Principles of neural network design

Francois Belletti, CS294 RISE
Human brains as metaphors of statistical models

**Biological analogies**

The visual cortex of mammals

Multiple sensing channels

Memory and attention

**Machine learning instantiations**

Deep convolutional neural networks

Multimodal neural networks

LSTMs and GRUs
Neural Networks For Computer Vision
Neural Networks in Computer Vision

Neural networks for classification of handwritten digits

Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.
Learning Mechanism: Correction of Mistakes

Nature used a single tool to get to today’s success: mistake

\[ W_k = W_{k-1} - \epsilon \frac{\partial E^p_k(W)}{\partial W} \]

1. receive new observation \( x = [x_1, \ldots x_d] \) and target \( y^* \)
2. **feed forward**: for each unit \( g_j \) in each layer 1…L
   compute \( g_j \) based on units \( f_k \) from previous layer:
   \[ g_j = \sigma \left( u_{j0} + \sum_k u_{jk} f_k \right) \]
3. get prediction \( y \) and error \( (y-y^*) \)
4. **back-propagate error**: for each unit \( g_j \) in each layer L…1

- **(a) compute error on \( g_j \)**
  \[ \frac{\partial E}{\partial g_j} = \sum_i \sigma'(h_i) v_{ij} \frac{\partial E}{\partial h_i} \]
  - should \( g_j \) be higher or lower?
  - how \( h_i \) will change as \( g_j \) changes
  - was \( h_i \) too high or too low?

- **(b) for each \( u_{jk} \) that affects \( g_j \)**
  \[ \frac{\partial E}{\partial u_{jk}} = \frac{\partial E}{\partial g_j} \sigma'(g_j) f_k \]
  - do we want \( g_j \) to be higher/lower?
  - how \( g_j \) will change if \( u_{jk} \) is higher/lower
  - update the weight
  \[ u_{jk} = u_{jk} - \eta \frac{\partial E}{\partial u_{jk}} \]

Back-propagation is a recursive algorithm

This module provides `function()`, commonly accessed as `theano.function`, the interface for compiling graphs into callable objects.

You’ve already seen example usage in the basic tutorial... something like this:

```python
>>> x = theano.tensor.dscalar()
>>> f = theano.function([x], 2*x)
>>> print f(4)
# prints 8.0
```
Image Classification
Successful Architecture In Computer Vision

An example of a wide network: AlexNet
Understanding What Happens Within A Deep NN

Examining convolution filter banks

Examining activations
Determining A Neuron’s Speciality

Images that triggered the highest activations of a neuron:
Another Successful Architecture For CV

Inception:
ILSVRC top-5 error on ImageNet
Recurrent Architectures
Learning To Leverage Context

Memory in Recurrent Architectures: LSTM (Long Short Term Memory Network)

Input $x$, output $y$, context $c$ (memory)
Other recurrent architectures

Gated recurrent units:

\[
\begin{align*}
    z_t &= \sigma (W_z \cdot [h_{t-1}, x_t]) \\
    r_t &= \sigma (W_r \cdot [h_{t-1}, x_t]) \\
    \tilde{h}_t &= \tanh (W \cdot [r_t \times h_{t-1}, x_t]) \\
    h_t &= (1 - z_t) \times h_{t-1} + z_t \times \tilde{h}_t
\end{align*}
\]
Why Is Context Important?

In language, most grammars are not context free

End-to-end translation, Alex Graves
Context Is Also Important In Control

Remembering what just happened is important for decision making
Memory is necessary for localization

Latest experiment in asynchronous deep RL: LSTMS for maze running

Memory comes at a cost: a lot of RAM or VRAM is necessary
Conclusion: the distributed brain
Interaction is crucial in enabling AI

AlphaGo Overview

Policy Network
- Expert Games
  - 130 000 Games
  - 30 M Positions
- Supervised Learning
  - SL Policy
  - Position $\rightarrow$ Next Move
  - Accuracy: 56%

Fast Policy Network
- Expert Games
  - 140 000 Patterns
  - 130 000 Games
  - 30 M Positions
- Supervised Learning
  - Fast Policy
  - Pattern $\rightarrow$ Next Move
  - Accuracy: 24%

Reinforcement Learning Policy Network
- Self-Play Games
  - 1.3 M Games by various versions of RL Policy
- Reinforcement Learning
  - RL Policy
  - Position $\rightarrow$ Next Move
  - Wins 89% vs. SL Policy

Value Network
- Self-Play Games
  - 30 M Positions by fixed version of RL Policy
- Reinforcement Learning
  - Position $\rightarrow$ Win Probability
  - 15 000 times faster than MCTS Rollouts evaluations

AlphaGo

Monte Carlo Tree Search
Playing versus computers before beating humans
Bootstrapping by interaction

Why would two androids casually chat one with another?
The distributed brain at the edge

Distributed RL is reminiscent of the philosophical omega point of knowledge

We are not human beings having a spiritual experience; we are spiritual beings having a human experience.

(Pierre Teilhard de Chardin)
Multiple Input Neural Networks
Multi Inputs For Inference

Youtube Video Auto-encoding

(a) Video-Only Deep Autoencoder

(b) Bimodal Deep Autoencoder
Multiple Input Control

Multiplexing Inputs
Multiplexing In The Human Brain

Frontal lobe
Executive functions, thinking, planning, organising and problem solving, emotions and behavioural control, personality

Motor cortex
Movement

Sensory cortex
Sensations

Parietal lobe
Perception, making sense of the world, arithmetic, spelling

Occipital lobe
Vision

Temporal lobe
Memory, understanding, language