

QBUS6850 Week 1

Introduction

Dr Stephen Tierney

The University of Sydney Business School

What is Machine Learning?

Arthur Samuel described machine learning as:

the field of study that gives computers the ability to learn without being explicitly programmed.

Machine Learning aims to build algorithms that can learn from and make predictions on data, and has evolved from the study of pattern recognition and computational learning theory in artificial intelligence.

Goal of Machine Learning

The goal is for computers to learn and think like humans - or better than humans.

This raises some questions:

- ▶ What is “learning”?
- ▶ How do we know that a machine is “learning”?
- ▶ Is a machine that learns different from a human?

These questions are out of scope of QBUS6850, but keep them in mind for the next few slides.

Types of Artificial Intelligence (AI)

Weak AI: intelligence that is limited to a specific task.

General AI: intelligence that can perform a wide range of tasks, at or above the level of human capability. Often called “strong AI”.

Perspective on General AI



Yann LeCun @ylecun · Dec 9, 2019

...

Let's start with Artificial Rat-level Intelligence (ARI), then move on to Artificial Cat-level Intelligence (ACI), and so on to Artificial Human-level Intelligence (AHI).

AGI does not exist, other than as Artificial {Gerbil,Goose,Giraffe,Gibbon,Gorilla}-level Intelligence....



Tony Zador @TonyZador · Dec 9, 2019

That does seem to be a common belief.

Whereas I believe that if we could achieve artificial mouse intelligence (AMI), it would be a relatively smaller step to AHI.

(And I don't think the term AGI even makes sense).
twitter.com/nicholdav/stat...



34



127



472



Yann LeCun @ylecun · Dec 10, 2019

...

There is no such thing as Artificial General Intelligence because there is no such thing as General Intelligence.

Human intelligence is very specialized.



39



79



372



Current State of Machine Learning

Currently Machine Learning is considered “**weak AI**”.

Almost all Machine Learning is an extension of statistics.

In many ways Machine Learning is statistics and the distinction between them is blurry and contended.

Current State of Machine Learning

Advances in machine techniques such as **deep learning** seemingly push us closer to general AI.

However these are just more complex models and are still limited to very narrow tasks.

Current State of Machine Learning

Gary Marcus, a professor of cognitive science at NYU gives great perspective on Machine Learning's place in AI:

Realistically, deep learning is only part of the larger challenge of building intelligent machines. Such techniques lack ways of representing causal relationships (...) have no obvious ways of performing logical inferences, and they are also still a long way from integrating abstract knowledge, such as information about what objects are, what they are for, and how they are typically used. The most powerful A.I. systems, like Watson (...) use techniques like deep learning as just one element in a very complicated ensemble of techniques, ranging from the statistical technique of Bayesian inference to deductive reasoning.

Deep Learning is Fragile

People with no idea
about AI, telling me my
AI will destroy the world

Me wondering why my
neural network is
classifying a cat as a dog..



Computer Science Memes

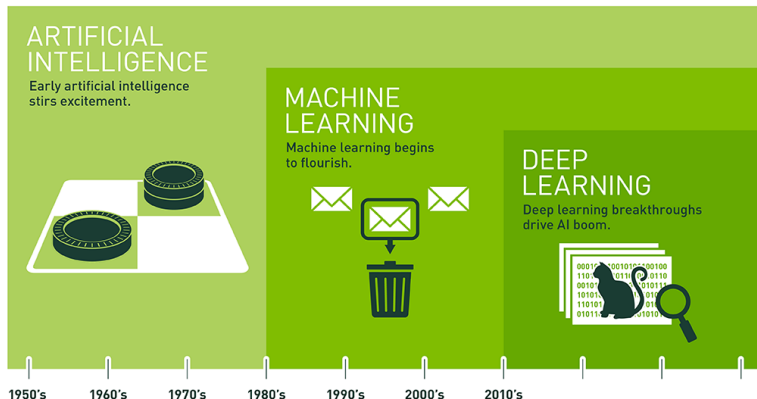
Deep Learning is Fragile

Many of the state of the art deep learning models are easily tricked.

- ▶ Early discovery: <https://arxiv.org/abs/1312.6199>
- ▶ Generating adversarial images:
<https://arxiv.org/abs/1412.1897>
- ▶ Adversarial Patch: <https://arxiv.org/abs/1712.09665>
- ▶ One pixel attack: <https://arxiv.org/abs/1710.08864>
- ▶ Countering adversarial images
<https://arxiv.org/abs/1711.00117>

Video Demo: <https://www.youtube.com/watch?v=MIbFvK2S9g8>

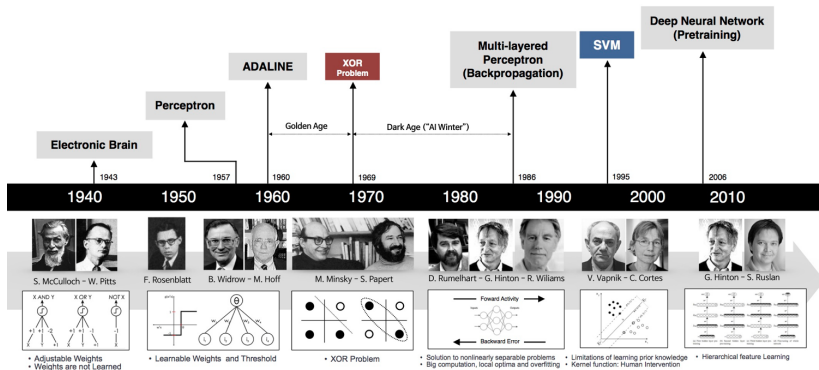
History of Machine Learning



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

<https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/>

History of Machine Learning



<https://www.slideshare.net/devue/251-implementing-deep-learning-using-cu-dnn/4>

Further explanation here:

<http://www.andreykurenkov.com/writing/ai/a-brief-history-of-neural-nets-and-deep-learning-part-4/>

Costs of Deep Learning

The estimated costs of training a model

	Date of original paper	Energy consumption (kWh)	Carbon footprint (lbs of CO ₂ e)	Cloud compute cost (USD)
Transformer (65M parameters)	Jun, 2017	27	26	\$41-\$140
Transformer (213M parameters)	Jun, 2017	201	192	\$289-\$981
ELMo	Feb, 2018	275	262	\$433-\$1,472
BERT (110M parameters)	Oct, 2018	1,507	1,438	\$3,751-\$12,571
Transformer (213M parameters) w/ neural architecture search	Jan, 2019	656,347	626,155	\$942,973-\$3,201,722
GPT-2	Feb, 2019	-	-	\$12,902-\$43,008

<https://arxiv.org/abs/1906.02243>

- ▶ Amazon GPU instances cost anywhere from \$1/hour to over \$10/hour. Training can take on the order of 10^4 hours.
- ▶ This does not include the man hours required to setup and maintain servers, research and implementation time for the models.

<https://openreview.net/forum?id=HyIVB3AqYm>

Weak AI is Valuable

Inspired by the quote from George Box:

All models are wrong but some are useful

In QBUS6850 **we will study a set of models** which can be applied to many problem domains in business.

Why Study Multiple Models?

If we are interested solely in the generalization performance, are there any reasons to prefer one classifier or learning algorithm over another?"

*If we make no prior assumptions about the nature of the classification task, can we expect **any classification method to be superior or inferior overall?***

... on the criterion of generalization performance, there are no context or problem-independent reasons to favor one learning or classification method over another.

Why Study Multiple Models?

The “No Free Lunch” theorem means that no single model will always be the best.

*The apparent superiority of one algorithm or set of algorithms is **due to the nature of the problems investigated and the distribution of data**”*

Chapter 9, Pattern Classification, Duda, Stork and Hart

Summary: <https://www.kdnuggets.com/2019/09/no-free-lunch-data-science.html>

Which Models Then?

We will study models which have a proven track record of performance and broad applicability.

The bulk of QBUS6850 will focus on:

- ▶ Neural Networks
- ▶ Ensembles

It has been well established that these particular model families satisfy most weak AI needs.

Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?