental questions about our existence - what does it mean to learn and what does it mean to be human? Reproducibility, while not always possible in science (consider the study of a transient astronomical phenomenon like a passing comet), is a powerful criteria for improving the quality of research. A result which is reproducible is more likely to be robust and meaningful and rules out many types of experimenter error (either fraud or accidental).

There are many interesting open questions about how reproducibility issues intersect with the Machine Learning community:

* How can we tell if papers in the Machine Learning community are reproducible even in theory? If a paper is about recommending news sites before a particular election, and the results come from running the system online in production - it will be impossible to reproduce the published results because the state of the world is irreversibly changed from when the experiment was ran.  

* What does it mean for a paper to be reproducible in theory but not in practice? For example, if a paper requires tens of thousands of GPUs to reproduce or a large closed-off dataset, then it can only be reproduced in reality by a few large labs.

* For papers which are reproducible both in theory and in practice - how can we ensure that papers published in ICML would actually be able to replicate if such an experiment were attempted?

* What does it mean for a paper to have successful or unsuccessful replications?

* Of the papers with attempted replications completed, how many have been published?

* What can be done to ensure that as many papers which are reproducible in theory fall into the last category?

* On the reproducibility issue, what can the Machine Learning community
learn from other fields?

Our aim in the following workshop is to raise the profile of these questions in the community and to search for their answers. In doing so we aim for papers focusing on the following topics:

* Analysis of the current state of reproducibility in machine learning venues
* Tools to help increase reproducibility
* Evidence that reproducibility is important for science
* Connections between the reproducibility situation in Machine Learning and other fields
* Replications, both failed and successful, of influential papers in the Machine Learning literature.

### Interactive Machine Learning and Semantic Information Retrieval

**Dorota Glowacka, Wray Buntine, Petri Myllymaki**

**C4.11, Fri Aug 11, 08:30 AM**

Retrieval techniques operating on text or semantic annotations have become the industry standard for retrieval from large document collections. However, traditional information retrieval techniques operate on the assumption that the user issues a single query and the system responds with a ranked list of documents. In recent years we have witnessed a substantial growth in text data coming from various online resources, such as online newspapers, blogs, specialised document collections (e.g. arXiv). Traditional information retrieval approaches often fail to provide users with adequate support when browsing such online resources, hence in recent years there has been a growing interest in developing new algorithms and design methods that can support interactive information retrieval. The aim of this workshop is to explore new methods and related system design for interactive data analytics and management in various domains, including specialised text collections (e.g. legal, medical, scientific) as well as for various tasks, such as semantic information retrieval, conceptual organization and clustering of data collections for sense making, semantic expert profiling, and document recommender systems.

Of interest, also, is probabilistic and machine learning formulations of the interactive information retrieval task above and beyond the simple “stochastic language models” framework developed in the information retrieval community.

The primary audience of the workshop are researchers and practitioners in the area of interactive and personalised system design as well as interactive machine learning both from academia and industry.

### Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00 AM</td>
<td>Welcome</td>
</tr>
<tr>
<td>09:15 AM</td>
<td>Using Web Text Based Analytics to Gather Customer Insights</td>
</tr>
<tr>
<td>09:35 AM</td>
<td>Intent Driven Dynamic Product Ranking System for Fashion E-commerce</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Coffee Break</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Towards End-to-End Reinforcement Learning of Dialogue Agents for Information Access</td>
</tr>
<tr>
<td>11:30 AM</td>
<td>Learning to Select Relevant and Diverse Subsets</td>
</tr>
<tr>
<td>12:00 PM</td>
<td>Lunch Break</td>
</tr>
<tr>
<td>02:00 PM</td>
<td>Focused Reading: Reinforcement Learning for What Documents to Read</td>
</tr>
<tr>
<td>02:20 PM</td>
<td>CCA: Attention Based Cross Modal Retrieval on Webpages</td>
</tr>
<tr>
<td>02:40 PM</td>
<td>User Study for Measuring Linguistic Complexity and its Reduction by Technology on a Patent Website</td>
</tr>
<tr>
<td>03:00 PM</td>
<td>Poster Session</td>
</tr>
</tbody>
</table>

### Machine Learning in Speech and Language Processing

**Karen Livescu, Tara Sainath, lianglu Lu, Anton Ragni**

**C4.3, Fri Aug 11, 08:30 AM**

This workshop continues a tradition of MLSLP workshops held as satellites of ICML, ACL, and Interspeech conferences. While research in speech and language processing has always involved machine learning (ML), current research is benefiting from even closer interaction between these fields. Speech and language processing is continually mining new ideas from ML and ML, in turn, is devoting more interest to speech and language applications. This workshop is a venue for locating and incubating the next waves of research directions for interaction and collaboration. The workshop will (1) discuss emerging research ideas with potential for impact in speech/language and (2) bring together relevant researchers from ML and speech/language who may not regularly interact at conferences. Example topics include new directions for deep learning in speech/language, reinforcement learning, unsupervised/semi-supervised learning, domain adaptation/transfer learning, and topics at the boundary of speech, text, and other modalities.

### Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:15 AM</td>
<td>Jason Weston: (Towards) Learning from Conversing</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Break</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Tasha Nagamine: Feature Representation and Transformation in Multilayer Perceptron Acoustic Models</td>
</tr>
</tbody>
</table>
Abstracts (3):

**Abstract 1:** *Jason Weston: (Towards) Learning from Conversing in Machine Learning in Speech and Language Processing, 09:15 AM*

An (end-to-end) dialogue agent might be far from knowing and understanding everything, but _if_ it can learn _while_ it is conversing with humans, maybe it can move towards that goal? We look at some key ingredients we will need: (i) ability to learn from textual feedback, i.e. the things said to it, (ii) ability to ask (useful) questions and learn from the replies, (iii) bootstrapping the model so that it doesn't take 6 months - 2 years to say something (like a human baby), (iv) making this a elegant unified system, and not a bunch of hacks. This talk describes joint work with Jiwei Li, Alexander H. Miller, Sumit Chopra and Marc'Aurelio Ranzato.

**Abstract 3:** *Tasha Nagamine: Feature Representation and Transformation in Multilayer Perceptron Acoustic Models in Machine Learning in Speech and Language Processing, 10:30 AM*

While deep learning has shown great success in recent years, how the nodes in different layers of neural networks represent both the input and the properties of the network function remain unknown. We present an empirical and joint framework to study the encoding properties of node activations in hidden and output layers of the network, and to construct the equivalent linear transformation applied to each data point. These methods are used to discern and quantify the properties of feed-forward neural networks trained to map acoustic features to phoneme labels. We show a selective and progressively nonlinear warping of the feature space in which the most discriminant dimensions of the input samples are emphasized. Analyzing the sample-dependent linear transforms applied to each data sample shows that categorization is achieved by forming prototypical templates to explicitly model the all the variations of each class. This work provides a comprehensive analysis of representation and computation of neural networks and provides an intuitive account of how deep neural networks perform classification.

**Abstract 4:** *Joelle Pineau: Neural Models for Interactive Dialogue Systems in Machine Learning in Speech and Language Processing, 11:15 AM*

Reinforcement learning provides a rich framework to cast the problem of learning a dialogue strategy for conversational AI agents. In this talk I will present recent results on building dialogue systems from large corpuses using neural architectures. I will highlight several challenges related to data acquisition, algorithmic development, performance evaluation, and user studies.
Abstract 1: Opening in Private and Secure Machine Learning, 08:30 AM

Introductory comments from organizers

Abstract 2: Deep Learning with Differential Privacy: Two Approaches in Private and Secure Machine Learning, 08:35 AM

We discuss two recently proposed approaches towards offering differential privacy for training data. The first approach modifies the SGD procedure so that the updates to the model's weights are provably differentially private. The second approach, called Private Aggregation of Teacher Ensembles (PATE) is particularly suitable for training classifiers. PATE combines, in a black-box fashion, multiple models trained with disjoint datasets and aggregates their results—enforcing differential privacy—to train a "student" model who is never directly exposed to sensitive data.

Abstract 3: PrivIT: Private and Sample Efficient Identity Testing in Private and Secure Machine Learning, 09:20 AM

We develop differentially private hypothesis testing methods for the small sample regime. Given a sample $D$ from a categorical distribution $p$ over some domain $\Sigma$, an explicitly described distribution $q$ over $\Sigma$, some privacy parameter epsilon, accuracy parameter alpha, and requirements beta_I and beta_II for the type I and type II errors of our test, the goal is to distinguish between $p=q$ and $\text{div}(p,q) \gg \alpha$. We provide theoretical bounds for the sample size $|D|$ so that our method both satisfies $(\epsilon,0)$-differential privacy, and guarantees $\text{beta}_I$ and $\text{beta}_II$ type I and type II errors. We show that differential privacy may come for free in some regimes of parameters, and we always beat the sample complexity resulting from running the chi^2-test with noisy counts, or standard approaches such as repetition for endowing non-private chi^2-style statistics with differential privacy guarantees. We experimentally compare the sample complexity of our method to that of recently proposed methods for private hypothesis testing.

Abstract 4: Differentially Private Learning of Undirected Graphical Models using CGMs in Private and Secure Machine Learning, 09:40 AM

We investigate the problem of learning discrete, undirected graphical models in a differentially private way. Approaches to this problem range from privileged algorithms that conduct learning completely behind the privacy barrier to schemes that release private summary statistics paired with algorithms to learn parameters from those statistics. We show that the approach of releasing noisy sufficient statistics using the Laplace mechanism achieves a good trade-off between privacy, utility, and practicality. Naive learning algorithms that use the noisy sufficient statistics “as is” outperform general-purpose differentially private learning algorithms. However, it ignores knowledge about the data generating process, is on uncertain theoretical foundations, and exhibits certain pathologies. We develop a more principled approach that applies the formalism of collective graphical models to perform inference over the true sufficient statistics within an expectation-maximization framework. We show that this learns better models than competing approaches on both synthetic data and on real human mobility data used as a case study.

Abstract 5: Deep Learning with Differential Privacy: Two Approaches in Private and Secure Machine Learning, 10:50 AM

Differential privacy has been recognized as the most suitable statistical data privacy definition in the academic community [1] and has spurred a decade-long effort to develop algorithms that satisfy the definition. Although by now, broadly speaking, differentially private algorithms are known for most machine learning primitives [2] including deep learning [3], they have seen limited adoption in practice. In particular, the two known large-scale commercial deployments of differential privacy (Google’s RAPPOR [4] and Apple’s learning system [5]) both operate in the so-called “local” model, where the data privatization occurs before it reaches the data collector. In contrast, most of the academic work has focused on developing algorithms for the “central” or “trusted data curator” models, in which all user data is collected by the curator before privatization techniques are applied [2].

In this talk, I will compare and contrast the “local” and the “central” models from the perspectives of privacy guarantees they provide for the users and the utility of the data for the data collector. I will then discuss implications these competing considerations may have for adoption of one model over the other by companies, depending on the size of the company’s user base [6].

I will then describe a hybrid model of differential privacy proposed by [7], that considers a combination of regular and opt-in users who desire the differential privacy guarantees of the local privacy model and the central model, respectively. Using the task of privately computing the head of a search log, I will demonstrate that within this model, it is possible to design a new type of blended algorithm, that provides significant improvements in the utility of obtained data. I will present both the new algorithm and experimental results of its performance (attaining NDCG values exceeding 95% for reasonable privacy parameter values) on two large search click data sets.

I will conclude by arguing that since many companies already rely on a group of beta testers with whom they have higher levels of mutual trust,
the hybrid model is appropriate for many real-world scenarios. Combined with the findings of the significant improvements in utility that operating in this model may bring even when the percentage of users opting-in to the “central” privacy model is low [7]. I will build a case for further algorithmic and ML research in the hybrid model. It can be a step to a viable approach for achieving broader practical adoption of differential privacy and a lens for putting the collaboration of differential privacy researchers with researchers from the secure distributed learning communities on a formal footing.

Abstract 10: Privacy-preserving machine learning and data mining using the Sharemind platform in Private and Secure Machine Learning, 02:00 PM

We have built the Sharemind platform for privacy-preserving computations, based on additively sharing the secrets among three parties and a large, passively secure protocol set built on top of this representation of private data. Using this protocol set, we have built various privacy-preserving numerical, statistical, anomaly detection, and combinatorial applications and prototypes on top of Sharemind, some of them for use-cases with a large number of inputs and correspondingly large computation size.

In this talk, I explain the basics of Sharemind and the construction of applications which may be of interest to the ML community. This covers the linear regression and principal component analysis, genetic algorithms, frequent itemset mining. I also explain how differentially private computations have been done on top of Sharemind: we implemented the sample-and-aggregate mechanism by Nissim et al. and Smith, as well as a technique to keep track of personalized differential privacy budgets by Ebadi et al. There are important differences between normal and privacy-preserving applications, when it comes to the relative efficiency of certain algorithmic steps. These have affected the construction of the applications that I’m going to talk about.

Abstract 11: Privacy-preserving entity resolution and logistic regression on encrypted data in Private and Secure Machine Learning, 02:45 PM

We consider a scenario of two data providers, A and B, each of whom manage a dataset of private information consisting of two different feature sets related to common customers/entities. They jointly aim to learn a linear model using stochastic gradient algorithms like SGD/SAG. The setting is federated learning, where data is kept locally and a shared model is learned on top of local computation. Notice that, in contrast with the large majority of work on distributed learning, in our scenario data is split vertically, i.e., by features. We also assume that only A knows the target variable. We propose a secure system solving the problem in two phases: privacy-preserving entity resolution and logistic regression over encrypted data. With the aid of a coordinator, C, we design a three-party protocol that is secure under the honest-but-curious adversary model. Our system allows A and B to learn a classifier collaboratively, without either exposing their data in the clear or even sharing which entities they have in common.

Privacy-preserving entity resolution. When the dataset is vertically partitioned across multiple organisations the problem arises of how to identify the corresponding entities, namely entity resolution. Entity resolution is usually done on identifying features such as name, address, etc. We perform privacy-preserving entity resolution using anonymous linkage codes, which map entity information onto a code from which it is impossible to reconstruct any entity data. We use the cryptographic longterm key (CLK) anonymous linkage code, which provides both privacy and error tolerance. The CLKs are used in a comparison function which estimates the likelihood that two entities match. Parties A and B create CLKs for each entry in their datasets and sent them to C, which performs the entity resolution. The protocol results in two permutations, one for each data provider, and a mask. The permutations describe how A and B should rearrange their dataset so as to be consistent with each other and the mask. The mask specifies whether a row corresponds to a record available in both datasets, thus a record which will be used for learning; it also implicitly excludes records that are not matched across A and B. The mask itself is only sent to data providers in encrypted form to prevent revealing the common entities. For simplicity, we omit mention of the permutations and mask in what follows.

Logistic regression on encrypted data Learning is performed on data encrypted with the Paillier partially homomorphic encryption scheme, an asymmetric scheme which permits both adding encrypted values and scaling encrypted values by unencrypted ones. These properties allow us to implement most of the linear algebra necessary for gradient descent optimization on encrypted data. Only C possesses the private key. We approximate the logistic loss and its gradient via Taylor expansion around 0 which results in polynomials that A and B can evaluate collaboratively and securely by only transmitting intermediate values that are encrypted with the Paillier scheme. Experimental results have shown that we can match the accuracy of exact logistic loss using a merely second-order Taylor approximation to the loss (hence linear approximation to the gradient) at the price of rescaling features into the interval [−1, 1] and of applying L1 /L2 regularization. Party C orchestrates the optimization algorithm, taking care of the stochastic learning parameters (regularization, learning rate, momentum, etc.), triggering gradient computations by A and B, and using the logistic loss on hold-out data to determine when to stop training so as to avoid overfitting. We have proven the practicality of our system in commercial deployments. Our system is capable of scaling to millions of records with hundreds of features.

Abstract 13: Detecting Causative Attacks using Data Provenance in Private and Secure Machine Learning, 03:30 PM

The reliance of machine learning methods on quality training data presents a security vulnerability in which adversaries may inject poisonous samples into the training dataset to manipulate the learned classifier. A highly-publicized example of this is the recent attack on Microsoft’s AI chat bot, Tay, which learned offensive and racist language from Twitter users. Defending against these types of attacks, called causative attacks, is particularly challenging in online learning and other environments where the model must be periodically retrained to account for dataset shifts.

One countermeasure, called Reject on Negative Impact (RONI) (Nelson et al., 2009), detects whether a given sample is poisoned by comparing the performance of the classifier on a trusted test set before and after the sample is added to a trusted training set. Compared to clustering-based methods, RONI is likely to perform better on heterogeneous datasets. However, this method requires that the classifier be re-trained for each sample, which may be infeasible in big-data settings. Moreover, it relies strongly on the coverage of the test set. Finally, it requires that some of the collected data is trusted.

In this talk, we demonstrate how data provenance can be used to aid in the detection of poisoned data. By utilizing a provenance framework,
cryptographically protected meta-data describing the origin and history of each data point can be collected. This may include information about the device from which the data was gathered, its firmware version, user id, and timestamp among others. Our method uses this provenance meta-data to segment the untrusted data into groups where the probability of poisoning is highly correlated across samples in each group. The data points in each group are then evaluated together by comparing the performance of the classifier trained with and without that group. Since groups of data points are evaluated together, this method reduces the number of times the classifier must be trained and amplifies the effect of the evaluated data points on the classifier, thereby improving accuracy.

Additionally, we present a methodology to address cases where the entire dataset is untrusted. Using provenance meta-data, the dataset is first segmented into groups that are evaluated together. It is then split into a training and a test set, and, for each group, classifiers are trained with and without the group. The performance of each classifier is then evaluated on a test set with data points from the group removed, preventing poisoned data in the test set from manipulating the evaluation of its own group. We present a detailed analysis of new attacks that arise when trusted data are unavailable and provide defense mechanisms to prevent them. Lastly, we show the results of simulations that evaluated the ability of our methods to detect poisoning attacks on logistic regression classifiers.

Deep Structured Prediction

Isabelle Augenstein, Kai-Wei Chang, Gal Chechik, Bert Huang, Andre Filipe Torres Martins, Ofer Meshi, Alex Schwing, Yishu Miao

C4.5, Fri Aug 11, 08:30 AM

In recent years, deep learning has revolutionized machine learning. Most successful applications of deep learning involve predicting single variables (e.g., univariate regression or multi-class classification). However, many real problems involve highly dependent, structured variables. In such scenarios, it is desired or even necessary to model correlations and dependencies between the multiple input and output variables. Such problems arise in a wide range of domains, from natural language processing, computer vision, computational biology and others.

Some approaches to these problems directly use deep learning concepts, such as those that generate sequences using recurrent neural networks or that output image segmentations through convolutions. Others adapt the concepts from structured output learning. These structured output prediction problems were traditionally handled using linear models and hand-crafted features, with a structured optimization such as inference. It has recently been proposed to combine the representational power of deep neural networks with modeling variable dependence in a structured prediction framework. There are numerous interesting research questions related to modeling and optimization that arise in this problem space.

This workshop will bring together experts in machine learning and application domains whose research focuses on combining deep learning and structured models. Specifically, we aim to provide an overview of existing approaches from various domains to distill from their success principles that can be more generally applicable. We will also discuss the main challenges that arise in this setting and outline potential directions for future progress. The target audience consists of researchers and practitioners in machine learning and application areas.

Picky Learners: Choosing Alternative Ways to Process Data.

Corinna Cortes, Kamalika Chaudhuri, Giulia DeSalvo, Ningshan Zhang, Chicheng Zhang

C4.6, Fri Aug 11, 08:30 AM

Picky Learners consists of a broad range of learning scenarios where the learner does not simply process every data point blindly, but instead can choose to incorporate them in alternative ways. Despite the growing costs of processing and labelling vast amounts of data, only isolated efforts have tackled this problem primarily in the areas of active learning, learning with rejection and on-line learning with feedback graphs.

In active learning, the learner can choose whether or not to query for a label of each data point, thereby paying different costs for each data point. A key advantage in this setting is that the number of examples queried to learn a concept may be much smaller than the number of examples needed in standard supervised learning. More recently, some have used variations of confidence-based models to determine which labels to query. Confidence-based models lie under the more general framework of learning with rejection, which is a key learning scenario where the algorithm can abstain from making a prediction, at the price of incurring a fixed cost. In this scenario, our picky learners can thus choose to abstain from providing a label. In the on-line setting, one can cast learning with rejection under the more general topic of on-line learning with feedback graphs, a setting that interpolates between bandit and full expert scenario in that the player observes a variety of different expert losses after choosing an action. On-line learning with feedback graphs can then in turn be connected back to active learning where depending on the feedback graph only certain labels are requested.

In short, our picky learners can choose to query for the label (active learning), choose to abstain on the label (learning with rejection) or choose to receive different expert losses (on-line learning with feedback graphs). All of three of these fields attempt in different ways to reduce the cost of processing the data by allowing for picky learners, but the connections between these topics has not been fully explored in terms of both theory and practice. The goal of this workshop is then to bring together researchers and practitioners in these three areas in order to bridge the gap between active learning, learning with rejection, and on-line learning with feedback graphs. We expect that the fruitful collaborations started in this workshop will result in novel research that will help develop each field.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Speaker(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00 AM</td>
<td>Opening Remarks</td>
<td></td>
</tr>
<tr>
<td>09:05 AM</td>
<td>Corralling a Band of Bandit Algorithms</td>
<td>Agarwal</td>
</tr>
<tr>
<td></td>
<td>Why adaptively collected data have negative bias and how to correct for it.</td>
<td>Nie</td>
</tr>
<tr>
<td>10:05 AM</td>
<td>Morning Break</td>
<td></td>
</tr>
</tbody>
</table>
Abstract 2: **Corralling a Band of Bandit Algorithms in Picky Learners: Choosing Alternative Ways to Process Data.**, Agarwal, 09:05 AM

Ensemble techniques are well known as a mechanism to increase the adaptivity and robustness of machine learning methods. However, for interactive learners that drive data acquisition, building an ensemble presents novel challenges, since each learner might seek to collect different data. In this talk, we examine this question from the lens of contextual bandit learning. We study the problem of combining multiple bandit algorithms (that is, online learning algorithms with partial feedback) with the goal of creating a master algorithm that performs almost as well as the best base algorithm if it were to be run on its own. We show how the task is hopeless in general, but present a new approach to achieve this goal under natural assumptions. As examples, we present two main applications. The first is to create an algorithm that enjoys worst-case robustness while at the same time performing much better when the environment is relatively easy. The second is to create an algorithm that works simultaneously under different assumptions of the environment, such as different priors or different loss structures.

Abstract 3: **Why adaptively collected data have negative bias and how to correct for it. In Picky Learners: Choosing Alternative Ways to Process Data.**, Nie, 09:45 AM

From scientific experiments to online A/B testing, the previously observed data often affects how future experiments are performed, which in turn affects which data will be collected. Such adaptivity introduces complex correlations between the data and the collection procedure. In this paper, we prove that when the data collection procedure satisfies natural conditions, then sample means of the data have systematic negative biases. As an example, consider an adaptive clinical trial where additional data points are more likely to be tested for treatments that show initial promise. Our surprising result implies that the average observed treatment effects would underestimate the true effects of each treatment. We quantitatively analyze the magnitude and behavior of this negative bias in a variety of settings. We also propose a novel debiasing algorithm based on selective inference techniques. In experiments, our method can effectively reduce bias and estimation error.

Abstract 5: **Label Efficient Learning by Exploiting Multi-class Output Codes. in Picky Learners: Choosing Alternative Ways to Process Data.**, Dick, 10:25 AM

We present a new perspective on the popular multi-class algorithmic techniques of one-vs-all and error correcting output codes. Rather than studying the behavior of these techniques for supervised learning, we establish a connection between the success of these methods and the existence of label efficient learning procedures. We show that in both the realizable and agnostic cases, if output codes are successful at learning from labeled data, they implicitly assume structure on how the classes are related. By making that structure explicit, we design learning algorithms to recover the classes with low label complexity. We provide results for the commonly studied cases of one-vs-all learning and when the codewords of the classes are well separated. We additionally consider the more challenging case where the codewords are not well separated, but satisfy a boundary features condition that captures the natural intuition that every bit of the codewords should be significant.

Abstract 6: **Learning with Rejection in Picky Learners: Choosing Alternative Ways to Process Data.**, Desalvo, 10:45 AM

We introduce a novel framework for classification with a rejection option that consists of simultaneously learning two functions: a classifier along with a rejection function. We present a full theoretical analysis of this framework including new data-dependent learning bounds in terms of the Rademacher complexities of the classifier and rejection families as well as consistency and calibration results. These theoretical guarantees guide us in designing new algorithms that can exploit different kernel-based hypothesis sets for the classifier and rejection functions. We compare and contrast our general framework with the special case of confidence-based rejection for which we devise alternative loss functions and algorithms as well. We report the results of several experiments showing that our kernel-based algorithms can yield a notable improvement over the best existing confidence-based rejection algorithm.

Abstract 7: **Active Learning in Expert Systems Experiments on StackExchange Data. in Picky Learners: Choosing Alternative Ways to Process Data.**, Hegde, 11:05 AM

We study adaptive matching in expert systems. Expert systems consist of a set of servers or experts of varying expertise, to which clients or tasks of varying types arrive. The task type is apriori unknown, and the task must be matched to the appropriate expert. We consider in particular the setting of Q&A platforms, focussing on real data from the...
We will talk about sequential learning problems where the available actions (arms) form a graph. These settings naturally arise in recommender systems, advertising, or sensor networks. Specifically, at each round, we are asked to pick a node representing an item to recommend or a sensor to take a reading from; after which we receive a bandit feedback in order to optimize a given performance measure (regret). To address these settings, we can always ignore the graph structure and use known algorithms for multi-armed bandits. However, their performance scales unfavorably with the number of nodes $N$ and $S$, for example, $O(\sqrt{NS})$, which is undesirable when $NS$ means a thousand of sensors or a million of movies. We will describe several graph-bandit problems and show how to use their graph structure to design new algorithms with faster learning rates, scaling not with $NS$ but with graph-dependent quantities, often much smaller than $NS$ in real-world graphs. In particular, when we receive side observations from the neighboring nodes, we can replace $NS$ in the regret guarantees with the independence number, or, in the noisy case, with the effective independence number.

Abstract 15: Active Learning from Peers in Picky Learners: Choosing Alternative Ways to Process Data., Murugesan 04:10 PM

This paper addresses the challenge of learning from peers in an online multitask setting. Instead of always requesting a label from a human oracle, the proposed method first determines if the learner for each task can acquire that label with sufficient confidence from its peers either as a task-similarity weighted sum, or from the single most similar task. If so, it saves the oracle query for later use in more difficult cases, and if not it queries the human oracle. Experiments over three multitask learning benchmark datasets show clearly superior performance over baselines such as assuming task independence, learning only from the oracle and not learning from peer tasks.

Abstract 17: Online Learning with Local Permutations and Delayed Feedback in Picky Learners: Choosing Alternative Ways to Process Data., 05:10 PM

We propose an Online Learning with Local Permutations (OLL) setting, in which the learner is allowed to slightly permute the order of the loss functions generated by an adversary. On one hand, this models natural situations where the exact order of the learner’s responses is not crucial, and on the other hand, might allow better learning and regret performance, by mitigating highly adversarial loss sequences. Also, with random permutations, this can be seen as a setting interpolating between adversarial and stochastic losses. In this paper, we consider the applicability of this setting to convex online learning with delayed feedback, in which the feedback on the prediction made in round $t$ arrives with some delay $\tau$. With such delayed feedback, the best possible regret bound is well-known to be $O(\sqrt{t\tau})$. We prove that by being able to permute losses by a distance of at most $M$ (for $M=\tau$), the regret can be improved to $O(\sqrt{t\tau}(1+\sqrt{\tau^2/M}))$, using a Mirror-Descent based algorithm which can be applied for both Euclidean and non-Euclidean geometries. We also prove a lower bound, showing that for $M<\sqrt{3}$, it is impossible to improve the standard $O(\sqrt{t\tau})$ regret bound by more than constant factors. Finally, we provide some experiments validating the performance of our algorithm.

Reliable Machine Learning in the Wild

Dylan Hadfield-Menell, Jacob Steinhardt, Adrian Weller, Smitha Milli

C4.7, Fri Aug 11, 08:30 AM

When can we trust that a system that has performed well in the past will continue to do so in the future? Designing systems that are reliable in the wild is essential for high stakes applications such as self-driving cars and automated surgical assistants. This workshop aims to bring together researchers in diverse areas such as reinforcement learning, human-robot interaction, game theory, cognitive science, and security to further the field of reliability in machine learning. We will focus on three aspects — robustness (to adversaries, distributional shift, model misspecification, corrupted data); awareness (of when a change has
Human in the Loop Machine Learning

Richard Nock, Cheng Soon Ong

C4.8, Fri Aug 11, 08:30 AM

For details see: http://machlearn.gitlab.io/hitl2017/

As machine learning systems become more ubiquitous in everybody’s
day-to-day life or work, society and industry is in an intermediate state
between fully manual and fully automatic systems. The gradient
undoubtedly points towards full automation, but moving forward in this
direction is going to face increasing challenges due to the fact that
current machine learning research tends to focus on end to end systems,
which puts aside the fact that for practical applications there are still gaps
or caveats in the automation. Parts of these come from the presence of
(or the necessity to have) the Human in the Loop.

There are two main locations for the Human in the automated system: (i)
upstream, in which case the focus is mainly in the inputs of the algorithm.
This can be essential for personalised assistants, that describe
environments where the machine learning method is tightly embedded
into the system. Such environments pose additional challenges related to
privacy at large; (ii) downstream: other domains have machine learning
approaches analyse parts of the data, and human experts use the results
and intuition to make decisions.

The Human dependences between these two locations is also neither
straightforward nor acrylic — some applications tend to have feedback
effects on data as actions or interventions are undertaken based on
machine learning predictions. Furthermore there are often very few
rounds of decision making in practice, but each round may affect the
statement of the problems related to the Human presence, as witnessed
for example by eventual privacy leakages.

This workshop aims to bring together people who are working on
systems where machine learning is only part of the solution. Participants
will exchange ideas and experiences on human in the loop machine
learning.

Topics of interest include:
- System architectures that allow for human decision making
- User interfaces for interacting with machine learning systems
- Validation of human in the loop software systems
- Viewpoints from traditional fields such as reinforcement learning and
Bayesian optimisation
- Challenges related to the human presence in the loop (privacy, bias,
fairness, etc.)
- Case studies of deployed machine learning

Machine Learning for Music Discovery

Erik Schmidt, Oriol Nieto, Fabien Gouyon, Gert Lanckriet

C4.9, Fri Aug 11, 08:30 AM

The ever-increasing size and accessibility of vast music libraries has
created a demand more than ever for machine learning systems that are
capable of understanding and organizing this complex data. While this
topic has received relatively little attention within the machine learning
community, it has been an area of intense focus within the community of
Music Information Retrieval (MIR), where significant progress has been
made, but these problems remain far from solved.

Furthermore, the recommender systems community has made great
progress in terms of collaborative feedback recommenders, but these
approaches suffer strongly from the cold-start problem. As such,
recommendation techniques often fall back on content-based machine
learning systems, but defining musical similarity is extremely challenging
as myriad features all play some role (e.g., cultural, emotional, timbral,
rhythmic).

We seek to use this workshop to bring together a group of world-class
experts to discuss these challenges and share them with the greater
machine learning community. In addition to making progress on these
challenges, we hope to engage the machine learning community with our
nebulous problem space, and connect them with the many available
datasets the MIR community has to offer (e.g., AcousticBrainz, Million
Song Dataset), which offer near commercial scale to the academic
research community.

Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Session Title</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30 AM</td>
<td>Welcome Remarks</td>
<td>Schmidt</td>
</tr>
<tr>
<td>08:40 AM</td>
<td>Matrix Co-Factorisation and Applications to Music Analysis</td>
<td>Essid</td>
</tr>
<tr>
<td>09:20 AM</td>
<td>Learning a Large-Scale Vocal Similarity Embedding</td>
<td>Kumar</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Aligned Hierarchies - A Multi-Scale Structure-Based Representation for Music</td>
<td>Kinnaird</td>
</tr>
<tr>
<td>11:10 AM</td>
<td>Mining Creation Methods from Music Data for Automated Content Generation</td>
<td>FUKAYAMA</td>
</tr>
<tr>
<td>02:00 PM</td>
<td>NSynth: Unsupervised Understanding of Musical Notes</td>
<td>Resnick</td>
</tr>
<tr>
<td>02:40 PM</td>
<td>Multi-Level and Multi-Scale Feature Aggregation Using Sample-level Deep Convolutional Neural Networks for Music Classification</td>
<td>Lee</td>
</tr>
</tbody>
</table>
Reinforcement Learning Workshop

Doina Precup, Balaraman Ravindran, Pierre-Luc Bacon

Parkside 1, Fri Aug 11, 08:30 AM

The workshop will contain presentations of late-breaking reinforcement learning results in all areas of the field, including deep reinforcement learning, exploration, transfer learning and using auxiliary tasks, theoretical result etc, as well as applications of reinforcement learning to various domains. A panel discussion on the most interesting and challenging current research directions will conclude the workshop.

Schedule

08:45 AM Introductory Remarks

09:00 AM Some experiments with learning hyperparameters, transfer, and multi-task leaning  Ravindran

09:30 AM Talk  Finn

10:00 AM Break

10:30 AM Achieving Above-Human Performance on Ms. Pac-Man by Reward Decomposition  van Seijen

11:00 AM Exploration methods for options  Lazaric

11:30 AM The effects of memory replay in reinforcement learning  Liu

12:00 PM Lunch

01:30 PM Asynchronous data aggregation for training end-to-end visual control networks  Monfort

01:50 PM On the reproducibility of policy gradient experiments  Islam

02:10 PM Reinforcement learning for the MALMO Minecraft challenge  Berariu

02:30 PM On the convergence of Tree backup and other reinforcement learning algorithms  Touati

02:50 PM Transfer learning using successor state features  Lehner

03:10 PM Coffee Break  Ravindran, Finn, Lazaric, Hofmann, Bellemare

03:30 PM Panel Discussion  Ravindran, Finn, Lazaric, Hofmann, Bellemare

04:10 PM Ephemeral Context to Support Robust and Diverse Recommendations  Rodriguez

04:30 PM Closing Remarks  Schmidt
CONVENTION CENTRE

International Convention Centre Sydney (ICCS)
Darling Road

All Events Will Take Place In The Main Building Here

Floor 2 - Parkside Ballrooms Only

Floor 3 - Darling Harbour Theater

Floor 4 - Most Events

Floor 4 - Workshops