Brief Introduction to Deep Learning Models & The Applications

Baoxun Wang

Before Starting

- 诚惶诚恐
  - Mathematics>Statistics>Machine Learning
  - There are NO Equations in the slides of this talk
  - But there are quite a number of figures

Agenda

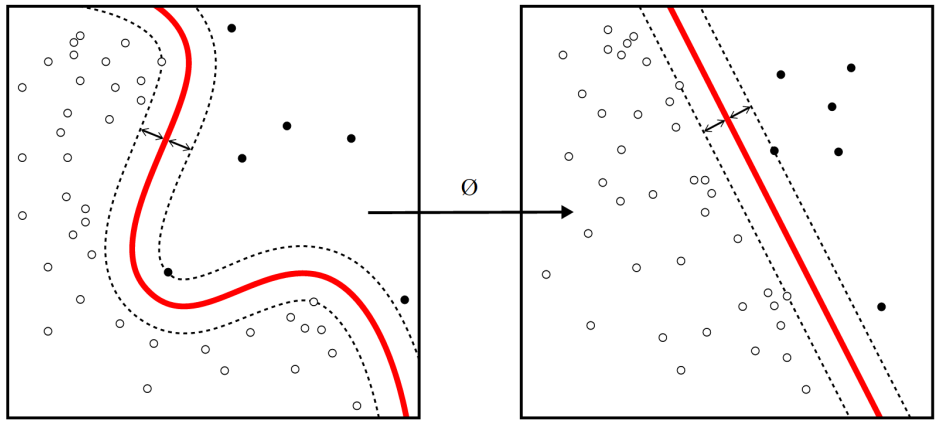
- Overview of Deep Learning
- Major DL frameworks & the essence
- Outstanding DL projects
- DL for Automatic Chatbots
- Remaining challenge & future
- Conclusions

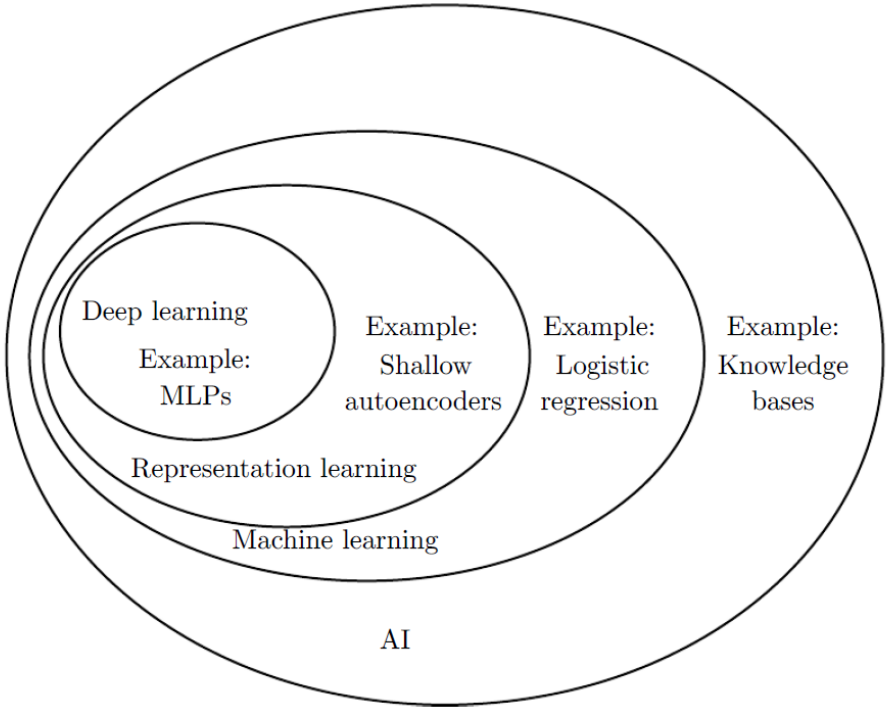
Agenda

- Overview of Deep Learning
  - Before DL …
  - Informal Definition & spirits
- Major DL frameworks & the essence
- Outstanding DL projects
- DL for Automatic Chatbots
- Remaining challenge & future
- Conclusions

Machine Learning

- Machine Learning
  - Supervised learning
    - Classification
    - Regression
  - Unsupervised learning
    - Clustering
  - Dimensionality Reduction
  - …



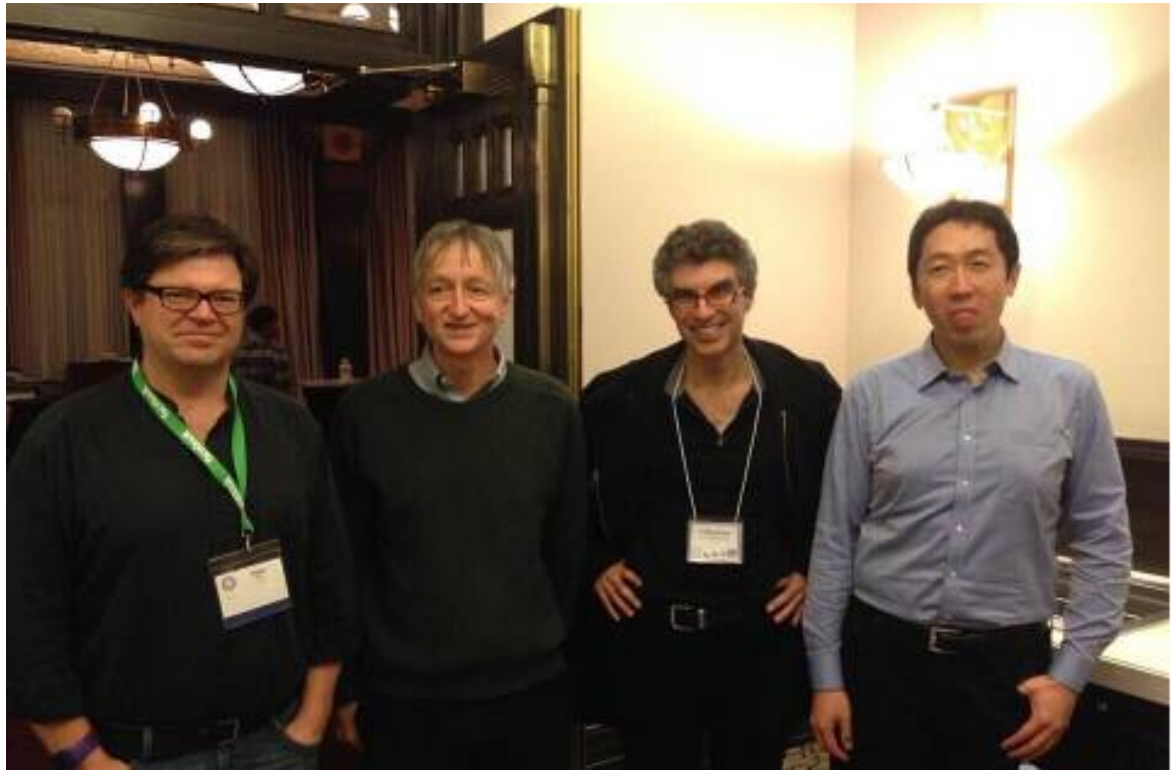


Before Deep Learning

- Before deep learning, the machine learning approaches are
  - Several sub-models coupled together
  - Large amount of work on feature engineering
  - Not specially designed for big data and unlabeled data

The Popular Deep Learning

- From left to right: Yann LeCun, G. Hinton, Yoshua Bengio, Andrew NG

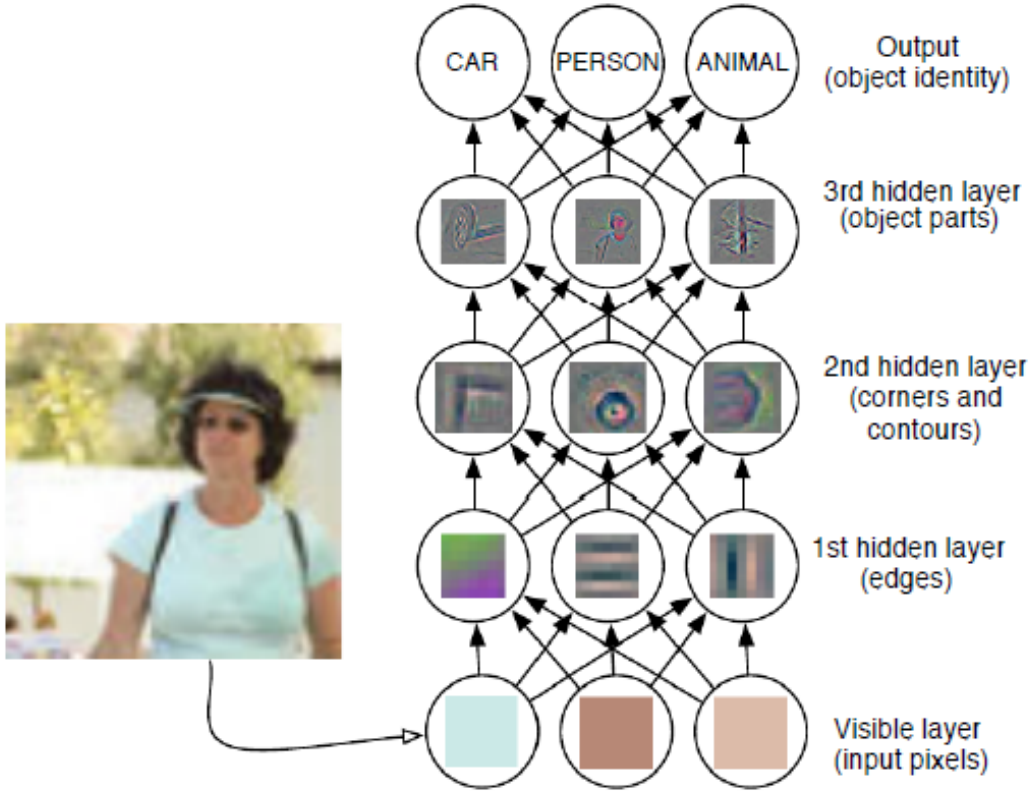


What is DL?

- Deep learning is a branch of machine learning algorithms that
  - using a deep graph with multiple processing layers
  - composed of multiple linear and non-linear transformations
  - attempt to model high level abstractions in data

The Spirit of DL

- End-to-End Learning
  - Instead of step-by-step learning
- Learning Representations
  - Free from feature Engineering
- Pre-training & Fine-tuning
  - Make full use of big data



From Starting to Boom

- Backpropagation (1960s-1970s)
- Restricted Boltzmann Machine (1986) → Deep Belief Networks for Image Recognition (2006)
- Convolutional Neural Networks (LeNet, 1998) → ImageNet with CNN (2012)
- Recurrent Neural Networks (LSTM, 1997) → RNN Language Model (2013)

Why now?

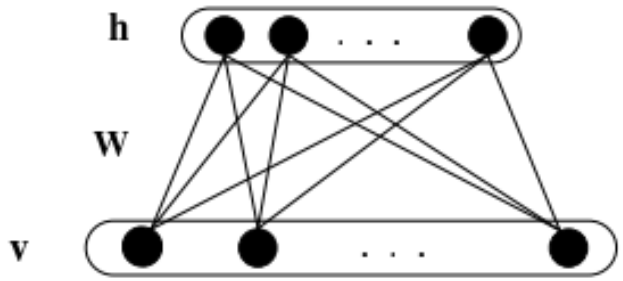
- Large amounts of data accumulated in Web
- Faster computing unit
  - GPU, FPGA, Special chips
- Development of Deep Learning models

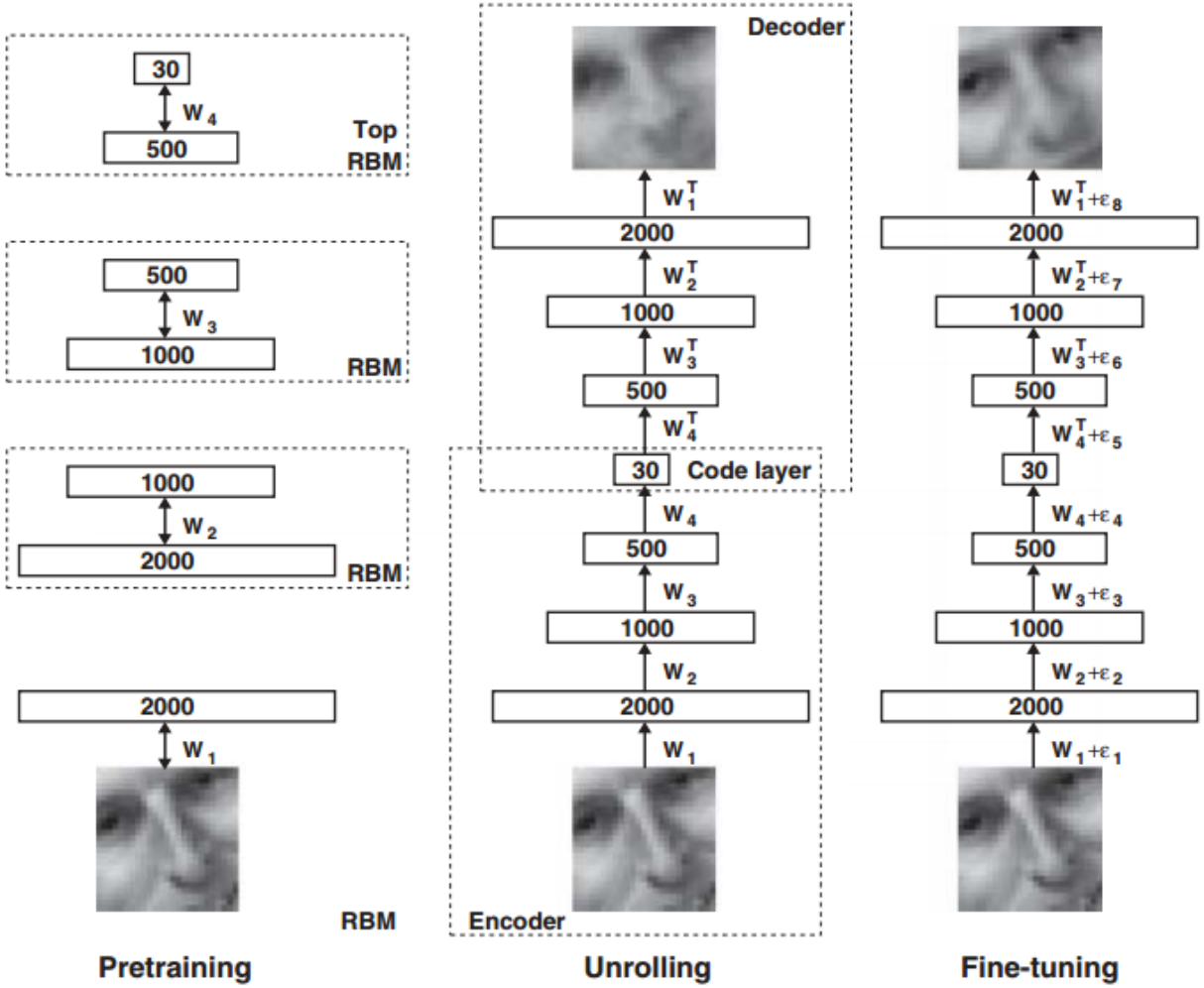
- AND…
  - Arxiv.org

Agenda

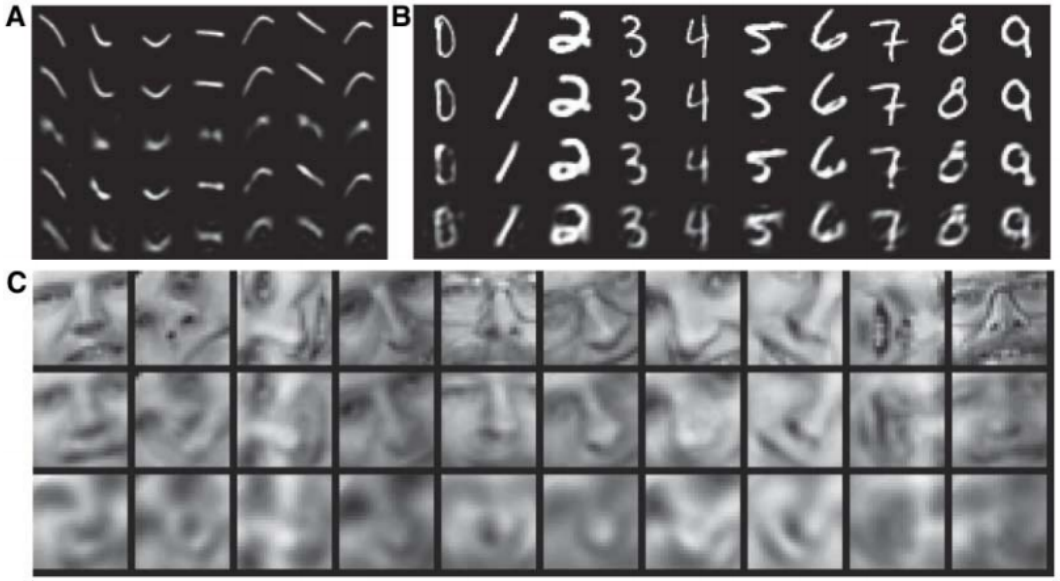
- Overview of Deep Learning
- Major DL frameworks & the essence
  - Deep Belief Network
  - Convolutional Neural Network
  - Recurrent Neural Network
- Outstanding DL projects
- DL for Automatic Chatbots
- Remaining challenge & future
- Conclusions

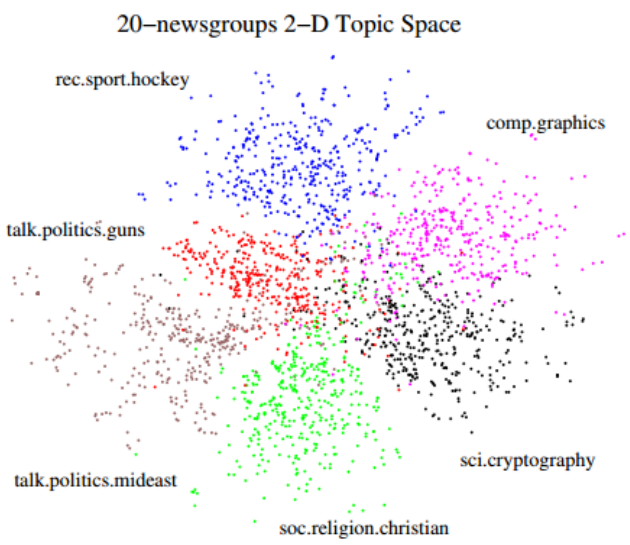
The very beginning – RBM & DBN

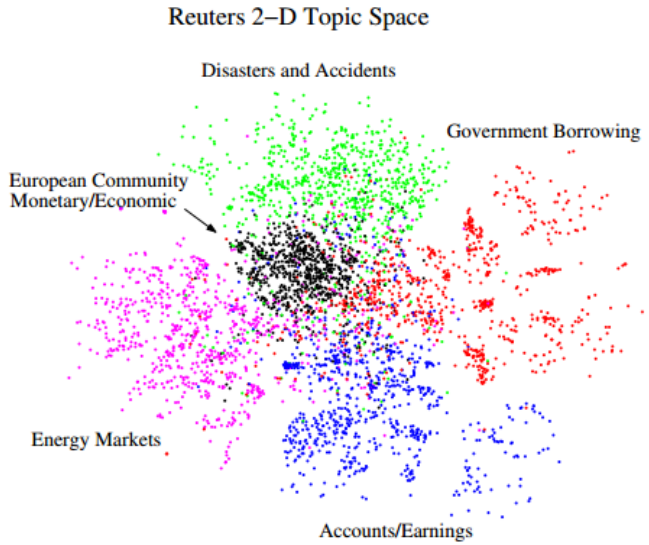




Applications of DBN

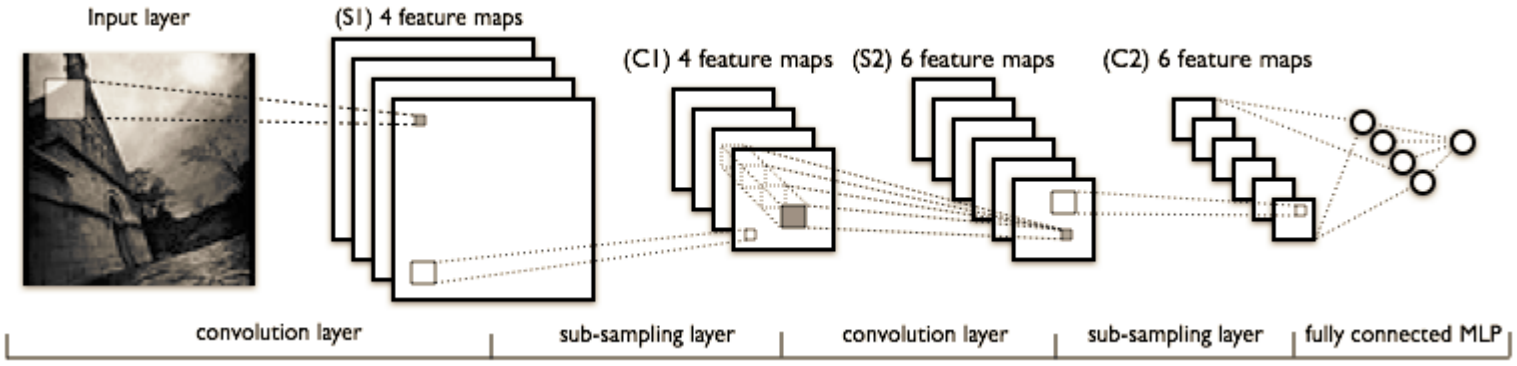






Classic Structure of CNN

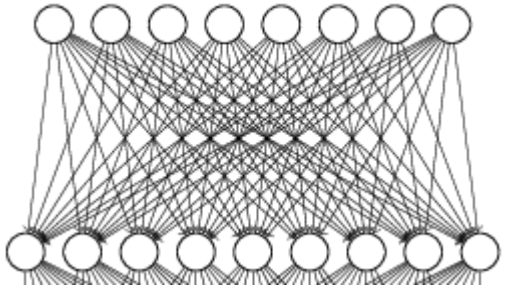
- CNN has changed the normal form of Computer Vision
- Makes CV much easier
- Simulation of animal visual cortex

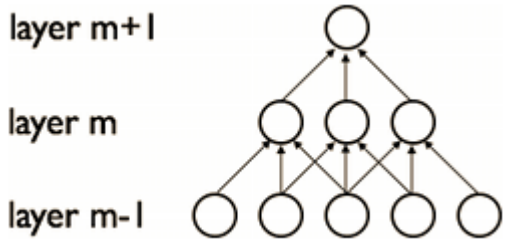


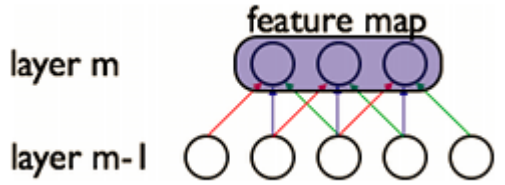
LeNet

Essence of CNN

- Sparse Connectivity
- Shared Weights

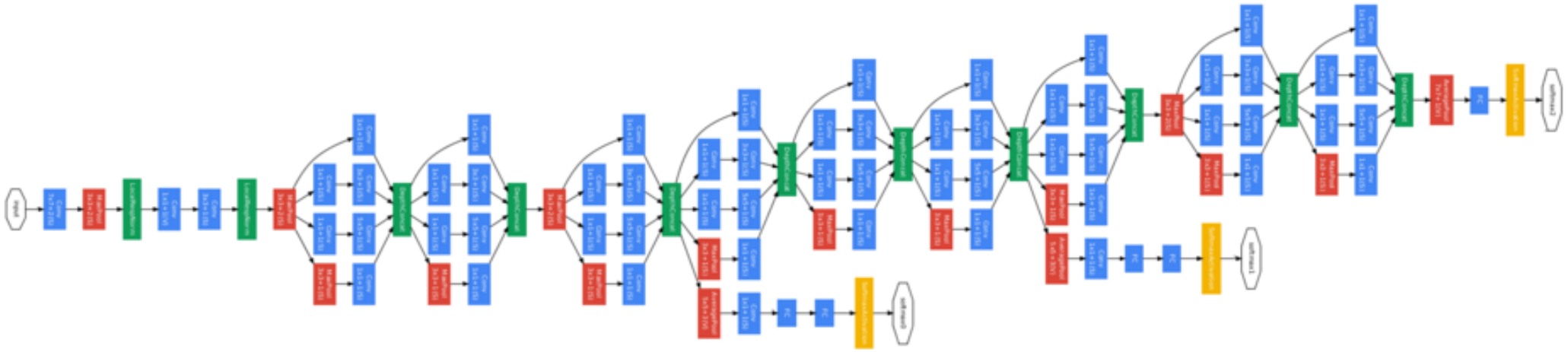






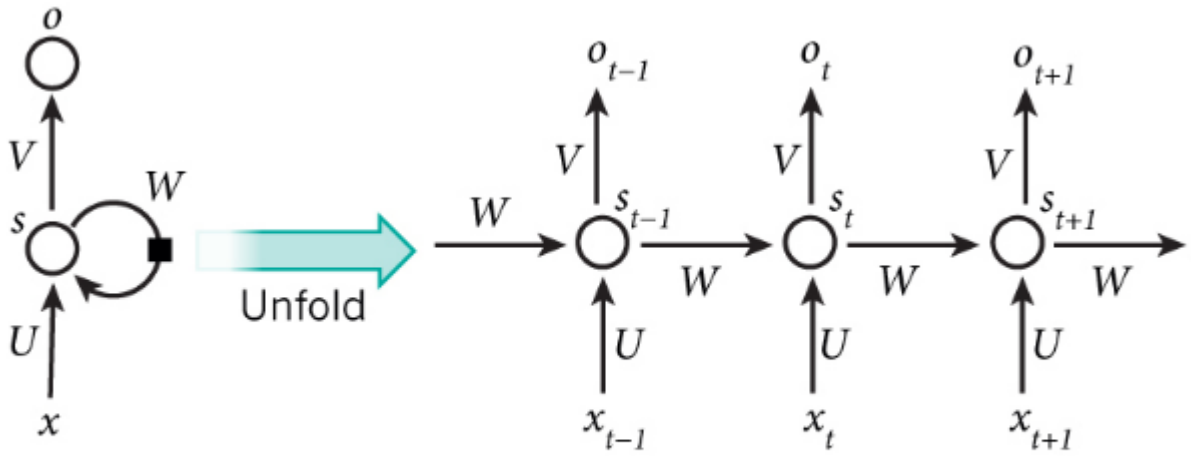
A very Deep CNN Architecture

- GoogLeNet
  - Champion of ImageNet 2014
  - Error rate=6.67%
  - 22 layers
  - Models today are even deeper!

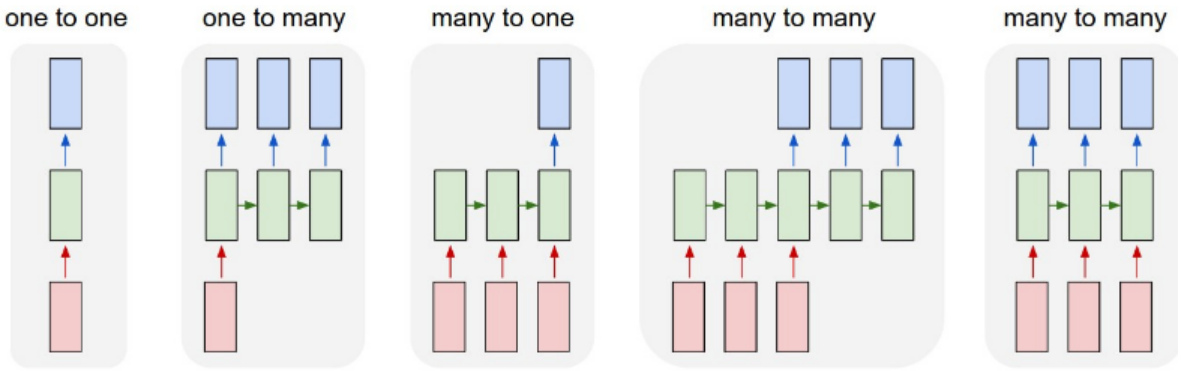


Basic Recurrent Neural Network

- Naturally suitable for sequence learning
- A very naïve simulation of “memory”
- Game changer of Speech Recognition and Machine Translation



A recurrent neural network and the unfolding in time of the computation involved in its forward computation. Source: Nature



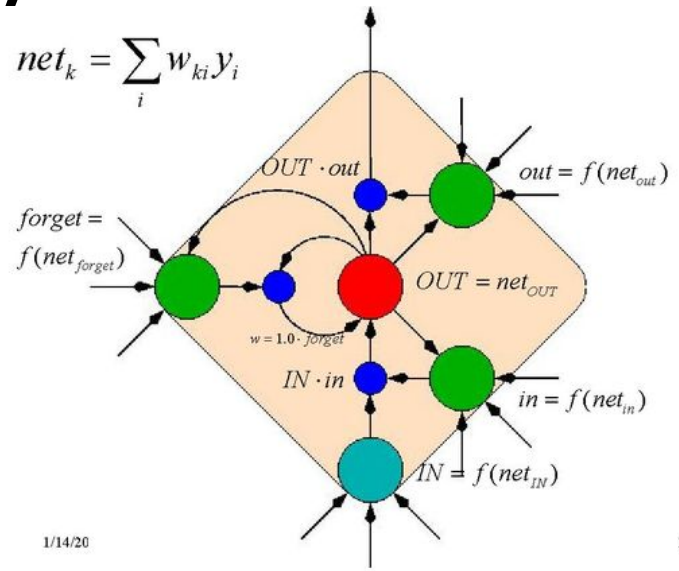
Long-Short Term Memory

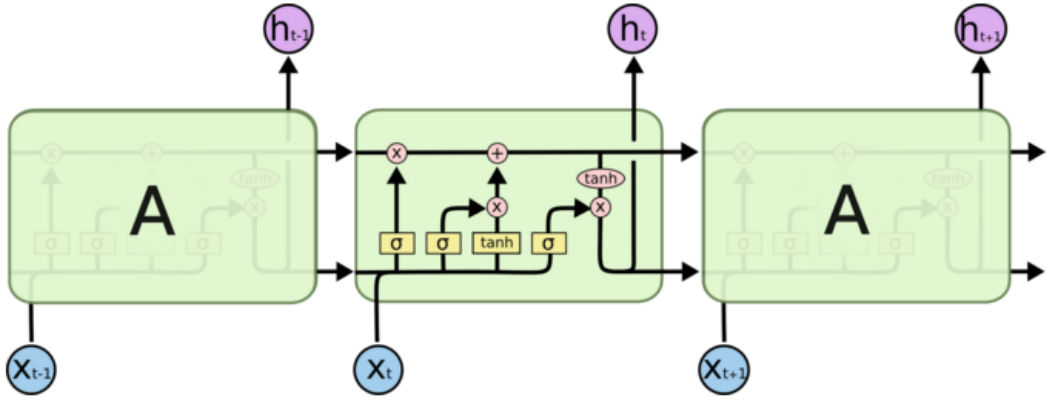
Allow each time step to modify

- Input gate (current cell matters) $i_t = \sigma\left(W^{(i)}x_t + U^{(i)}h_{t-1}\right)$
- Forget (gate 0, forget past) $f_t = \sigma\left(W^{(f)}x_t + U^{(f)}h_{t-1}\right)$
- Output (how much cell is exposed) $o_t = \sigma\left(W^{(o)}x_t + U^{(o)}h_{t-1}\right)$
- New memory cell $\tilde{c}_t = \tanh\left(W^{(c)}x_t + U^{(c)}h_{t-1}\right)$

Final memory cell: $c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t$

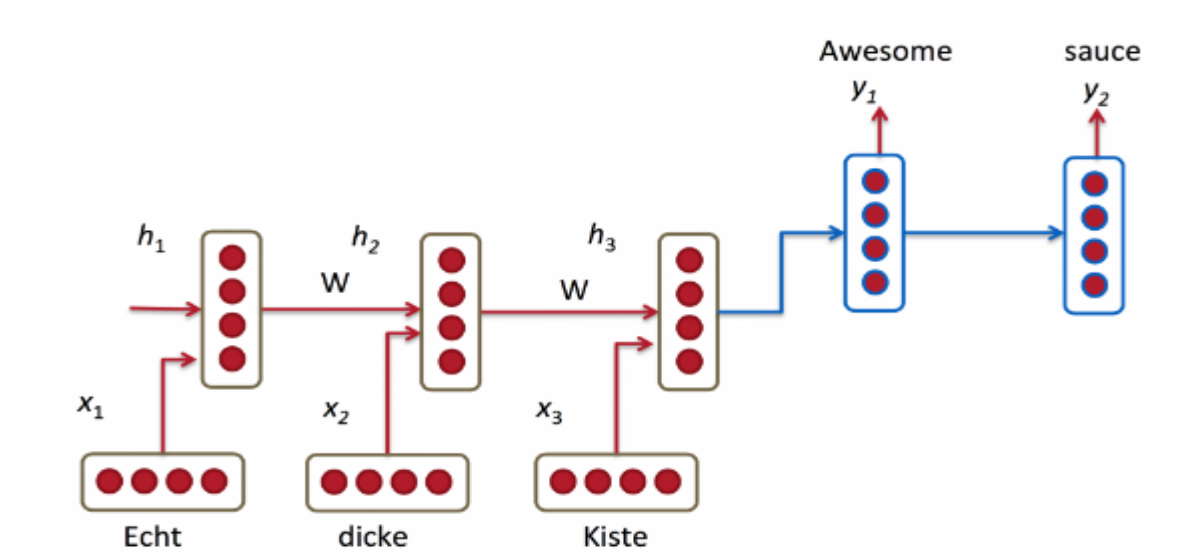
Final hidden state: $h_t = o_t \circ \tanh(c_t)$





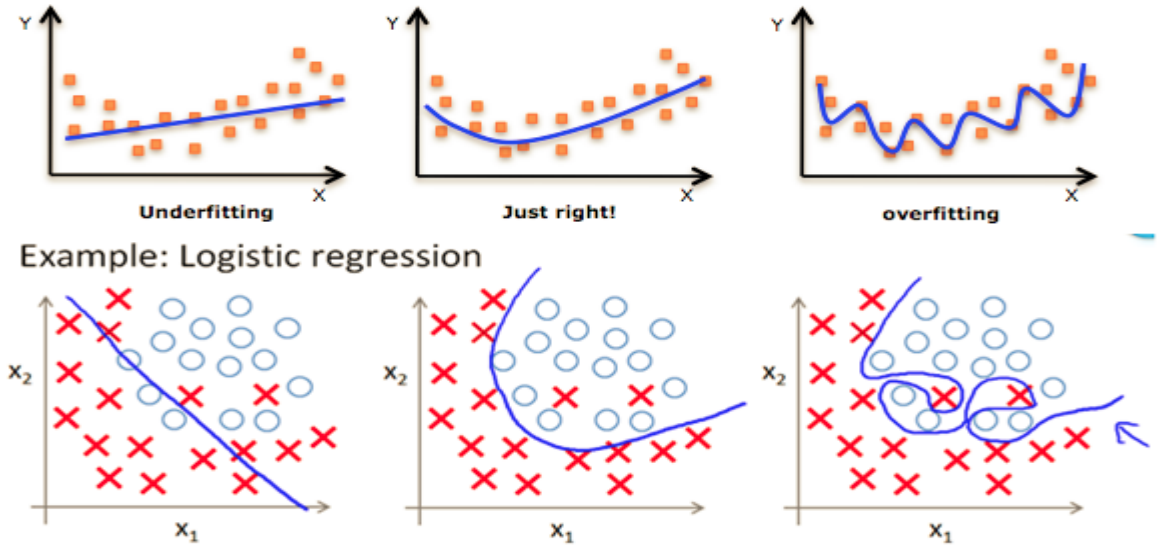
The Magic S2S

- Comes from Machine Translation, begins the NMT time
- Composed of an Encoder & a Decoder
- Perfectly matching the basic idea of MT
- Is believed to be the key to Automatic Chatbots

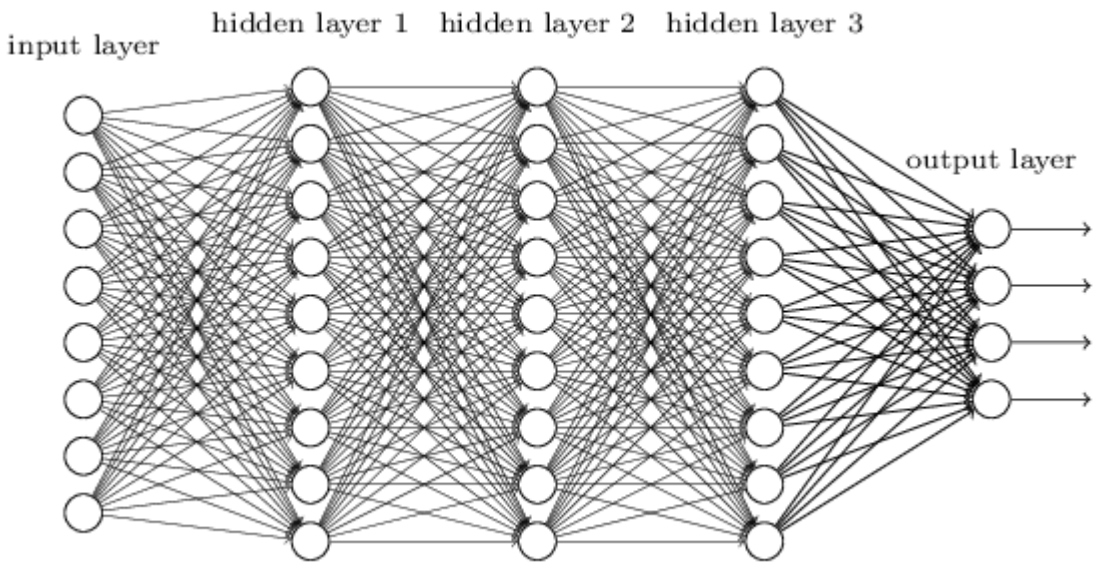


Another way to consider DL

- What are Machine Learning researchers fighting against?
  - Overfitting
- Overfitting is mainly caused by
  - Amount of parameters in models



To reduce number of parameters



- DBN
  - Layer-by-layer training
- CNN
  - Locally connection & sharing Weights
- RNN
  - Reusing parameters

Agenda

- Overview of Deep Learning
- Major DL frameworks & the essence
- Outstanding DL projects
  - CNN for ImageNet
  - Image caption generation
  - AlphaGo
  - Google’s Neural Machine Translation System
- DL for Automatic Chatbots
- Remaining challenge & future
- Conclusions

Outstanding DL Projects

- DL changes the normal forms of some machine learning problems
  - CV, Machine Translation
- DL provides solutions to some difficult tasks
  - Go, Chatbots
- DL makes everyone seem to know AI well…

Large Scale Visual Recognition Challenge 2016

- ImageNet

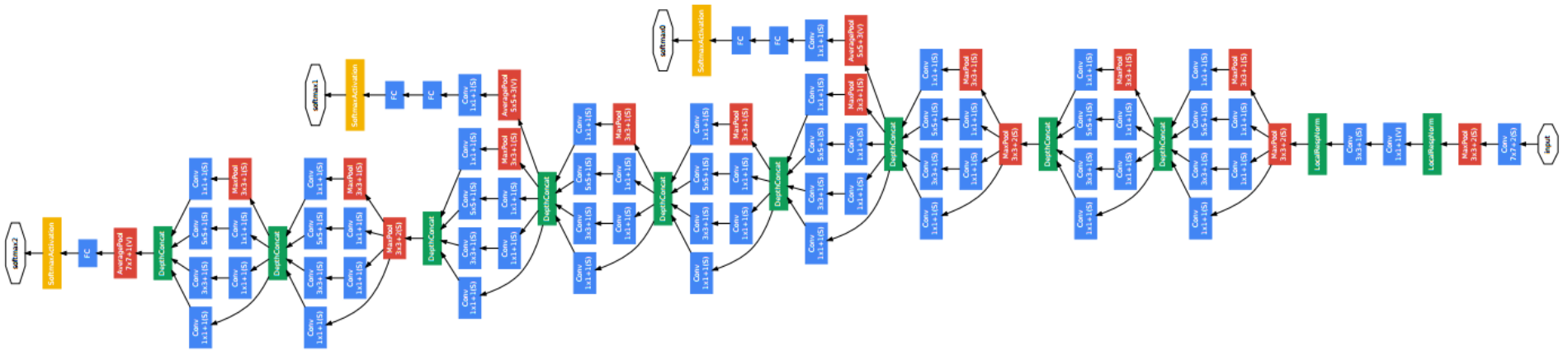
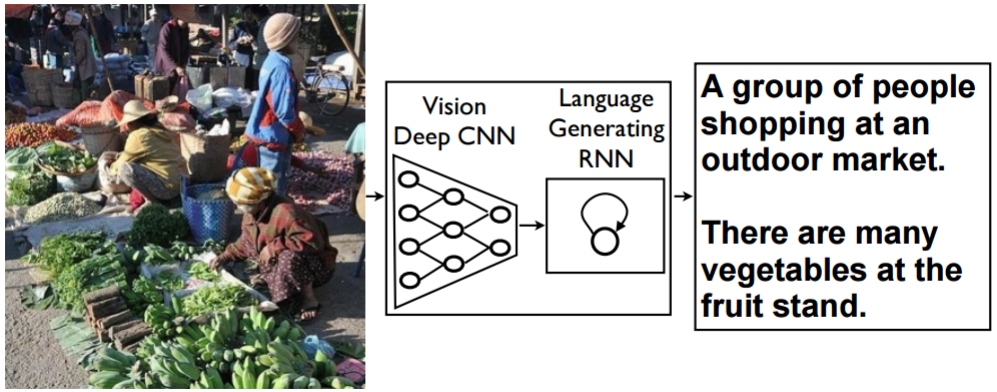
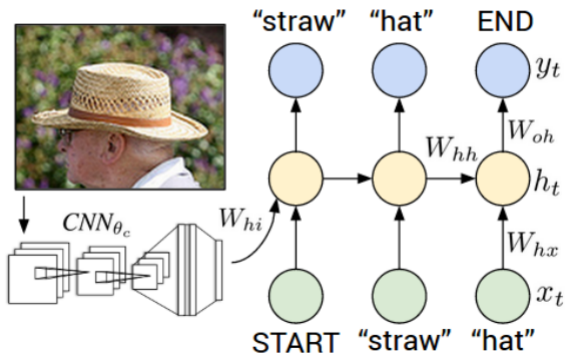


Image Caption Generation





man in black shirt is playing guitar.

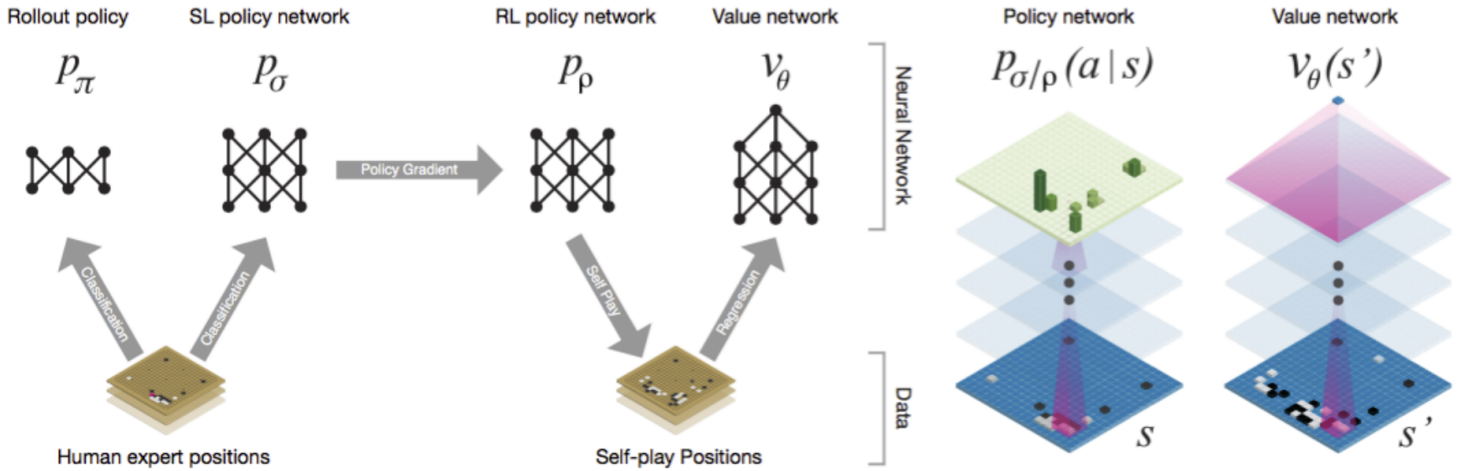
construction worker in orange safety vest is working on road.

two young girls are playing with lego toy.

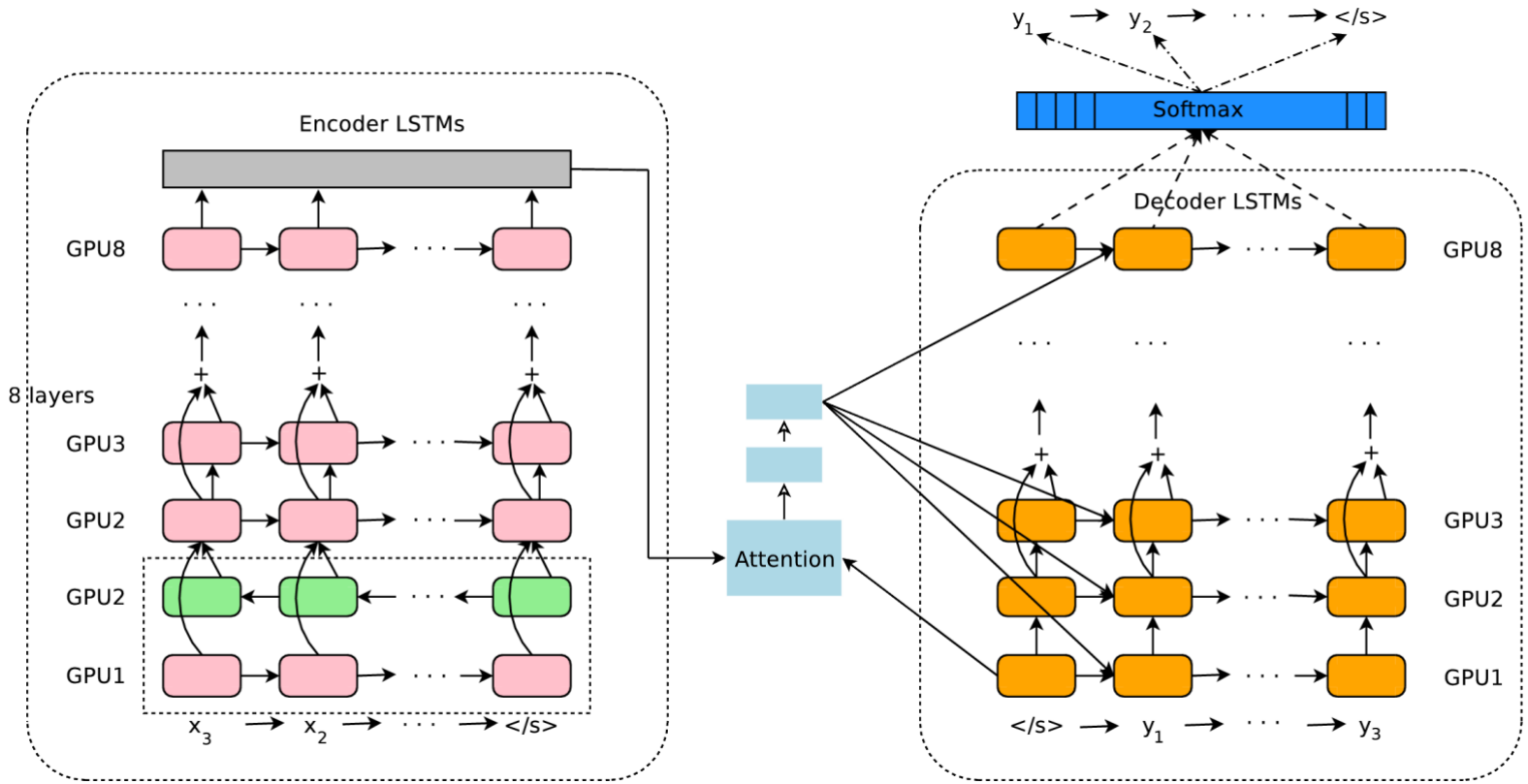
boy is doing backflip on wakeboard.

AlphaGo


AlphaGo


Rollout policy
SL policy network
RL policy network
Value network
p_π
p_σ
p_ρ
v_θ
Policy Gradient
Classification
Classification
Self Play
Regression
Neural Network
Data
Human expert positions
Self-play Positions
Policy network
Value network
$p_{\sigma/\rho}(a \mid s)$
$v_\theta(s')$
s
s'

Google's Neural Machine Translation Sys.



Agenda

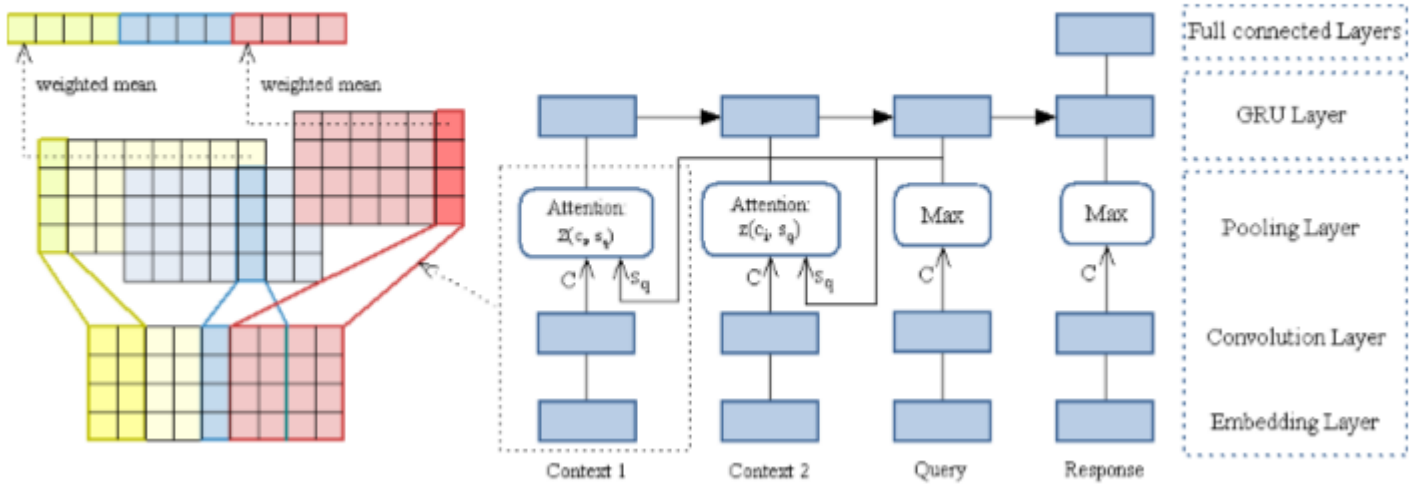
- Overview of Deep Learning
- Major DL frameworks & the essence
- Outstanding DL projects
- DL for Automatic Chatbots
- Remaining challenge & future
- Conclusions

What we have done

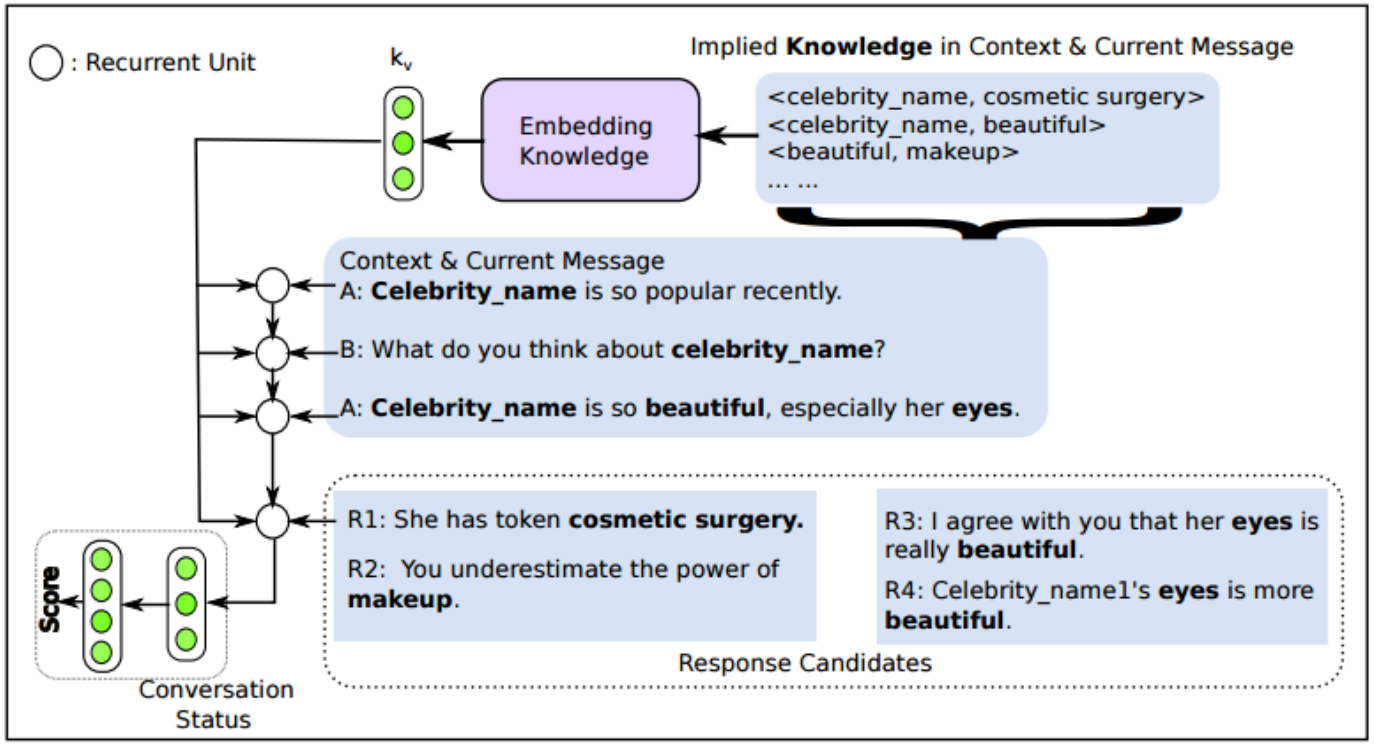
- We are concentrating on the deep learning models for NLP tasks
  - Short-text semantic relevance quantification
  - Context-aware conversation modeling
  - Task oriented dialog system
  - Automatic reading comprehension
  - …

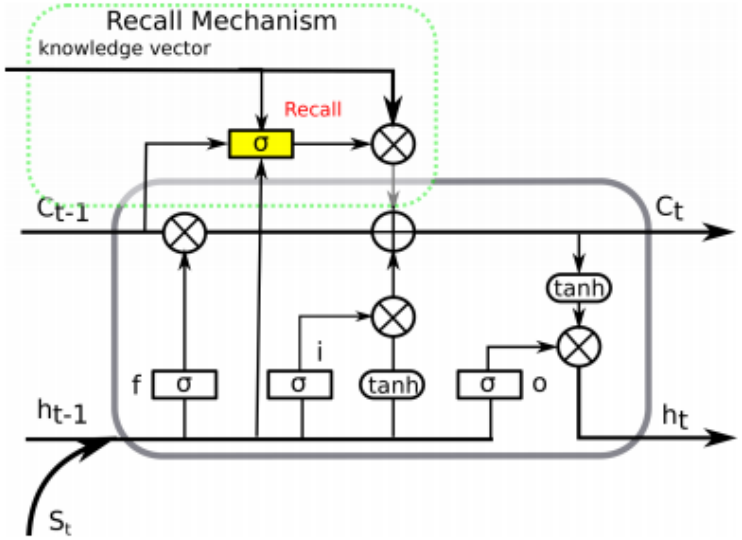
Context-aware Response Selection

- Multi-column convolution kernel + attention GRU



Incorporating Domain Knowledge into Context-aware response selection





Agenda

- Overview of Deep Learning
- Major DL frameworks & the essence
- Outstanding DL projects
- Remaining challenge & future
- Conclusions

Remaining Challenges

- Current Deep Learning models are not good at
  - Learning on small datasets
  - Incorporating knowledge/priors
  - Unsupervised learning
  - Understanding natural language

What’s after DL?

- Attention mechanism
- Deep Reinforcement Learning
- Small dataset based learning
- Unsupervised Learning
  - Generative Adversarial Networks
- Neural Turing Machine
- Learning the structure of network
- …

Agenda

- Overview of Deep Learning
- Major DL frameworks & the essence
- Outstanding DL projects
- Remaining challenge & future
- Conclusions

Conclusions

- Deep Learning is the inevitable result of machine learning development
- Deep Learning does bring some changes to machine learning
- Deep Learning is NOT the simulation of human brains
- CV & SR have benefited from DL significantly, but NLP has not
- The situation changes very quickly

References

- Conferences
  - Machine Learning: NIPS, ICML, IJCAI, AAAI
  - CV & Speech: CVPR, ICCV, InterSpeech
  - NLP: ACL, EMNLP, Coling
- Papers
  - Alex et al., ImageNet Classification with Deep Convolutional Neural Networks, NIPS 2012
  - G. Hinton, Reducing the Dimensionality of Data with Neural Networks, Science, 2006
  - Hochreiter., LONG SHORT-TERM MEMORY, Neural Computation, 1997
  - Zachary et al., A Critical Review of Recurrent Neural Networks for Sequence Learning, arxiv, 2015
  - C. Szegedy et al., Going Deeper with Convolutions, CVPR 2015
  - Andrej Karpathy and Li Fei-Fei, Deep Visual-Semantic Alignments for Generating Image Descriptions, CVPR 2015
  - Y. Wu et al., Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation, arxiv, 2016
  - D. Silver et al., Mastering the game of Go with deep neural networks and tree search, Nature, 2016

Thanks!
Q&A