Computational Thinking and AI for Philosophers

Dr. Mohamed Siala
homepages.laas.fr/msiala

INSA-Toulouse & LAAS-CNRS

April 21, 2020

Context

- Recent personnel interest in ethical AI
- This is a talk for non computer scientist
- Many notions are introduced in a non-formal way
- The most challenging talk I ever prepared : So many things to talk about, I should avoid any technical material, present only important issues, ...
- Three parts: computational thinking, fair AI, explainable AI
- Purpose of the talk : To present ethical AI issues from a computer scientist point of view

Context


ACM FAccT Conference
2020
2019
2018
Network
Organization
ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)
A computer science conference with a cross-disciplinary focus that brings together researchers and practitioners interested in fairness, accountability, and transparency in socio-technical systems.

What is Computational Thinking?

Definition

- Computational Thinking: How to solve problems using a finite sequence of steps

What is Computational Thinking?

Example 1: The Fried Egg Problem





Example 1: The Fried Egg Problem

First Method

1. Break the egg (1 second)
2. Put salt/pepper (1 second)
3. Pour the oil on the pan (1 second)
4. Turn on the cooker (1 second)
5. Put the pan on the cooker (1 second)
6. Wait for the oil to heat (60 seconds)
7. Pour the egg (1 second)
8. Wait two minutes (120 seconds)
9. Total duration: 186 seconds

Example 1: The Fried Egg Problem

Second Method

1. Turn on the cooker : 1 second
2. Put the pan on the cooker : 1 second
3. Pour the oil on the pan : 1 second
4. Wait for the oil to heat : 60 seconds
5. Break the egg : 1 second $\rightarrow$ 0 second
6. Put salt/pepper 1 second $\rightarrow$ 0 second
7. Pour the egg 1 second $\rightarrow$ 0 second
8. Wait two minutes 120 seconds
9. Total duration: 183 seconds

The Fried Egg Problem

- With method 1 or method 2, the result is exactly the same: (a fried egg)
- The second method is better because we **gained** 3 seconds

The choice of the sequence of steps (or method) matters

Example 2: How to Sort a Deck of Cards

Example 2: How to Sort Cards

- Method 1 :
  - Pick the smallest, put it in first position
  - Pick the second smallest, put it in second position
  - Repeat the same process . . . until all cards are sorted
- Method 2 :
  - Split the cards in 2 groups
  - Sort each group separately
  - Merge the two sorted groups

The time difference between the two methods can be significant. **This is particularly true when dealing with a large number of cards**

Computational thinking

Definition

Computational Thinking: how to **solve problems** using a **finite sequence of steps**

Definition : Problem

A problem is defined by an **input** (a set of objects) and a **question** on the input

Example: The Sorting Cards Problem

- **Input**: a set of numbered cards
- **Question**: Find a sequence of the cards such that the cards are sorted in an increasing order

Algorithms

Definition : Algorithm

Given a problem $\mathcal{P}$, an **algorithm** is a finite sequence of steps to answer the question of $\mathcal{P}$

Example: Algorithms for the Sorting Cards Problem

Algorithm 1

- Pick the smallest, put it in first position
- Pick the second smallest, put it in second position
- Repeat the same process . . . until all cards are sorted

Algorithm 2

- Split the cards into two groups
- Sort each part separately
- Merge the two sorted parts

Computational Thinking

- A problem can occur in many situations (Sorting cards problem is essentially the same problem as sorting the prices of an amazon search)
- For one problem, we can have different algorithms
- The different algorithms may have different runtimes
- Some problems are non computable. That is, there is no algorithm to solve them

Back to Algorithms

What matters to a computer scientist?

- Since many algorithms can exist to solve a given problem, which one should we choose?
- We evaluate the algorithms based on their **runtime** and memory space
- Theoretical analysis before implementing the algorithm based on the number of steps (operations) and the memory needed for each operation
- Mostly worst-case reasoning
- **Time-complexity** and **Space-complexity** as mathematical functions on the size of the input

Some properties

- **Problem Hardness:** A problem A is **harder** that another problem B if the best known algorithm to solve A takes longer time than the best algorithm to solve B
- Tractable/Intractable problems = short/long time to solve
- **A slightest change in the definition of a problem can have a significant impact on its hardness**
- **Many "tools" are developed to deal with intractable problems**

Intractable problems?

- Intractable problems are everywhere (timetabling, scheduling, management, predictions, cryptography)
- The community has developed many “advanced tools” to deal with them

What to remember

- Problem
- Algorithm
- Some problems are inherently harder than others

Artificial Intelligence

- Artificial Intelligence: "The science of making machines do things that would require intelligence if done by men." [Minsky, 1988]
- The definition of AI is still a debate (philosophically, but even within the computer science/mathematics community)..
- Personnel point of view: I used to see AI as **"proposing smart algorithms for hard problems"**. Recently I consider it also from what comes out of the applications of the algorithms (using algorithms for health care/driving cars/music recommendation, etc).

Increasing Number of Real Life and Social AI Applications

AI: Increasing Number of Real Life and Social Applications

- The diverse applications of AI raised many ethical issues and questions
  - Job applications: AI that parses CVs for software engineers and recommends to hire mostly men
  - Credit scoring: AI that gives a credit score (for bank loans and credit applications) that recommends people from a particular geographical region
  - Compass tool: (2016) used by judges in the US to predict which criminals are likely to re-offend is found to be biased by the skin color ("race" African-American/Caucasian).

Later I'll discuss why this happened

COMPASS data and Rule-based Predictions

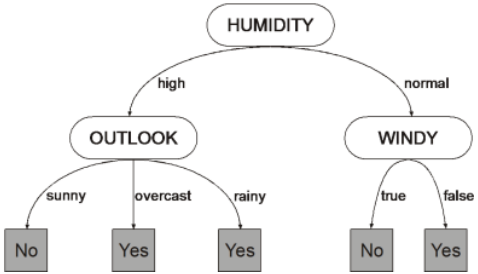
Sex	Age	Priors	Juvenile Felonies	Juvenile Crimes	Race
Male	15	1	0	1	Caucasian
Male	15	1	0	1	African-American
Female	33	1	0	1	African-American
Female	27	0	1	0	Caucasian
Male	41	0	1	0	Caucasian
...	...	...	...	...	...

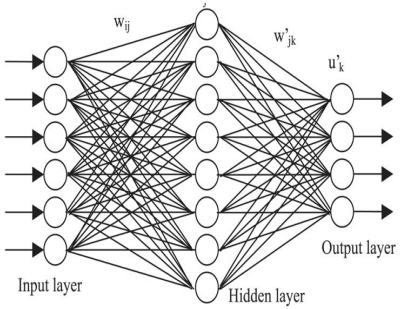
The problem is to predict recidivism. That is, the tendency of a convicted criminal to re-offend.

Example of rule list for the above dataset

```
if [priors:>3] then recidivism
else if [juvenile-felonies:>0] then recidivism
else if [age:26-45] then not recidivism
else if [age:>45] then not recidivism
else if [priors:2-3] then recidivism
else if [age:18-20] then recidivism
else if [race:Caucasian] then not recidivism
else if [juvenile-crimes:>0] then recidivism
else if [priors:0] then not recidivism
else if [sex:Female] then not recidivism
else recidivism
```

Learning Models (Examples)





Computationally Speaking.. What is a Prediction Problem?

- **Input:** Some (historical) data coming from an unknown distribution and a fixed learning model
- **Question** What is the best way to adapt the learning model to my data in order to minimize the prediction error

Why Unfair Predictions?

- The data used to build the learning model is biased
- The perfect learning model is therefore doomed to be biased
- **Don't blame the algorithm, blame the source of the data (society)?**
- How (ideally) should such an issue be addressed?
  - Clean the data (removing bias)
  - Build unbiased algorithms

Computationally Speaking.. How a Fair Prediction Problem should be?

- **Input:** Some (historical) data coming from an unknown distribution and a fixed learning model
- **Question** What is the best way to adapt the learning model to my data in order to minimize the prediction error and satisfy some fairness requirements

Fairness Requirements

Paper [Verma and Rubin, 2018]: *Fairness definitions explained*, Verma, Sahil and Rubin, Julia, 2018 IEEE/ACM International Workshop on Software Fairness

	Definition	Paper	Citation #	Result
3.1.1	Group fairness or statistical parity	[12]	208	×
3.1.2	Conditional statistical parity	[11]	29	✓
3.2.1	Predictive parity	[10]	57	✓
3.2.2	False positive error rate balance	[10]	57	×
3.2.3	False negative error rate balance	[10]	57	✓
3.2.4	Equalised odds	[14]	106	×
3.2.5	Conditional use accuracy equality	[8]	18	×
3.2.6	Overall accuracy equality	[8]	18	✓
3.2.7	Treatment equality	[8]	18	×
3.3.1	Test-fairness or calibration	[10]	57	✓
3.3.2	Well calibration	[16]	81	✓
3.3.3	Balance for positive class	[16]	81	✓
3.3.4	Balance for negative class	[16]	81	×
4.1	Causal discrimination	[13]	1	×
4.2	Fairness through unawareness	[17]	14	✓
4.3	Fairness through awareness	[12]	208	×
5.1	Counterfactual fairness	[17]	14	–
5.2	No unresolved discrimination	[15]	14	–
5.3	No proxy discrimination	[15]	14	–
5.4	Fair inference	[19]	6	–

Fairness Requirements

Some Questions

- Which measure to choose?
- What should be done in case of conflict between different measures?

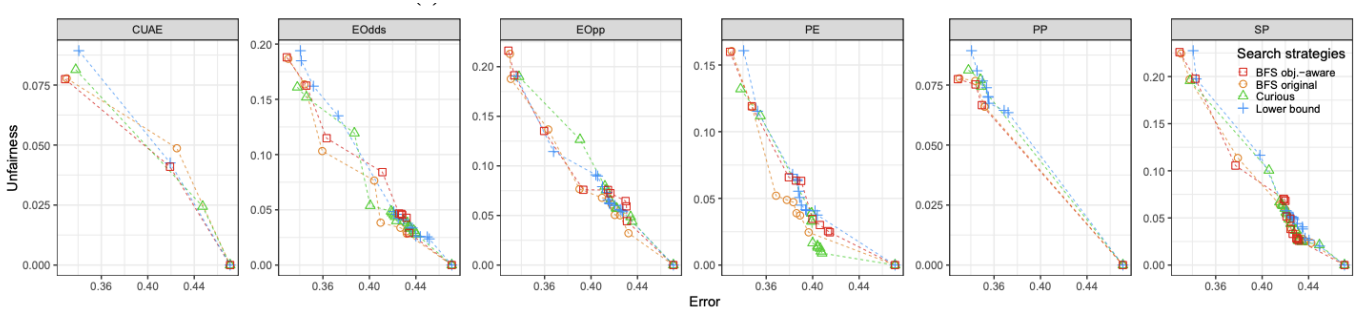
Fairness Requirements: A Personnel Point of View

- It's the job of computer scientists/mathematicians to develop algorithms that support as many fairness measures as needed
- The choice of the fairness measure should be done at the decision-making level

Discussion Question: In an ideal scenario, how should fairness be treated in AI?

Some Work in Progress

Pre-print: *Learning Fair Rule Lists*,
Ulrich Aïvodji, Julien Ferry, Sébastien Gambs, Marie-José Huguet, Mohamed Siala
`https://arxiv.org/abs/1909.03977`



(b) Performances of `FairCORELS` on COMPAS dataset.

Explainability

- The GDPR (General Data Protection Regulation) explicitly mentions that, in the context of algorithmic decision-making, every user has the right to **explanation**.
  `Link:https://gdpr-info.eu/` s
- First question: What does it mean to "explain"?
- Second question: What if we cannot "explain"?

What does it mean to "explain"?

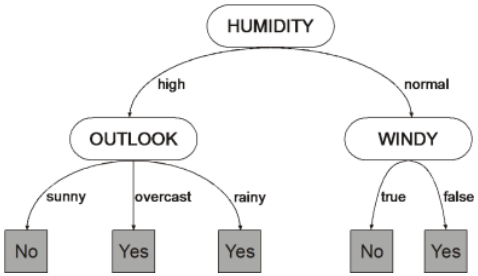
- Personnel view: it depends on the context : The nature of explanation can differ from application to application
- Explain to who? See an interview of Richard Feynman on the why question `https://www.youtube.com/watch?v=Q1lL-hXO27Q&feature=emb_title`
- In the case of prediction: the community seems to agree that an explanation for a prediction is a set of input features that contributed to the prediction

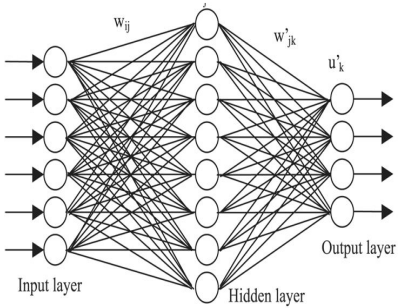
Example

Example of rule list

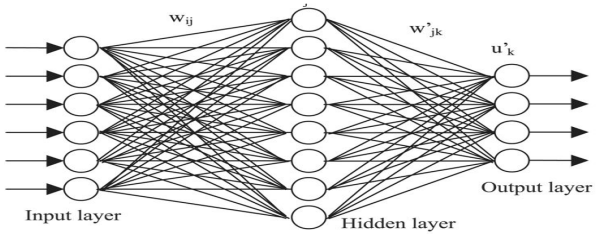
```
if [priors:>3] then recidivism
else if [juvenile-felonies:>0] then recidivism
else if [age:26-45] then not recidivism
else if [age:>45] then not recidivism
else if [priors:2-3] then recidivism
else if [age:18-20] then recidivism
else if [race:Caucasian] then not recidivism
else if [juvenile-crimes:>0] then recidivism
else if [priors:0] then not recidivism
else if [sex:Female] then not recidivism
else recidivism
```

Explaining Prediction: Interpretable vs. Black-Box Models

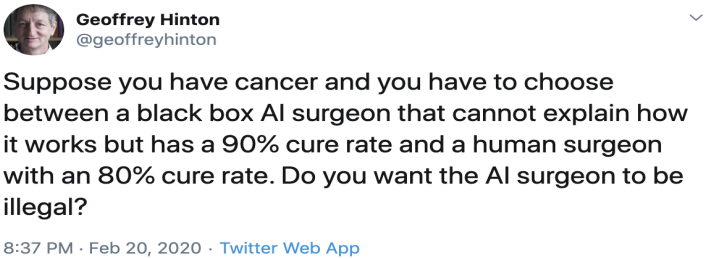




What if we cannot "explain"?



- Deep learning models: Huge success in many applications: image recognition problems, autonomous cars, heath-care systems, etc
- Hard to explain the predictions (no obvious way to figure out with part of the input influenced the prediction)



Do we really need to explain?

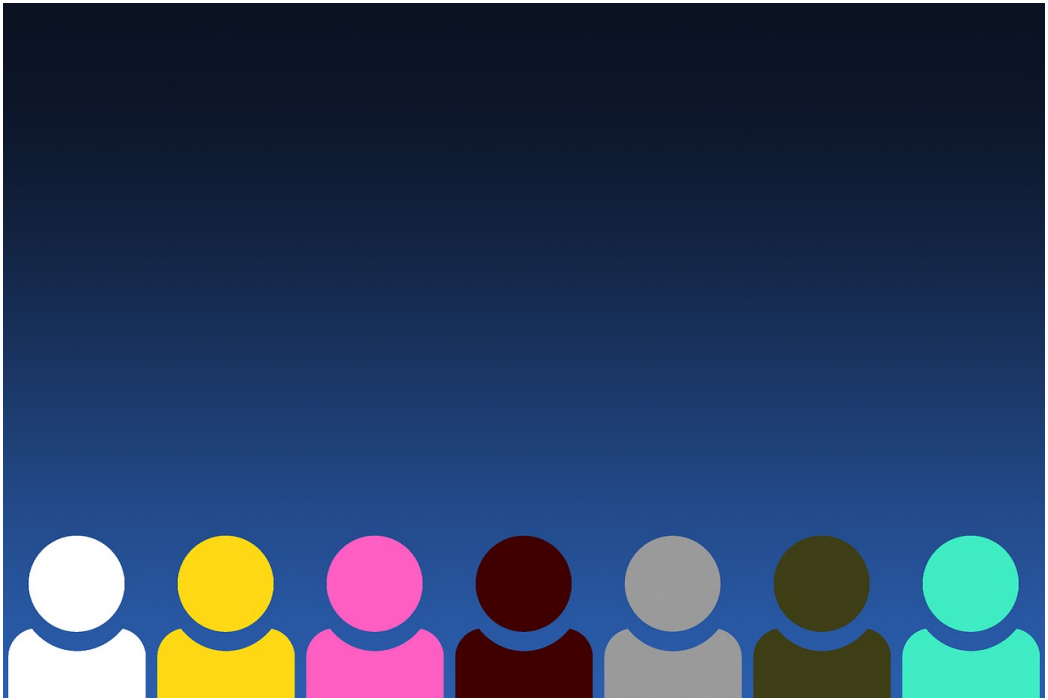
FairWashing [Aïvodji et al., 2019]

- Imagine situations where it is possible to find multiple explanations for the same action. Which one to choose?
- What if some explanations are "more fair" than others?
- A company proposing an AI decision making service can always present the fairest explanations to the user..
- **Interpretable Models are well encouraged to avoid such a problem. However, the cost could be huge in terms of quality of predictions.**

Some Discussion Questions

- Do we really need to explain AI algorithms?
- Explainable and Fair AI algorithms are not for free: **computational challenges**, and important trade-off with the quality of predictions
- For computer scientists/mathematicians, we need formal definitions of fairness, explanability, etc
- Ethical AI is no longer an option, it is mandatory today.
- We (computer scientists/mathematicians) need to work with philosophers on the decision-making level
- Other ethical issues related to AI?

Thank you!



References


Aïvodji, U., Arai, H., Fortineau, O., Gambs, S., Hara, S., and Tapp, A. (2019).
Fairwashing: the risk of rationalization.
arXiv preprint arXiv:1901.09749.


Minsky, M. (1988).
Society of mind.
Simon and Schuster.


Verma, S. and Rubin, J. (2018).
Fairness definitions explained.
In *2018 IEEE/ACM International Workshop on Software Fairness (FairWare)*, pages 1–7. IEEE.