CS372 Lecture 1

Machine Learning on Big and Small Data @ Google, DeepQ, Stanford

Edward Y. Chang
Poor Suffering Neglected

People Street
Caffe Demos

The Caffe neural network library makes implementing state-of-the-art computer vision

Classification

Click for a Quick Example

Maximally accurate	Maximally specific

dog

domestic animal

canine

carnivore

hunting dog

CNN took 0.224 seconds.
Hope
Caffe Demos

The Caffe neural network library makes implementing state-of-the-art computer vision algorithms easy.

Classification

Click for a Quick Example

Maximally accurate
- clothing
- musical instrument
- covering
- consumer goods
- commodity

Maximally specific

CNN took 0.061 seconds.
Caffe Demos

The Caffe neural network library makes implementing state-of-the-art computer vision systems easy.

Classification

Click for a Quick Example

Maximally accurate  Maximally specific

- dog
- domestic animal
- canine
- hunting dog
- carnivore
Lessons Learned

• Machine does not discriminate
• Machine cannot *predict* if it has not *seen*
• Successful Machine learning demands
  – Scale of data
  – Diversity of data
Generating Data

Using Generative Adversarial Networks (GANs)

Drawing?

From Hung-Yi Lee
GANs: Generative Adversarial Networks
[I. Goodfellow, 2014]
How to put DL into practice?

w/ Big Diversified Training Data
- Image Processing
- Machine Translation
- Games

w/o Labeled Big Data
- NLP tasks (e.g., summarization and classification)
- Healthcare
Machine Learning Handles Big Data
The state between 2006 & 2012 (Google)

• Linear & Sub-linear Algorithms Preferred
  – All data must be processed timely >> “accuracy”
    • E.g., Web site ranking & ads matching
  – Φ: Learning algorithm \( \rightarrow \) pick low complexity algorithms \( O(N) \) or \( O(\log N) \)

• Quadratic Algorithms Considered not Practical
  – \( O(N^2) \) takes \( N \) times longer time than \( O(N) \), where \( N \) can be 1,000,000,000
    – One billion second = 31.7 years!
More Data vs. Better Algorithms

Banko & Brill, 2001

Figure 2. Learning Curves for Confusable Disambiguation
AI Renaissance

- Scale of data
- Scale of computation

- Approximate $N^2$ problem to NP, where $P = \sqrt{N}$
- Employ $P$ threads/cores
- E.g., 1,000,000,000 data instances, $P = 32k$ threads, each GPU has $> 1k$ cores
Google AI Infrastructure
2006 to 2012

• Scalable ML Algorithms
  • Support Vector Machines [NIPS 07]
  • Latent Dirichlet Allocation [TIST 10]
  • Frequent Itemset Mining [ACM RS 08]
  • Spectral Clustering [PAMI 10]
  • Data Driven CNNs [US Patents 11]
    • US9547914, filed in 08/2011
    • US8798375B1, filed in 09/2011
**AlexNet by Krizhevsky** [Krizhevsky NIPS 2012]

- Deep CNN architecture proposed
  - 5 convolutional layers (with pooling and ReLU)
  - 3 fully-connected layers

- Parameters: 300,000+ for 1,000 classes

- ImageNet sponsorship 2010 - 2012
The Renaissance of Deep Learning

NVIDIA Corporation
NASDAQ: NVDA

195.04 USD +4.55 (2.39%) ↑
Oct 21, 2:36 PM EDT · Disclaimer

1 day 5 days 1 month 6 months YTD 1 year 5 years Max

46.20 USD Jun 10, 2016
GPipe: Efficient Training of Giant Neural Networks using Pipeline Parallelism

Yanping Huang
Google Brain
huangyp@google.com

Youlong Cheng
Google Brain
ylc@google.com

Dehao Chen
Google Brain
dehao@google.com

HyoukJoong Lee
Google Brain
hyouklee@google.com

Jiquan Ngiam
Google Brain
jngiam@google.com

Quoc V. Le
Google Brain
qvl@google.com

Zhifeng Chen
Google Brain
zhifengc@google.com

550M parameters

ImageNet Results
1,000,000x Increase in Computation in 2019 since AlexNet (1 week → 1 second)
AI is Hard, AI Deployment is Harder

Andrew Ng

- Non-IT Company Using AI
  Many companies are figuring out how to use AI, but this is not easy. The technology is still complex, and few teams understand AI well enough to implement it effectively. Outside the IT industry, almost no companies have enough access to AI talent.
DEEP LEARNING SIMPLIFIED
Deep Learning Frameworks

- Caffe
- CNTK
- DEEPLearning4J
- dmlc
- mxnet
- TensorFlow
- Theano
- torch
HOW TO CHOOSE?

GCE/AWS

Configurations

Models

Hyper-parameters
Automated AI (AutoML)
Scalable Deep Learning [AIMA 17]

- Cost effective training, ease of use
- Automation in
  - Model selection (shuffleNet, ResNet etc.)
  - Parameter tuning (hyper-parameters)
  - System configuration (# GPUs, # servers)
  - Framework selection (TensorFlow, Caffe, PyTorch, MxNet, etc.)
  - Cloud platform selection (GCE, AWS, Azure etc.)
  - IO/GPU scheduling
  - Data augmentation
- Model compaction using
  - Quantization and compression
  - Information-theory based model reduction [IEEE CVPR 19]
Detecting Diabetic Retinopathy

Model selection/automatic parameter tuning within top 3%
How to put DL into practice?

w/ Big Diversified Training Data
• Image Processing
• Natural Language Processing
• Games

w/o Labeled Big Data
• NLP tasks (e.g., summarization and classification)
• Healthcare
TRICORDER, STAR TREK ENTERPRISE 2270

Competition Rules
- Total weight < 5 pounds (2.22 kg)
- Diagnose 12 diseases & measure 5 vital signs

Competing Teams
- 310 entries in 2013
XPRIZE Tricorder Winners
Vital Sense

Blood/Urine Sense

Optical Sense

Symptom Checker

Breath Sense

Vital Sense
Disease Coverage

• **Blood/Urine Sense**
  – Diabetes
  – Anemia
  – Pneumonia
  – Leukocytosis
  – Urinary Track Infection
  – HIV Screen

• **Vital Sense**
  – Heart Rate
  – Respiratory Rate
  – Body Temperature
  – Blood Pressure
  – Oxygen Saturation
  – Atrial Fibrillation
  – Sleep Apnea
  – Hypertension

• **Optical Sense**
  – Otitis Media
  – Melanoma

• **Symptom Checker**
  – COPD

• **Breath Sense**
  – COPD
Symptom Checker

What symptoms do you have?

- Fever
- Headache
- Joint pain
- Chill
- Lost Taste
- Sore throat
- Weakness

Diseases or further tests

Symptom Probing
From Go to Cure
Reinforcement Learning

Environment = Patient

State = Disease

Action = Symptom Probing

Reward

Agent = Doctor
Reinforcement Learning in AlphaGo

- Policy Network
  \[ p_{\sigma|\rho}(a|s) \]

- Value Network
  \[ v_\theta(s') \]
Exploring vs. Exploiting

• For 3,000 years, placing a stone inside the 4th line is considered risky, and not encouraged.

• AlphaGo: placing on the 5th line could be an ingenious move!
Technical Challenges

• AlphaGo
  – Countable discrete actions
  – Outcome is binary (win or loss)
  – Can explore (self playing with whatever moves)

• Symptom Checker
  – A body system consists of uncountable continuous attributes
  – Outcome is one or two out of 800 (diseases)
  – No avenue for exploration
DeepQ Solutions

[NIPS RL 2017, AAAI 2018]

• Hierarchical Reinforcement Learning
  – A committee of doctors
  – Master agent picks a part/system agent to select a symptom to probe

• Consider Context
  – Who (demographic, heredity, medical history)
  – Where (distribution of diseases by locations)
  – When (seasonal diseases)
DeepQ Solutions

[NIPS RL 2015, AAAI 2018]

• Hierarchical Reinforcement Learning
  – A committee of doctors
  – Master agent picks a part/system agent to select a symptom to probe

• Consider Context
  – Who (demographic, heredity, medical history)
  – Where (distribution of diseases by locations)
  – When (seasonal diseases)
Accuracy

- 35% in 2015
- 50% in 2016
- 63% in 2017
- 76% in 2018
REFUEL: Sparse Feature Exploration in Reinforcement Learning with Application to Disease Diagnosis

[AAAI 2018, NeurIPS 2018]

• Reward shaping
  – Give positive symptoms additional reward
  – While still observing MDP optimality

• Feature rebuilding
  – Use correlation to rebalance features
  – Aim to find key symptoms

• Accuracy Improvement
  – 76% → 85%
Launched at 12 Hospitals 2017-19
Performing Triage
AI EXAMPLE: OPTICAL SENSE TEST FLOW

Preparation

Take out Otoscope and the Base
Please find above device from the right drawer

Place the scope

Plug the scope on your selected ear
gently grab your ear tip and place the scope on your selected ear

Ear Drum Analysis

Analysing...

Find symptoms

Find skin symptoms on your body
Find the skin symptom on your body as your target shooting spot, and cover the skin scope on top of it

Confirm

Please confirm the photo you want to submit
Otitis Media

No

Yes
CNNs and DNNs
Small Data & Human Engineering Features

Accuracy 75% \rightarrow 80.11%
Our Deep + Transfer Learning for OM

[IEEE EMBC 2015]

Accuracy 80.11% → 90.96%
Learning Part-based Representation

Faces  Cars  Elephants  Chairs

Lee et.al., ICML 2009
Visual Cortex Like Bases
Transfer Useful Features for OM

Explicit Similes
Cardiovascular Disease Detection

[ACM MMHealth 2017; IEEE MIPR 2018]
Small Data Learning

- Dimension Reduction
- Transfer Representation Learning
- Active Learning
- Self-Supervised Learning
- Various GAN (generative adversarial network) Models
Original GAN

random noise ➔ Generator ➔ Discriminator ➔ real or fake

fake image

real image
Hemorrhagic vs. Ischemic stroke
Thoracic Diseases
Using GANs to Generate Lung Images [IEEE AIVR 2018]
Empirical Study

1. Random Initialization
2. Transfer Learning
3. Pre-trained w/ Images Generated by Unsupervised GANs
4. Trained w/ Images Generated by Supervised GANs

### TABLE II
**EXPERIMENTAL RESULTS**

<table>
<thead>
<tr>
<th>Methods</th>
<th>Scale of Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
</tr>
<tr>
<td>Methods</td>
<td>AUROC (std)</td>
</tr>
<tr>
<td>Method 1</td>
<td>0.708</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
</tr>
<tr>
<td>Method 2</td>
<td>0.756</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
</tr>
<tr>
<td>Method 3</td>
<td>0.726</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
</tr>
<tr>
<td>Method 4 (2x)</td>
<td>0.713</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
</tr>
<tr>
<td>Method 4 (5x)</td>
<td>0.693</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
</tr>
</tbody>
</table>
KG-GAN: Knowledge Guided GANs

[IEEE ICIP 2019 keynote]

• Guiding the generator to explore diversity productively.

• Allowing the discriminator to tolerate diversity reasonably.
IEEE ICIP keynote & panel
September 2019
Precision Surgery with VR/AR
UCLA, Stanford, Mt. Sinai, Cleveland H., UCSF, etc.
Summary

• Data Driven AI wins out Model Based
• Data, both volume and diversity
• Big data $\rightarrow$ AutoML
• Small data? (future lectures)
• User Privacy (Soteria, a multi-layer blockchain)
Publications

- **Artificial Intelligence in XPRIZE DeepQ Tricorder (XPRIZE Award)**, EY Chang, MH Wu, KF Tang, HC Kao, CN Chou, Multimedia for Personal Health and Health Care, 2017.
COURSE LOGISTICS
Disease Diagnosis and Treatment Using AI

- COVID-19 (April)
- Cancer (May)
- Depression/Anxiety (May)
- Brain Surgery (June)
Course Policies

• No Audit
  – Proprietary data of sponsors
  – Novel ideas of yours

• Lectures are recorded and can be viewed any time to accommodate different time zones, etc..

• No public lecture link and please do not distribute course materials outside the class.
Grading

• Assignments 10%
• Participation/Discussion 10%
• Term Project 80% broken down into five factors
  o Proposal (functional specifications) 10%
  o Literature survey and algorithm design 10%
  o Implementation & Demo quality 40%
  o Presentation/documentation 10%
  o Effort 10% (ranked within a team)
Potential Projects

• COVID-19 & Economy Joint Prediction
• Image-based Diagnosis: Thoracic Diseases
• Covert News into Comic Strip
• Disease Diagnosis
• Knowledge + Perception Fusion, Multimodal BERT
• Virtual Assistant for News or COVID Q&A
  – Collaborate with CS271S
• Your passionate topics
Stay Home!
Stay Safe!