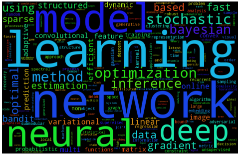
Highlights of NIPS 2016

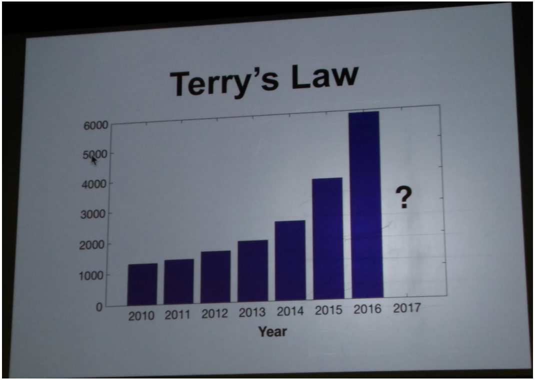
Nishanth Koganti

Research Student, Shibata Lab
Kyushu Institute of Technology

January 13, 2017



- Largest International Conference on Machine Learning.
- Held since 1987 and focuses on AI, ML, Mathematics, Statistics.

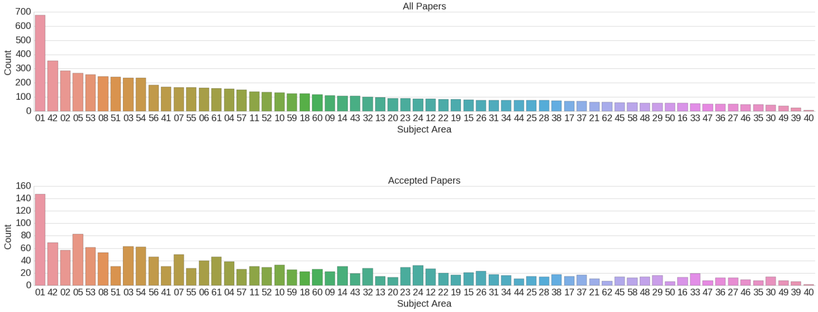


Terry's Law on NIPS Participants [1]

[1] Proposed by Terrence Sejnowski, NIPS 2016 Chair

NIPS 2016: Topics Convered

- 568 Papers (46 oral) accepted among 2400 papers with 6 reviewers per submission.
- Most popular topics were Deep Learning (1 in 4) with application to Computer Vision (1 in 10).



Long-tailed Distribution of Topics [1]

[1] NIPS 2016 Review Process

NIPS 2016: Impressions

- **Strong industry presence**: Google Deepmind, Facebook AI Research, OpenAI etc
- **Philosophical panel discussions**: Healthy debates on general trends and future of community.
- **Emphasis on Deep Learning**: Studies on applications as well as **theoretical foundations** of Deep Learning.

Tutorial on GANs[1]

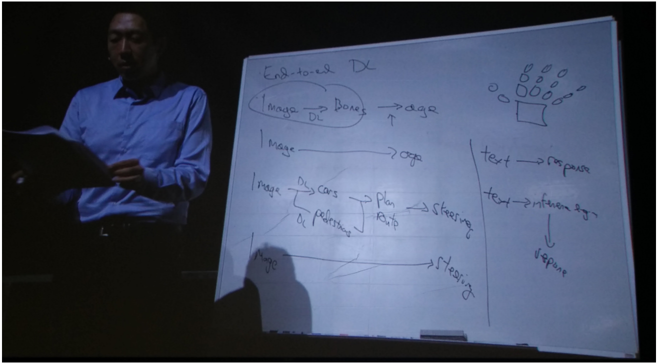
Conference Venue[1]

Boston Dynamics[1]

[1] Ian Goodfellow's Tutorial on GAN, Image of Conference Venue, Boston Dynamics Spot Demo

Nuts and Bolts of ML: Andrew Ng

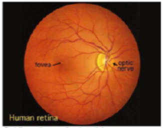
- Practical advice on machine learning research
- Importance of human-level accuracy as metric
- Techniques for collecting data to train models.



[1] https://www.youtube.com/watch?v=F1ka6a13S9I

Predictive Learning: Yann Lecun

General AI means agents should have **Common Sense**[1]



- Infer the state of the world from partial information
- Infer the future from the past and present
- Infer past events from the present state

- Filling in the visual field at the retinal blind spot
- Filling in occluded images
- Fillling in missing segments in text, missing words in speech.
- Predicting the consequences of our actions
- Predicting the sequence of actions leading to a result

- Predicting any part of the past, present or future percepts from whatever information is available.

- That's what predictive learning is
- But really, that's what many people mean by unsupervised learning

[1] https://nips.cc/Conferences/2016/Schedule?showEvent=6197

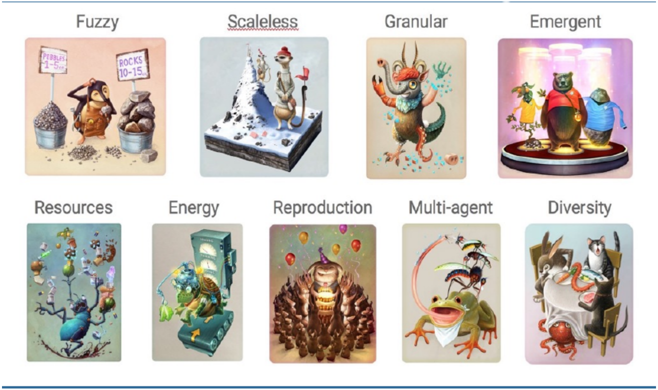
Predictive Learning: Yann Lecun

Key to General AI will be **Unsupervised Learning**[1]

[1] https://nips.cc/Conferences/2016/Schedule?showEvent=6197

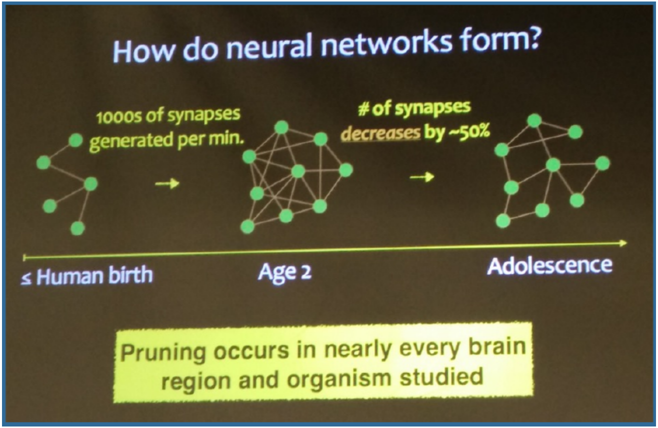
Intelligent Biosphere: Drew Purves

Intelligent life strongly depends on the environment and vice versa[1]



[1] https://nips.cc/Conferences/2016/Schedule?showEvent=6193

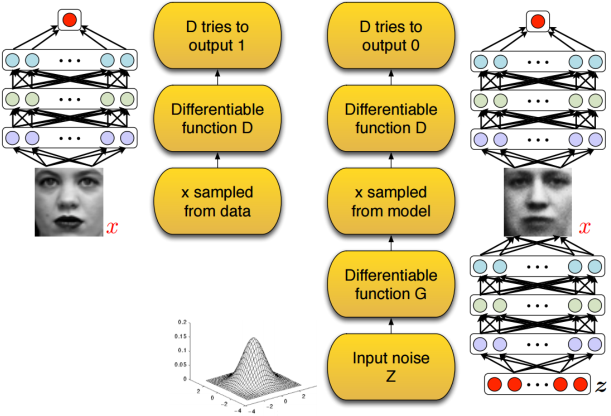
Pruning of neural connections is crucial for Learning in Brain[1]



[1] https://nips.cc/Conferences/2016/Schedule?showEvent=6192

Generative Adversarial Networks: Overview

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}(x)} \log D(x) + \mathbb{E}_{x\sim p_x(z)} \log(1 - D(G(z)))$$



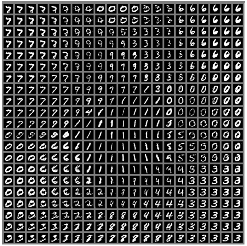
---

[1] Goodfellow et al. "Generative adversarial nets" in *NIPS 2016*

GAN: Tutorials

- Tutorial on GANs by Ian Goodfellow:
  - Applications to generate realistic images.
  - Learn latent manifolds encoding task variability.
  - Theoretical challenges for training models.
- Tutorial on Practical Tips by Soumith Chintala:
  - Tips on data preprocessing, network design and optimizers.

Manifold on MNIST[1]

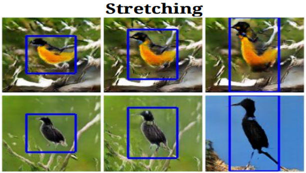


Tutorial by Chintala[1]

[1] Presentation at DLSS 2015, How to Train Your GAN

GAN: Applications

Learning What and Where[1]



- Use text input to generate novel images.
- Architecture for location and content controllable synthesis.

Information Maximizing GANs[2]

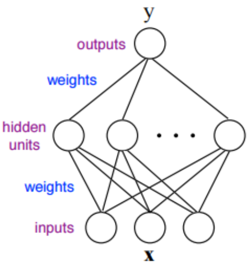
- Unsupervised learning of **disentagled** representations.
- Maximizes mutual information between latent variables and observations.

[1] Reed et al., "Learning what and where to draw" in *NIPS 2016*

[2] Chen et al., "Infogan: Interpretable representation learning by information maximizing generative adversarial nets" in *NIPS 2016*

Bayesian Deep Learning

- History of Bayesian Neural Networks by Zoubin Gharamani[1]:
  - Derives from Gaussian Process Regression
  - Feature selection possible using Automatic Relevance Determination
  - Current research on variational inference techniques



Bayesian neural network

Data: $\mathcal{D} = \{(\mathbf{x}^{(n)}, y^{(n)})\}_{n=1}^{N} = (X, \mathbf{y})$

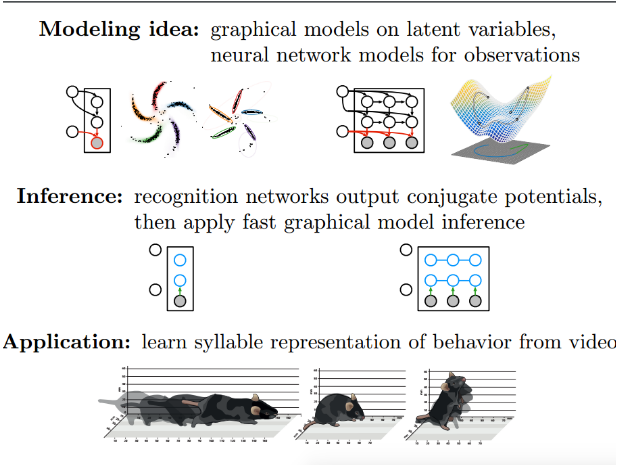
Parameters $\boldsymbol{\theta}$ are weights of neural net

prior	$p(\boldsymbol{\theta}\|\boldsymbol{\alpha})$
posterior	$p(\boldsymbol{\theta}\|\boldsymbol{\alpha},\mathcal{D}) \propto p(\mathbf{y}\|X,\boldsymbol{\theta})p(\boldsymbol{\theta}\|\boldsymbol{\alpha})$
prediction	$p(y'\|\mathcal{D},\mathbf{x}',\boldsymbol{\alpha}) = \int p(y'\|\mathbf{x}',\boldsymbol{\theta})p(\boldsymbol{\theta}\|\mathcal{D},\boldsymbol{\alpha})\, d\boldsymbol{\theta}$

[1] Keynote Talk by Zoubin Gharamani

Deep Learning and Graphical Models

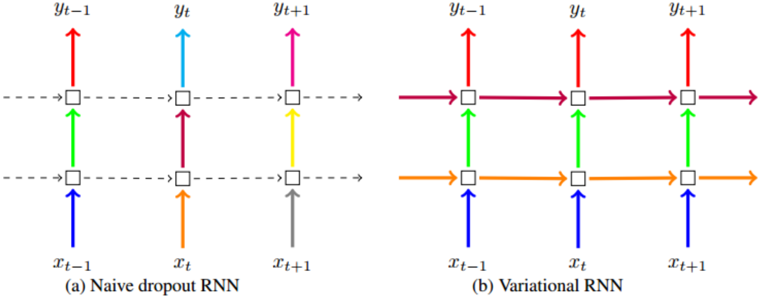
Combined training of Graphical Models and Deep Neural Networks using Variational Inference[1]



[1] Johnson et al., "Composing graphical models with neural networks for structured representations and fast inference" in *NIPS 2016*

Dropout as a Bayesian Approximation

Dropout is equivalent to variational inference in Deep GPs with specific kernel function



[1] Gal et al., "A theoretically grounded application of dropout in recurrent neural networks" in *NIPS 2016*

Deep Reinforcement Learning: Tutorials

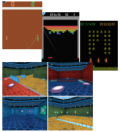
- Deep RL through Policy Optimization[1]:
  - History of policy search RL focusing on Deep RL
  - Future work on **Skill Transfer** and **Exploration**
- Nuts and Bolts of Deep RL Research[2]:
  - Tips on implementing existing algorithms
  - Problem formulation for novel tasks


Kohl and Stone, 2004


Ng et al, 2004

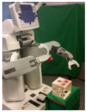

Tedrake et al, 2005

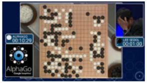

Kober and Peters, 2009


Mnih et al, 2015 (A3C)


Silver et al, 2014 (DPG)
Lillicrap et al, 2015 (DDPG)

Schulman et al, 2016 (TRPO + GAE)


Levine*, Finn*, et al, 2016 (GPS)



Silver*, Huang*, et al, 2016 (AlphaGo**)

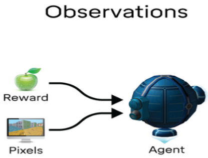
[1] Tutorial by Pieter Abbeel and John Schulman

[2] Tutorial by John Schulman in Deep RL Workshop

- OpenAI Universe[1]: Develop and implement RL algorithms on many environments.
- Google Deepmind[2]: Simulator of 3D environments such as **Quake** and puzzle tasks.
- Facebook TorchCraft[3]: Bridge between Torch Library and Starcraft Game.

OpenAI Universe

Google Deepmind



Facebook Starcraft

[1] OpenAI Universe Press Release
[2] Google Deepmind Blog
[3] TorchCraft Repository on Github

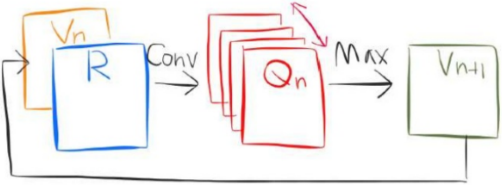
Represent Bellman equations with Convolutional Neural Networks to learn Value function

Max-Pooling

$$V_{n+1}(s) = \boxed{\max_{a}} Q_n(s,a)$$

Convolution

$$Q_n(s,a) = R(s,a) + \boxed{\gamma \sum_{s'} P(s'|s,a)} V_n(s')$$



[1]Tamar et al., "Value Iteration Networks" in *NIPS 2016*

Recurrent Neural Networks: Overview

- Papers on theoretical foundations (**better architectures**[1]) and applications (**Reinforcement Learning**[2]).
- Strong correspondence to Neuroscience research (Fast and Slow time scales[3]).

[1] http://papers.nips.cc/paper/6114-weight-normalization-a-simple-reparameterization-to-accelerate-training-of-deep-neural-networks

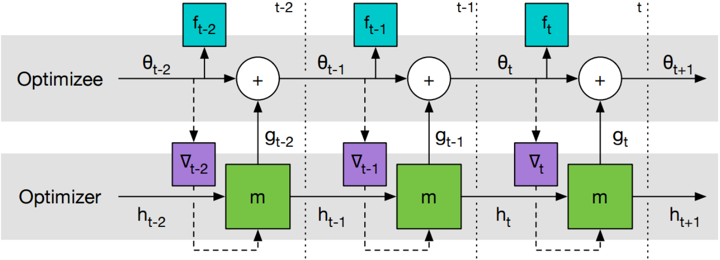
[2] https://arxiv.org/pdf/1611.02779.pdf

[3] http://papers.nips.cc/paper/6310-phased-lstm-accelerating-recurrent-network-training-for-long-or-event-based-sequences

Learning to (learn by gradient descent) by gradient descent

- Current optimizers are parametric and domain specific.
- Attempt to learn the optimizer using RNN.

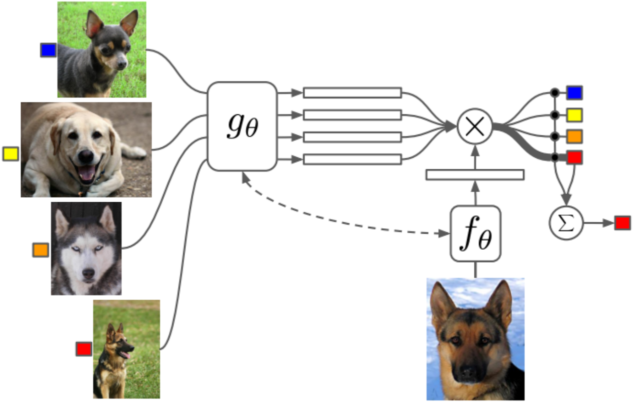
$$\theta_{t+1} := \theta_t + g(\Delta f(\theta_t), \phi) \tag{1}$$



[1] Andrychowicz et al., "Learning to learn by gradient descent by gradient descent" in *NIPS 2016*

One-shot Learning

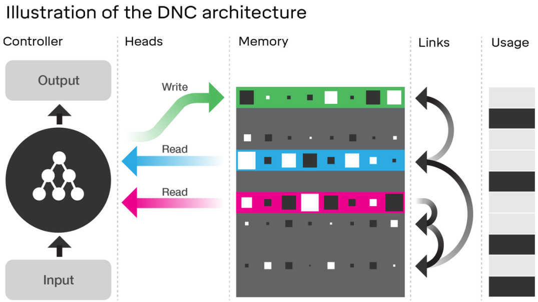
- Novel architectures proposed for one-shot learning.
- The idea is a differentiable nearest neighbor classifier.



[1] Vinyals et al., "Matching networks for one shot learning" in *NIPS 2016*

Differentiable Neural Computers

- Neural Network with memory access (Read, Write, Link).
- Used for skill transfer, reinforcement learning, reasoning etc.

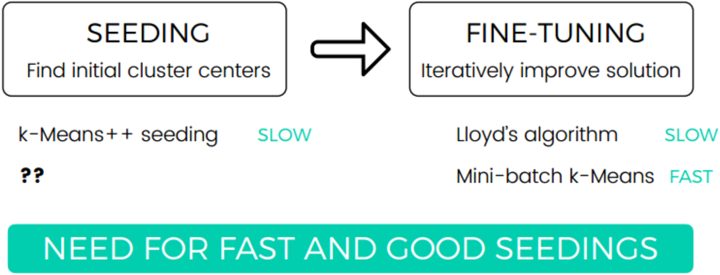


[1] https://deepmind.com/blog/differentiable-neural-computers/

Boston Robotics Demo

[1] https://nips.cc/Conferences/2016/Schedule?showEvent=6194

- Matrix completion: Non-convex objective for matrix completion has no spurious local minima [1]
- Clustering: Seedings for large-scale k-means++ clustering [2]

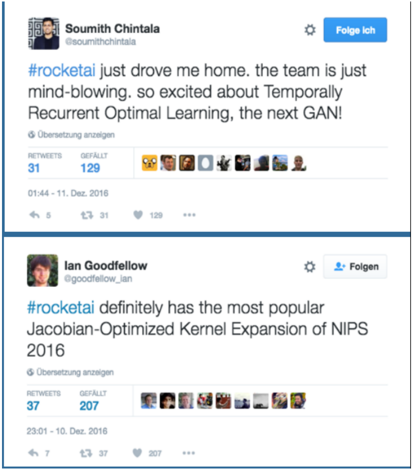


[1] http://papers.nips.cc/paper/6048-matrix-completion-has-no-spurious-local-minimum

[2] http://papers.nips.cc/paper/6478-fast-and-provably-good-seedings-for-k-means

RocketAI: The AI Bubble

Prank by NIPS organizers to warn about the AI Bubble



[1] https://medium.com/the-mission/rocket-ai-2016s-most-notorious-ai-launch-and-the-problem-with-ai-hype-d7908013f8c9

Sightseeing in Barcelona

Thank you!

Take home Messages

- Deep Learning is no longer just a tool!
- Emphasis is changing to General AI and Life long learning
- Recurrent Neural Networks are very important!