Big Data & Deep Learning: A Powerful Mix

Max Welling

Universiteit van Amsterdam

University of California, Irvine
Overview

- Three exponential growth laws
- Machine learning 101
  - Unsupervised learning
  - Supervised (deep) learning
  - Generative vs. discriminative models

(ML as a "datascope")
Three Exponential Growth Laws
Exponential Growth: Big Data

- Moore's law
- Computer Power
- Data Volume
- Model Capacity
- deep learning
- big data
How much Data = Big Data?

• Total amount generated by humanity so far:
  4 zettabyte = 4,000,000,000,000,000,000,000 byte.

• That's 8 billion hard disks or a pile as high as the moon if every hard disc is 5 cm thick.

• Amount of data doubles about every two years.
Big Data in Astronomy

• "Square Kilometer Array": 1 exabyte per day in 2024.

(1 exabyte = 1000 petabytes = 1 million terabytes = 1 million hard discs per day)
Exponential Growth: Compute Power

Moore's law

Computer Power

Data Volume

Model Capacity

big data

deep learning
Moore's Law Powers Big Data
Exponential Growth: Model Capacity

- Moore's law
- Computer Power
- Data Volume
- Model Capacity
- big data
- deep learning
1943: First NN (+/- N=10)

1988: NetTalk (+/- N=20K)

2009: Hinton’s Deep Belief Net (+/- N=10M)

2013: Google/Y! (N=+/- 1B)

2030-2050 capacity human brain is reached

Deep Neural Networks are Big!
The Power of Deep Learning

Conversational Speech: *Switchboard* challenge

Eric Horwitz
Machine Learning 101
Data-mining: Digging for Information

data = ore
informatie = gold
machine learning = pickaxe
Useful Information is information with which you can make predictions
The Information Sieve

data = ore

useful information = gold

noise = grit
What is this?
“To Know is to Forget”

To generalize one needs to forget the details and remember the essence.
Overfitting: An Experiment
Wisdom of the Crowd

- Everyone guesses the weight of this cow.
- Order all estimates.
- Take the middle guess (1, 2, 3, 4, 5)

Answer: 600 kg
Unsupervised Learning
(no labels)
Hierarchical Clustering of Birds

(with Dilan Görür)
Topic Models

• Every document consists of a small number of topics.
• The algorithm learns the topics distribution per document as well as the words in a topic.
Supervised Learning
(with labels)
Training a Classifier

**Step 1: Train model**

Patients: 3D scan + profile

Controls: 3D scan + profile

Deep learning, multi layer network

**Step 2: Model Toepassen**

New Subjects: 3D scan + profile

Deep learning, multi layer network

Patient! No Patient!
Deep Learning

- Neural networks with many layers of artificial neurons (10B parameters)
- Trained on GPUs (supercomputing at home)

Forward: Filter, subsample, filter, subsample, ...., classify
Backward: backpropagation
Basic Convolutional Network Operations

(slided borrowed from Li Deng)
Increasingly Abstract Features

Features learned from training on different object classes.
Generative vs Discriminative?

Deep neural network

• Advantages generative models:
  • inject expert knowledge
  • model causal relations
  • Interpretable
  • probabilities
  • unlabeled data (semi-supervised learning)

Bayesian network

• Advantages discriminative models:
  • low bias / flexible map from input to target
  • efficient training algorithms available
  • solve the problem you are evaluating on.
  • don't need Bayes rule to classify
Conclusions

• Machine Learning tries to make predictions on future data by learning models from historical data.

• Big data, Moore's law and deep learning are revolutionizing AI.

• So, much, that people are getting scared....(so jump on the bandwagon!)

Don't Let Artificial Intelligence Take Over, Top Scientists Warn
Questions