An Introduction to Machine Learning and AI and a brief discussion on their relevance in the global order

Andalo January 9th 2018 – Isodarco School

SAPIENZA
UNIVERSITÀ DI ROMA

Marco Schaerf

Department of Computer, Control and Management Engineering Antonio Ruberti

What is Intelligence?

- Intelligence:
  - “the capacity to learn and solve problems” (Webster's dictionary)
  - in particular,
    - *the ability to solve novel problems*
    - *the ability to act rationally*
    - *the ability to act like humans*

- Artificial Intelligence
  - build and understand intelligent entities or agents
  - 2 main approaches: “engineering” versus “cognitive modeling”

What's involved in Intelligence?

- Ability to interact with the real world
  - to perceive, understand, and act
  - e.g., speech recognition and understanding and synthesis
  - e.g., image understanding
  - e.g., ability to take actions, have an effect
- Reasoning and Planning
  - modeling the external world, given input
  - solving new problems, planning, and making decisions
  - ability to deal with unexpected problems, uncertainties
- Learning and Adaptation
  - we are continuously learning and adapting
  - our internal models are always being "updated"
    - e.g., a baby learning to categorize and recognize animals

Academic Disciplines relevant to AI

• Philosophy	Logic, methods of reasoning, mind as physical system, foundations of learning, language, rationality.
• Mathematics	Formal representation and proof, algorithms, computation, (un)decidability, (in)tractability
• Probability/Statistics	modeling uncertainty, learning from data
• Economics	utility, decision theory, rational economic agents
• Neuroscience	neurons as information processing units.
• Psychology/ Cognitive Science	how do people behave, perceive, process cognitive information, represent knowledge.
• Computer engineering	building fast computers
• Control theory	design systems that maximize an objective function over time
• Linguistics	knowledge representation, grammars

Artificial Intelligence

- Most important pioneer for AI (and Computer Science):

  Alan M. Turing, 1912-1954
- From his 1950 paper: “Computing Machinery and Intelligence”
- “I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.”
- Crux of paper: A compelling philosophical analysis for the feasibility of intelligent machines.

AI: Early Optimism

- 1958, H. A. Simon and A. Newell: "within ten years a digital computer will be the world's chess champion”

- 1967, M. Minsky: "Within a generation ... the problem of creating 'artificial intelligence' will substantially be solved."

“AI Winters”

- The first AI winter 1974−1980: slow progress and dearth of research funding

- The second AI winter 1987−1993: the “Japanese Fifth-Generation bust” and dearth of research funding

AI Breakthroughs, I

- 1997: IBM's Deep Blue beats Kasparov.

AI Breakthroughs, II

- 2011: IBM's Watson defeats the two greatest Jeopardy! champions, Brad Rutter and Ken Jennings, by a significant margin.

AI Breakthroughs, III

- 2016: AlphaGo beats Lee Se-dol to take Google DeepMind Challenge series!

Automated Driving, I

- 2005: DARPA Grand Challenge - Stanford autonomous vehicle drives 131 miles along an unrehearsed desert trail.

Automated Driving, II

Artificial Intelligence Topics

Artificial Intelligence: A Modern Approach
(Third edition) by Stuart Russell and Peter Norvig

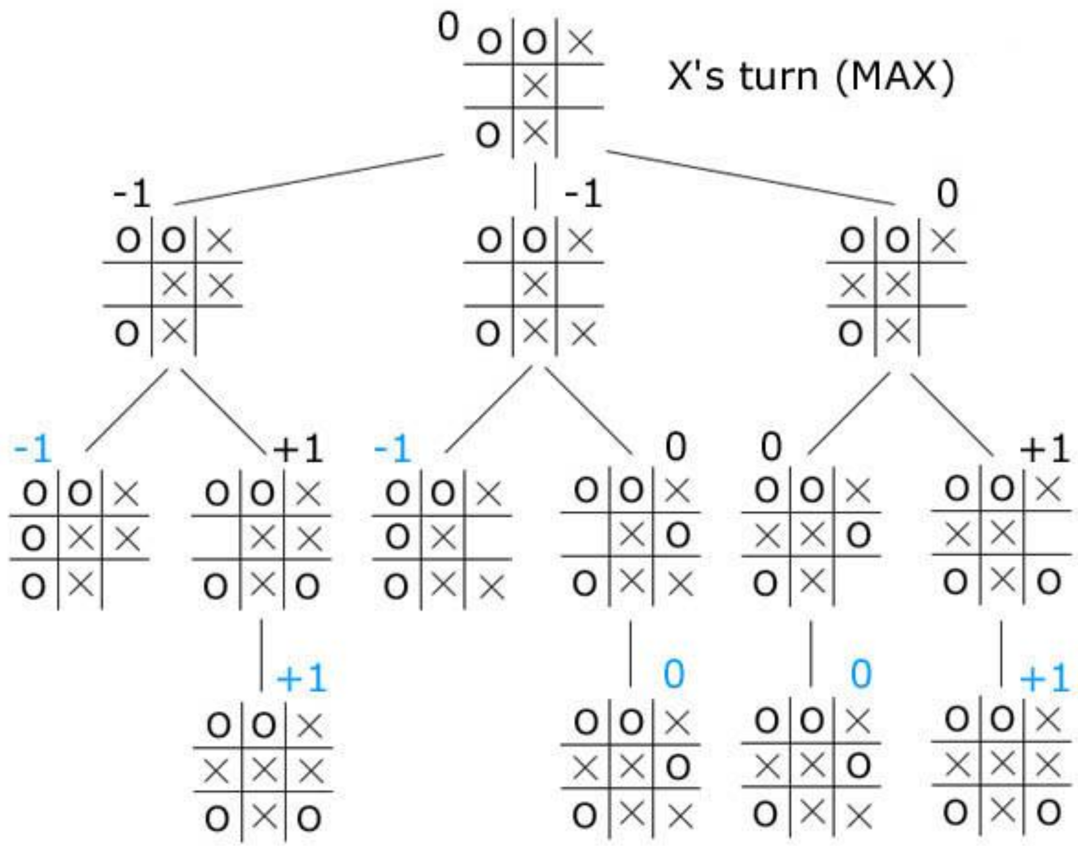
1. Artificial Intelligence
2. Problem Solving
3. Knowledge and Reasoning
4. Uncertain Knowledge and Reasoning
5. (Machine) Learning
6. Communicating, Perceiving, and Acting

Brief survey of AI Techniques

- Use (mostly) games to explain the techniques.
- Search
  - Tic-Tac-Toe
  - Chess
- Game Theory
  - Poker
- Planning
  - Blocks world
- (Deep) Learning
  - Go

Tic-Tac-Toe: Game Tree

Simple game, game tree can be completely explored. Number of states bounded by b^d where b (branch) is the number of available moves (at most 9) and d (depth) is the length of the game (at most 9)



Chess: Game Tree

Complexity of Chess

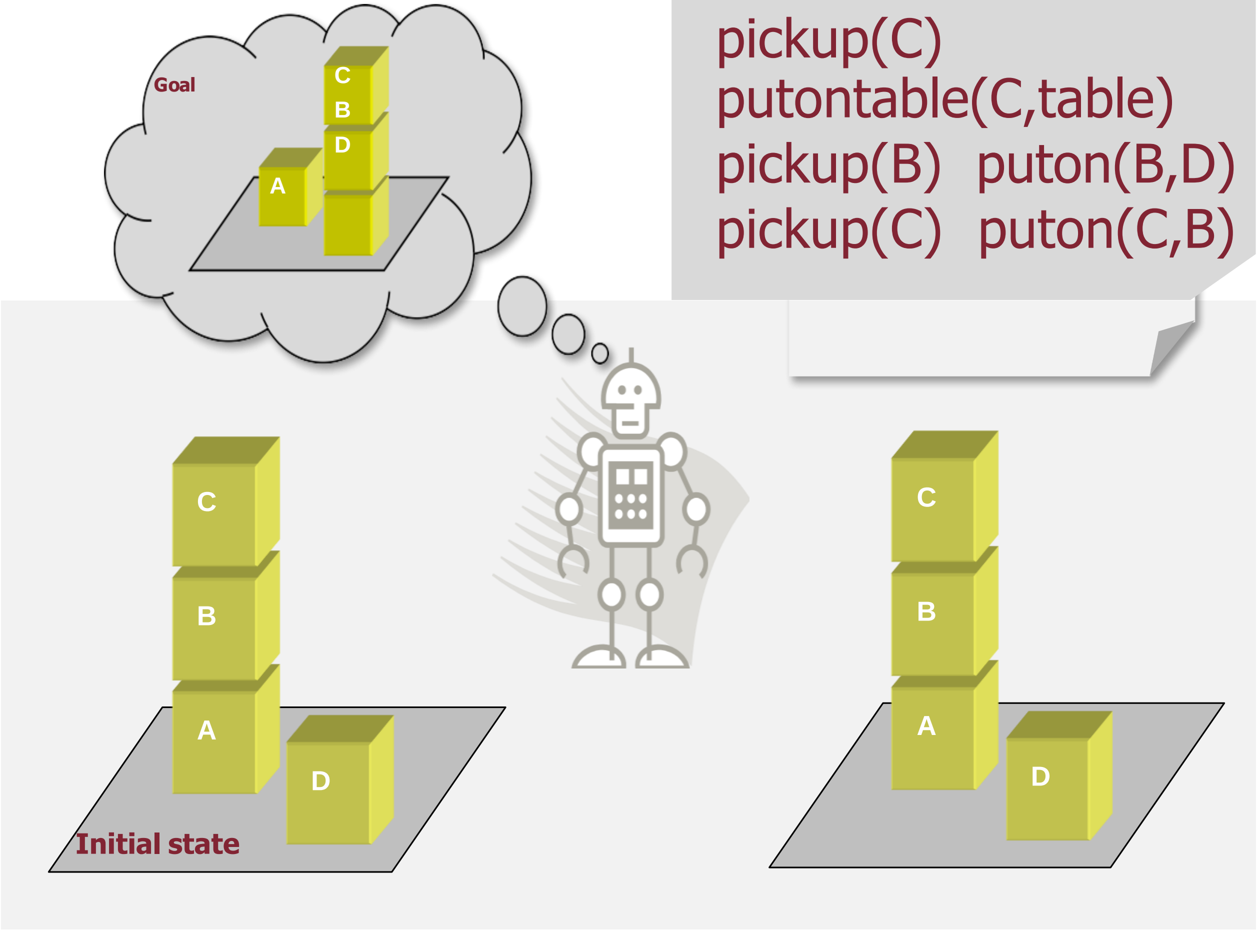
- High complexity: branch approximately 30, depth approximately 60. Even considering all equivalent states, number of states at least 10^{50}.
- Analyzing all states is unfeasibile, we need «some intelligence»
- Deep Blue is a combination of brute-force (expensive and specialized HardWare) and clever position-valuation algorithms
- Developed techniques are quite specific and tailored to chess

Poker: A game of incomplete knowledge

- Moves are limited, but ignorance about the state makes state space HUGE. Techniques used:
  - Space reduction
  - Game theory
  - Agents behaviour modeling
  - Statistics on players
  - Montecarlo runs
  - LEARNING

Automated Planning

Blockworld



Plan

pickup(C)
putontable(C,table)
pickup(B) puton(B,D)
pickup(C) puton(C,B)

Planning problem specification

Input:

– initial (current) state of the world
– description of actions that can change the world
– desired state of the world

Output:

– a sequence of actions (a plan)

Properties:

– actions in the plan are unknown
– time and resources are not assumed

Go: a recent breakthrough

- The search space of Go is HUGE (10^{170} states)
- Techniques used for chess were not much effective in developing a Go program
- Techniques used in AlphaGo:
  - Value network assessing the current state (specific for Go)
  - Policy network choosing the next move (specific for Go)
  - Montecarlo tree search
  - Deep Learning using supervised learning (with the help of human experts) and reinforcement learning (from games of self-play)

Learning and Deep Learning

- Learning is the process of acquiring new information from experience. Many differeent techniques:

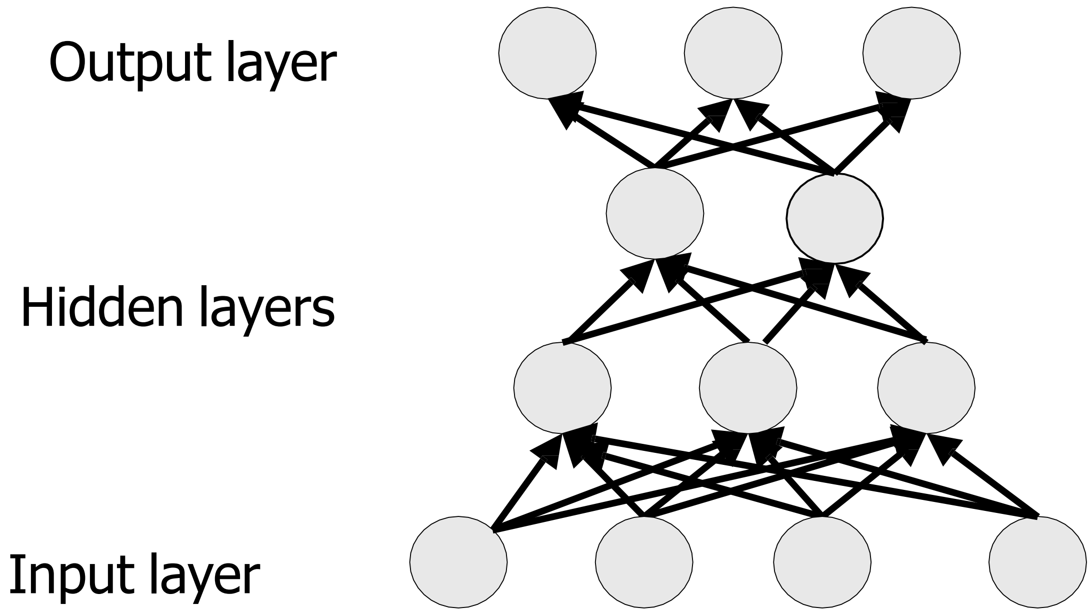
Decision tree learning
Artificial neural networks
Inductive logic programming
Clustering
Reinforcement learning
Similarity and metric learning
Genetic algorithms
Learning classifier systems
Association rule learning
Deep learning
Support vector machines
Bayesian networks
Representation learning
Sparse dictionary learning
Rule-based machine learning

Differences between approaches

- Some are mathematically well-understood and the properties are well-known, as an example:
  - Decision tree learning
  - Inductive logic programming
  - Bayesan networks
- Others are less well-understood and their properties less formally analyzed, as an example:
  - Artificial neural networks
  - Deep learning

Deep architectures

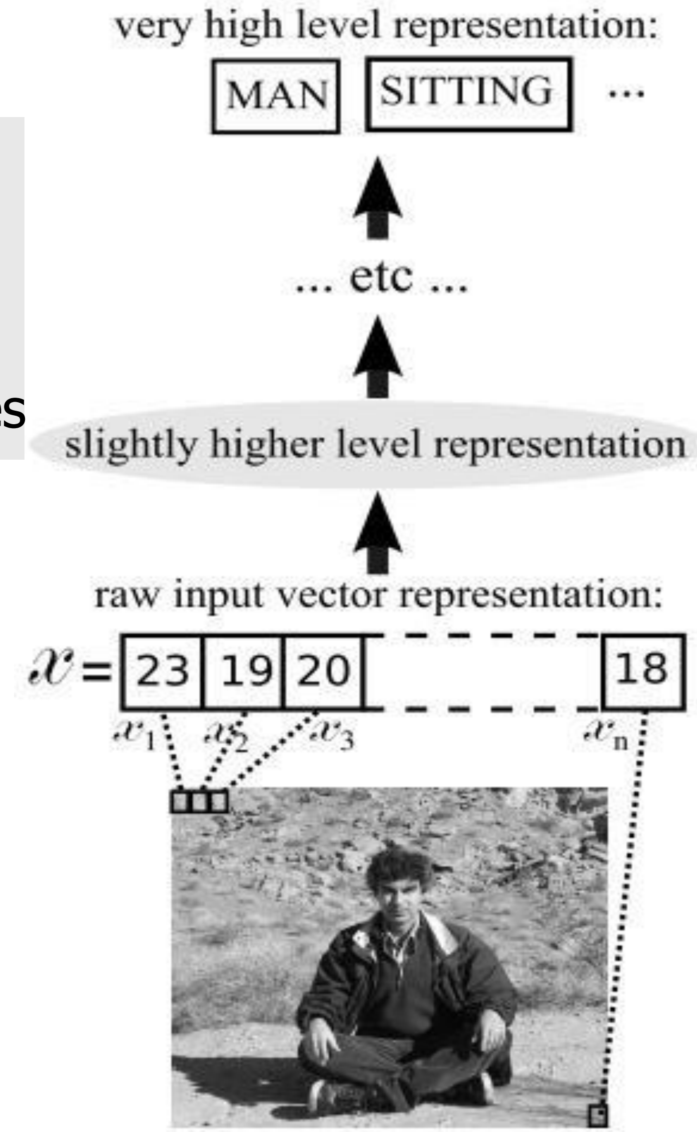
Defintion: Deep architectures are composed of *multiple levels* of non-linear operations, such as neural nets with many hidden layers.



Goal of Deep architectures

Goal: Deep learning methods aim at

- learning *feature hierarchies*
- where features from higher levels of the hierarchy are formed by lower level features



edges, local shapes, object parts

Low level representation

Deep Learning History

- **Inspired** by the architectural depth of the brain, researchers wanted for decades to train deep multi-layer neural networks.
- **No success**ful attempts were reported before 2006 …

  Researchers reported positive experimental results with typically two or three levels (i.e. one or two hidden layers), but training deeper networks consistently yielded poorer results.
- **Exception**: convolutional neural networks, LeCun 1998
- **SVM**: Vapnik and his co-workers developed the Support Vector Machine (1993). It is a shallow architecture.
- **Digression**: In the 1990's, many researchers abandoned neural networks with multiple adaptive hidden layers because SVMs worked better, and there was no successful attempts to train deep networks.
- **Breakthrough in 2006**

Breakthrough

Deep Belief Networks (DBN)

Hinton, G. E, Osindero, S., and Teh, Y. W. (2006).
A fast learning algorithm for deep belief nets.
Neural Computation, 18:1527-1554.

Autoencoders

Bengio, Y., Lamblin, P., Popovici, P., Larochelle, H. (2007).
Greedy Layer-Wise Training of Deep Networks,
Advances in Neural Information Processing Systems 19

Theoretical Advantages of Deep Architectures

- Some functions cannot be efficiently represented (in terms of number of tunable elements) by architectures that are too shallow.
- Deep architectures might be able to represent some functions otherwise not efficiently representable.
- **More formally**:

  Functions that can be compactly represented by a depth k architecture might require an exponential number of computational elements to be represented by a depth k – 1 architecture

- The consequences are
  - **Computational**: We don't need exponentially many elements in the layers
  - **Statistical**: poor generalization may be expected when using an insufficiently deep architecture for representing some functions.

Limits of Deep learning (thus far) (Gary Marcus, 2018)

- is data hungry
- is shallow and has limited capacity for transfer
- has no natural way to deal with hierarchical structure
- has struggled with open-ended inference
- is not sufficiently transparent
- has not been well integrated with prior knowledge
- cannot inherently distinguish causation from correlation
- presumes a largely stable world, in ways that may be problematic
- works well as an approximation, but its answers often cannot be fully trusted
- is difficult to engineer

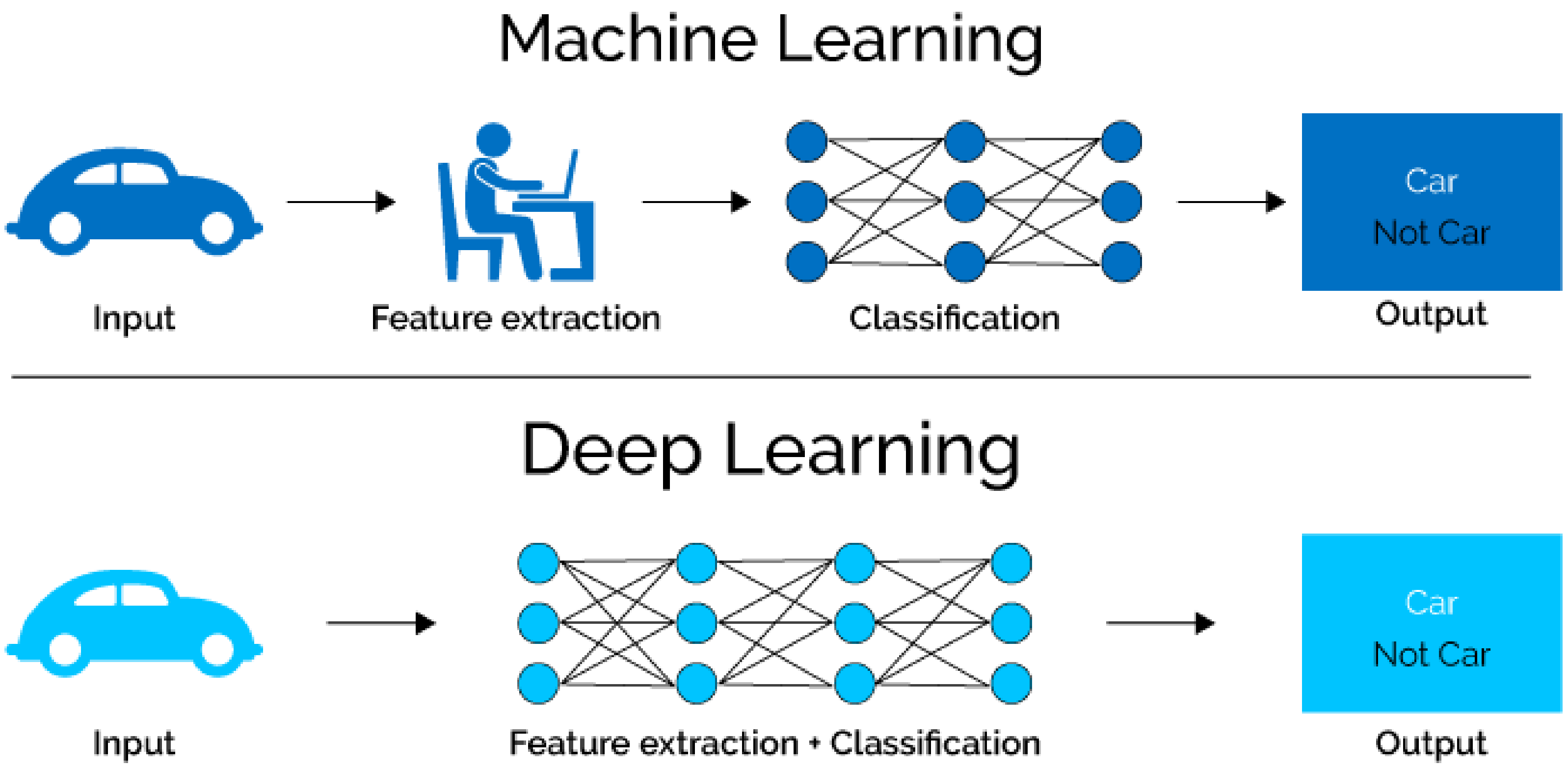
Advances in 2017

- Libratus AI: Poker-playing system beating the best human players at heads-up no-limit Texas hold’em
- Automatic extraction of features
- AlphaGoZero and AlphaZero
- Theory of “information bottleneck” by Naftali Tishby
- Acknowledgement of the importance of bias

Libratus

- Very advanced poker-playing AI software using:
  - Game Theory
  - Game decomposition
  - Optimization techniques
  - Learning (only a bit)
- It beat 4 top professional players in a 120.000 (duplicated) hands of **no-limit** heads-up Texas hold’em. It ended up as the only player with a surplus (over 1.7M chips)

Automatic extraction of features



- It works but you need tons of (reliable, unbiased) (hand)-labelled data

AlphaGoZero and AlphaZero

- Developed by DeepMind, a company acquired by Google in 2014.
- Differently from AlphaGo, AlphaGoZero has no feature-seection phase but it learnt the features by playing repeatedly against AlphaGo.
- It overplayed AlphaGo after 3 days of training (on a multi-milion dollar hardware!)
- Using the same technique, AlphaZero outplayed the best chess engine (Stockfish) after a few hours of training

The general idea

1. Start from an existing system
2. Design a (more complex but ‘featureless’) new system and use it to learn from the existing system
3. Beat the existing system !!

- Repeat points 1-3 over and over

- Works best when it is possible to assign a reward (reinforcement-learning style)
- Improvements obviously decrease over iterations

Theory of Deep NN “information bottleneck”

- Based on the “information bottleneck” framework developed by Naftali Tishby (Hebrew Univ. Jerusalem)
- Uses information theory (first introduced by Shannon) to model the behaviour of deep neural networks
- Formalizes the task of understanding the «relevant» information and discarding (forgetting) the «details»
- Provides a more mathematically-sound way to interpret the results of Deep neural networks

Important problem to acknowledge: Data Bias

- 14th century: an oblique or diagonal line
- 16th century: undue prejudice
- 20th century: systematic differences between the sample and a population
- In ML: underfitting vs overfitting
- In Law: judgments based on preconceived notions or prejudices as opposed to the impartial evaluation of facts. Impartiality underpins jury selection, due process, limitations placed on judges etc. Bias is hard to fix with model validation techniques alone. So you can have an unbiased system in an ML sense producing a biased result in a legal sense.
- Bias is a skew that produces a type of harm.

Lessons on bias

1. Data is not neutral. Data cannot always be neutralized. There is no silver bullet for solving bias in ML & AI systems.
2. Bias in MLaaS is harder to identify and also correct as we do not build them from scratch and are not always privy to how it works under the hood.
3. There are two main kinds of harms caused by bias: Harms of allocation and harms of representation. The former takes an economically oriented view while the latter is more cultural.
4. Structural bias is a social issue first and a technical issue second. If we are unable to consider both and see it as inherently socio-technical, then these problems of bias are going to continue to plague the ML field.
5. Instead of just thinking about ML contributing to decision making in say hiring or criminal justice, we also need to think of the role of ML in the harmful representation of human identity.

Examples of Bias

	denigration	stereotype	recognition	under-representation	ex-nomination
Image search for 'CEO' yields all white men on first page of results.			x	x	x
Google Photo mislabels black people as 'gorillas'	x				
YouTube speech-to-text does not recognize women's voices			x		x
HP Cameras' facial recognition unable to recognize Asian people's faces			x	x	x
Amazon labels LGBTQ literature as 'adult content' and removes sales rankings		x	x		x
Word embeddings contain implicit biases [Bolukbasi et al.]	x	x	x	x	x
Searches for African American-sounding names yield ads for criminal background checks [Sweeney]	x	x		x	

Source: Kate Crawford's NIPS 2017 Keynote presentation: Trouble with Bias

Applications of AI in the defence sector

- (Autonomous) weaponized robots
- (Autonomous) control systems for nuclear arms
- Training and simulation systems
- Surveillance systems (project Maven, similar projects by other countries)

Deep Learning in Defense systems

- Needs tons of annotated data or a reward system to assess the quality of the proposed answer.
- Simulation systems can provide rewards
- Recognition data is now widely available, but it needs to be annotated
- Data must also be of good quality, unbiased and not accessible (immune to cybersecurity threats)

Project MAVEN

- Pilot project of the Defense Department to interpret the videos coming from drones, US spy planes and satellites.
- A significat part is already annotated, so that it can be used as a training set. The goal is to have at least 1M images in the training set
- Very small team and dedicated team, short developing time through the use of agile methodology (highly unusual for DoD)
- Very successusful but limited in scope

Criticalities

- Companies invest much more in AI than defense sector. Best people go to companies
- Defense departments tend to develop in-house (for obvious reasons) but this means that private companies have better tools. Terrorists might have access to better technologies than the armies (e.g. UAVs)
- Use of external technology (like project MAVEN) raises issues of maintenance and adaptation
- AI systems do not satisfy the current verification and validation processes of DoD

Conclusions

- AI techniques will be heavily used in the defense industry.
- There are many useful techniques, but the most important one is (Deep) Learning
- Availability, reliability and unbiasedness of data is crucial

My greatest fear

«Expertise is overrated»

Belief held by many politicians, defense officials, AI developers, public opinion leaders and many more around the world.

If technology is used improperly many issues will arise