How can we control what we can’t exactly measure?

On the Accountability of Black Boxes

30/04/2019, Strata London

Yiannis Kanellopoulos, @ykanellopoulos
Difficult to do context switching
About Code4Thought

Code4Thought is a startup that aspires to help organisations be in control of their automated decision making systems.

It does that by analysing black-box models as follows:

- Assessing them based on their FAT (Fairness, Accountability, Transparency) principles,
- Explaining the decisions these models make (current focus on those utilising structured data).
How it all started: “You have the right not to know”
Spicing things up - Weapons of Math Destruction*:

- The CDO/CDS Subprime Mortgage Crisis brought the world close to a financial meltdown;

- In 2012 JP Morgan loses 6BN in the famous Whale Debacle case;

- Knight Capital Group lost 440M overnight due to a small error in their model getting the firm close to insolvency.

* Cathy O’Neil, “Weapons of Math Destruction”,

<code4thought>
As a result: Costs are increasing

- Cost of doing business: Regulations such as TRIM, GDPR impose an administrative burden;

- Cost of opportunity (not innovating): Risk Compliance departments are not allowing the deployment of any state of the art model in a financial institution that cannot be explained/interpreted;

Authority is increasingly expressed algorithmically

“Already today, ‘truth’ is defined by the top results of the Google search.”

Uval Noah Harrari, “21 lessons for the 21st century”
Accountability

The degree of which one decides when and how an algorithmic model should be guided (or restrained) in the risk of crucial or expensive errors, discrimination, unfair denials, or censorship.

Nicholas Diakopoulos, “Accountability in Algorithmic Decision Making.”

The challenges for Algorithmic Accountability

- Current state of the art techniques (e.g. deep learning) can be extremely precise but opaque;
- Organizations push-back: Secrecy and protection of intellectual property as counter-arguments;
- Several initiatives and practices but not yet an industry standard on Accountability.
References/Sources of Inspiration

- Algorithmic Accountability Reporting: On the Investigation of Black Boxes
  NICHOLAS DIAPOLPOULOS, PH.D.

- Outnumbered
  DAVID SUMPTER
  From Facebook and Google to fake news and filter-bubbles - the algorithms that control our lives

- Weapons of Math Destruction
  CATHY O'NEIL
  NEW YORK TIMES BESTSELLER

- Artificial Unintelligence
  MEREDITH BOUSSARD
  HOW COMPUTERS MISUNDERSTAND THE WORLD

- The Black Box Society
  FRANK PASQUALE
  THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION

HBR's 10 Must Reads 2017
The definitive management ideas of the year from Harvard Business Review.
Our principles when dealing with Algorithmic Systems

- Humans should be in-the-loop;
- Data is socially constructed;
- Accuracy versus Interpretability is a false dilemma.
Our approach (1/2): A model agnostic evaluation framework

Organisations (Cater for)
- Responsibility/Human Involvement
- Explainability
- Accuracy
- Auditability
- Fairness

Algorithmic Systems Accountability

Algorithms (Designed, Implemented and Evaluated for)
- Algorithmic Presence
- Data
- Algorithm Input
- Performance Evaluation
- Inferencing
Our approach (2/2): A model agnostic explanation mechanism (Surrogate Tree)
Goals of our Pilot projects (Case Studies)

The last 12 months Code4Thought has been conducting R&D activities in order to create an assessment framework and explanation mechanism for black-box models and develop an MVP.

Clients have accepted to test and validate Code4Thought’s MVP in their newly developed models, which are part of a series of business services. The goals for these projects were:

- For Code4Thought to gather feedback and validate our work so far,
- For Clients to see whether there can be any lessons learnt that they can apply in their models.
Delivery Process and Timeline

- Week n: Kick-Off
- Week n+2: Management/Technical Sessions
- Week n+4: Validation Session
- Week n+6: Final Presentation Session
Scope of the analyses

For the purposes of both case studies the following data were requested, collected and processed:

- Answers to the questionnaires regarding the model’s implementation and Client’s ability to control it,
- Input data for the model and the respective labels as prescribed from Code4Thought (only in case study II).
Case Study I:

Client is a European financial institution with 22,000 employees and a revenue of 2 Billion Euro.

Client was developing a classification algorithm whose goal was to rate the users of their Mobile and Internet Banking platform based on how digitally literate they are.
Research Questions:

The organization was trying to improve its users’ experience by providing advanced functionality to the literate ones and basic to those who are rated as non-literate.

- At what extent does Client follow best-practices regarding:
  - Their selected approach to solve the given problem,
  - The organisation’s maturity to hold their model accountable and be able to control it.
Delivered Results

- Data without labels
- Partial performance evaluation
- The employed technique is not fit for purpose
- No priorities
Case Study*:

Client is a U.S. multinational software company with a multi-billions turnover that provides among other technologies server, virtualization and networking software as a service (SaaS).

Client’s service is employing a Machine Learning model that allows network administrators to monitor and identify Security threats.

*For more details read: Yiannis Kanellopoulos, “Accountability of Algorithmic Systems: How We Can Control What We Can’t Exactly Measure”
https://www.cutter.com/offer/accountability-algorithmic-systems-how-we-can-control-what-we-can't-exactly-measure
Cutter Business Technology Journal, March 2019
Research Questions:

The following research questions were posed from Client to our team:

- At what extent does Client follow best-practices regarding:
  - The transparency of its model behind their Service,
  - The organisation’s maturity to hold their model accountable.

- Can Code4Thought provide a fast, user-friendly, post-hoc explanation of the decisions made by Client’s black-box model?
Delivered Results

Unsupervised model

Not priorities

No annotations
Conclusions

- Accountability and Ethics are an afterthought, not embedded already from the inception of a Machine Learning project;
- Proper model selection is crucial for the success of a Machine Learning project;
- Collecting the data is one thing, creating value out of them is another;
Stay in touch

- **Use:** explain.code4thought.eu
- **Contact:** yiannis@code4thought.eu - https://www.linkedin.com/in/ykanellopoulos/
- **Follow:** @code4thought.eu
...one more thing

Rate today’s session

Session page on conference website

O’Reilly Events App