Bolei's ICML'17 Memo (Aug.16, 2017)

Memo Highlights:

- **Tutorial on interpretable machine learning:**
  - A very high-level overview on the the interpretability of machine learning
  - slide:http://people.csail.mit.edu/beenkim/papers/BeenK_FinaleDV_ICML2017_tutorial.pdf
- **Best paper: Understanding the black-box predictions via influence functions** (https://arxiv.org/pdf/1703.04730.pdf)
  - Topic: How can we explain the predictions of a black-box model.
  - Use the influence function to trace a model's prediction back to its training data, and identify the training points most responsible for a given prediction.
- **A closer look at memorization in deep neural network** (https://arxiv.org/pdf/1706.05394.pdf)
  - The generalization of neural network becomes an important issue. This paper analyzes what is memorization and how it happens.
  - Another relevant paper (Dziagite and Roy 2017): Computing nonvacuous generalization bounds for deep (stochastic) neural networks with many more parameters than training data" Dziugaite, Gintaire and Roy. Arxiv 2017
  - It mentioned another relevant paper: unsupervised learning by predicting noise, ICML'17
  - It designed an experiment, to compare the properties of the networks learned on real data and random shuffled labeled data.
- **Cognitive psychology for deep neural networks: a shape bias case study (https://arxiv.org/pdf/1706.08606.pdf)**
  - DeepMind's project page: Interpreting Deep Neural networks using Psychology: https://deepmind.com/blog/cognitive-psychology/
  - Several psychology experiments are evaluated on googlenet with different initial training seeds.
  - They claim that though the final accuracy is the same, the variation among different models shows the initialization and stopping time matter to the model training.
  - They show that the deep CNN also get human-like bias (shape bias).

- **On expressive power of deep neural networks** (http://proceedings.mlr.press/v70/raghu17a/raghu17a.pdf)
    - On the problem of neural network expressivity, which seeks to characterize how structural properties of a neural network affect the functions it is able to compute
- **Test of time award paper: Combining online and offline knowledge in UST** (http://www.machinelearning.org/proceedings/icml2007/papers/387.pdf)

David Silver has been working on the Monte-Carlo Tree search for playing Go for more than 10 years! Such amazing!

- Online: Monte-Carlo Simulation
- Offline knowledge: offline value function, trained from TD(lamda), offline policy
- MoGO = MCTS + RAVE + Simple value function + rollout with custom policy
- Then there is AlphaGo, deep RL to train policy network and value network, then use the value network to reduce the depth and breadth in MCTS.
- QA: the interpretability of deep neural networks asked by Andrew from Umass.

Memo Details:

Overview of the ICML'17 conference:
~400 papers are accepted, 2500 people registered to attend, 8 parallel oral sessions. Each paper has an oral presentation, the poster session is at night.

Workshops & Tutorials

Tutorial on Interpretable Machine Learning

- Webpage: (http://people.csail.mit.edu/beenkim/icml_tutorial.html).
- Given by Been Kim (slide: http://people.csail.mit.edu/beenkim/papers/BeenK_FinaleDV_ICML2017_tutorial.pdf)

Tutorial on Seq2seq learning

- https://sites.google.com/view/seq2seq-icml17, Given by Oriol Vinyals
- Neural embedding + Recurrent language models
- A review on RNN, attention, PixelRNN for modeling the natural image statistics.

- Slides: https://docs.google.com/presentation/d/1quIMxEEPEf5EkRHc2USQaoJRC4QNX6_KomdZTBMBWjk/edit#slide=id.g2349e758b6_0_293
- One mentioned paper (https://arxiv.org/pdf/1603.08983.pdf): Adaptive computation time for recursive neural network, it could automatically decide how many recursive steps taken to reach some prediction. Similarly, the size and the depth of the receptive fields are relevant to the complexity of visual patterns. And recognizing different images require different difficulty, some are easier and some are harder. Is there a way to adaptively adjust the number of total layers used for inferring a given image?
- I have another relevant idea: how to evaluate the expressiveness of a receptive field. The expressiveness of a receptive field is decided by the RF size and the depth of layer. How to evaluate the extreme case of visual patterns which could be represented by a receptive field?

Tutorial on Deep Reinforcement Learning, decision making, and control

- Given by Sergey Levine and Chelsea Finn.
- https://sites.google.com/view/icml17deeprl
- Slide: https://drive.google.com/file/d/0B_j5EZzjlxchV2l3TGJPdTljM1k/view
- Last several slides about the frontiers of deep learning for decision making, open problems:
    - How can our agents learn effectively using (much) less experience?
    - How do we ensure that our agents explore safely (ensuring safety)
    - Where do rewards come from

Workshops:

- ICML’17 workshop on Human Interpretability in Machine Learning
    - **https://sites.google.com/view/whi2017/home**
- ICML’17 workshop on Visualization for Deep Learning.
    - http://icmlviz.github.io/schedule/
    - Bolei on Interpreting Deep Visual Representations: http://icmlviz.github.io/assets/papers/6.pdf
- ICML’17 workshop on Video gaming and machine learning
    - https://syhw.github.io/vgml_workshop_icml2017/
    - Video gaming becomes a frontier for industrial labs such as Google DeepMind, Facebook AI Research.

Day1

Keynote 1: Causal learning given by Bernard Schlopf

- Mentioned one new textbook on causal inference: Elements of Causaul inference: PDF draft: http://www.math.ku.dk/~peters/jonas_files/bookDRAFT11-online-2017-06-28.pdf
- Mentioned a benchmark dataset with 106 cause-effect relations. How to build a better causality model?

Some deep learning theory papers:

- **The loss surface of deep and wide neural networks:** http://proceedings.mlr.press/v70/nguyen17a/nguyen17a.pdf
  - it argues that all local minima are global optimal given that the number of hidden units of one layer of the network is larger than the number of training points.
- **On expressive power of deep neural networks:** http://proceedings.mlr.press/v70/raghu17a/raghu17a.pdf
  - On the problem of neural network expressivity, which seeks to characterize how structural properties of a neural network affect the functions it is able to compute

PixelCNN

- **Video Pixel CNN** (https://arxiv.org/pdf/1610.00527.pdf)
  - learning the distribution of the natural videos.
  - p(x) = sum_t sum_h sum_c P(x_t,h,c| x_>), dependency across time, RGB, and space .
  - Model is evaluated on moving MNIST and robot manipulation dataset

Meta-learning

- **Model-agnostic meta-learning for fast adaption of deep networks** (https://arxiv.org/pdf/1703.03400.pdf)
  - How to learn from few samples: few-shot learning.
  - Meta-learner learns how to update the parameters of the learner's model.
  - Algorithm of MAML is proposed for meta-learning.
  - This paper does a good job to generalize the MAML to a lot of scenario, both in supervised learning and reinforcement learning, image classification and virtual environment movement.
  - Hmm, actually it looks like a very simple heuristics, but it still can write it into a decent paper…

Meta-learning

- **AdaNet: adaptive structural learning of artificial neural networks** (http://www.cs.nyu.edu/~mohri/pub/adanet.pdf)
  - Algorithms for adaptively learning both the structure of the network and its weights.

Best paper:

- **Understanding the black-box predictions via influence functions** (https://arxiv.org/pdf/1703.04730.pdf)
  - How can we explain the predictions of a black-box model:
  - Use the influence function to trace a model's prediction back to its training data, and identify the training points most responsible for a given prediction.

Other interesting papers:

- Parallel multiscale autoregressive density estimation: could generate very high quality images in a parallel way, improvement over previous pixelCNN.
- Warped convolutions: efficient invariance to spatial transformations.
- Curiosity-driven exploration by self-supervised prediction (https://arxiv.org/pdf/1705.05363.pdf) : very interesting paper from Alexei Efros and Trevor Darrell, to let the agent always explore some unexplored actions by learning features.
- Schema Networks from Vicarious AI (https://arxiv.org/pdf/1706.04317.pdf): Zero-shot transfer with a generative causal model of intuitive physics: pretty interesting paper to model the environment explicitly, then learning the links among different attributes, so the learned weights are human-interpretable and could be transferred to other tasks.

Day2:

- **Test of time paper: Combining online and offline knowledge in UST** (http://www.machinelearning.org/proceedings/icml2007/papers/387.pdf)

David Silver has been working on the Monte-Carlo Tree search for GO for more than 10 years!

  - Online: Monte-Carlo Simulation
  - Offline knowledge: offline value function, trained from TD(lamda), offline policy

- MoGO = MCTS + RAVE + Simple value function + rollout with custom policy
- Then there is AlphaGo, deep RL to train policy network and value network, then use the value network to reduce the depth and breadth in MCTS.
- QA: the interpretability of deep neural networks asked by Andrew from Umass.

- **Input switched affine network: designed for interpretability (http://proceedings.mlr.press/v70/foerster17a/foerster17a.pdf)** :
  - From Google Brain.
  - use a very simple linear system to model the state transition, no nonlinearity is included, the transition matrix is changed based on the input char.

- **Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders** (https://arxiv.org/pdf/1704.01279.pdf) :
  - From Google DeepMind
  - very impressive results on wave synthesis, they also provide a new dataset called Nsynth Dataset.
  - Nsynth Dataset https://magenta.tensorflow.org/nsynth

- **A closer look at memorization in deep neural network** (https://arxiv.org/pdf/1706.05394.pdf)
  - What is memorization.
  - Another relevant paper (Dziagite and Roy 2017): Computing nonvacuous generalization bounds for deep (stochastic) neural networks with many more parameters than training data" Dziugaite, Gintaire and Roy. Arxiv 2017
  - It mentioned another relevant paper: unsupervised learning by predicting noise.
  - DNN learns simple patterns,
  - It designed an experiment, to compare the properties of the networks learned on real data and random shuffled labeled data.

- **Cognitive psychology for deep neural networks: a shape bias case study (https://arxiv.org/pdf/1706.08606.pdf)**

- DeepMind’s project page: Interpreting Deep Neural networks using Psychology: https://deepmind.com/blog/cognitive-psychology/
- Interesting paper from Google DeepMind
- Several psychology experiments are evaluated on googlenet with different initial training seeds.
- They claim that though the final accuracy is the same, but the variation among different models shows the initialization and stopping time matter to the model training.
- They show that the deep CNN also get human-like bias (shape bia).

- **Understanding and visualizing the MLP: a speech case study.**
  - Trained on WSJ corpse ~ 80 hours of speech.
  - A relevant paper S.Wang et al analysis of deep network with extended data Jacobian matrix. ICML’16.
  - Some interesting modeling for the weights: consider the learned weights as linear mapping function, then visualize the SLT from input to output.
  - Using SVD to measure the diversity of linear mapping functions

- **On Calibration of modern neural networks (https://arxiv.org/pdf/1706.04599.pdf) :**
  - Very nice structured paper: pose the problem, then plot the problem, then provide a simple solution to the problem
  - Over-confidence sample = mis-calibration
  - Reduce the miscalibration: temperature scaling.
  - Kilian Q. Weinberger has done a lot of interesting works on interpreting deep neural networks.

- **Coordinated multi-agent imitation learning (https://arxiv.org/pdf/1703.03121.pdf)**
  - Imitation learning from multiply coordinating agents.

Day 3:

- Keynote given by Raia Hadsell from DeepMind: http://raiahadsell.com/index.html,
  - a lot of deep reinforcement learning tasks such as navigation,
  - Emergence of Locomotion Behaviours in Rich Environments
  - navigation in streetView environment through Taxi pickup simulation.

Application session: There are the following cool applications of Neural networks, such as graphics, physical simulation, and quantum chemistry.

- **Dance Dance Convolution** (https://arxiv.org/pdf/1703.06891.pdf)
  - http://deepx.ucsd.edu/#/home/main

- **Accelerating Eulerian Fluid Simulation With Convolutional Networks** (http://proceedings.mlr.press/v70/tompson17a/tompson17a.pdf) :
  - use the deep learning technique to speed up the rendering of graphics, it could get pretty cool rendering results.
  - http://cims.nyu.edu/~schlacht/CNNFluids.htm

- **World of Bits: An Open-Domain Platform for Web-Based Agents** (http://proceedings.mlr.press/v70/shi17a/shi17a.pdf)
  - From Stanford Percy Liang's group, Tim Shi. and Andrey Karpathy OpenAI
  - Deep reinforcement learning agent for web browsing.

- **Neural Message Passing for Quantum Chemistry (https://arxiv.org/pdf/1704.01212.pdf)**