Standard Models for Explainable ML

Abstract

Explainable Machine Learning (ML) is a property not just of models but of ML's scientific identity. I analyze this scientific identity through a case study of the work of the ACM task force on Data Science and of four seminal articles including Vaswani et. al. 'Attention is all you Need' (Vaswani et al., 2017). I conclude that ML's scientific identity is disparate and reflects a preference for probabilistic and heuristic rather than general statements. I challenge this preference and discuss a number of ways to potentially render ML less disparate and more explainable. I amend the Standard Model for Machine Learning, originally proposed by Hu et al. in 2022, in order to facilitate a structured discussion of accuracy, model desiderata, context and ethical and legal requirements across all ML model categories. I propose a stronger articulation of deductive versus inductive processes when models learn from data. I propose more consistent symbolic generalizations as well as more intuitive methods for teaching ML. These initiatives will promote Explainable ML.

What is a standard model and why do we need it?

- One model that comprises all ML models as special cases
- Why?
  - Words, definitions, notation matter in innovation
  - Confirmed by literature such as Hadamard (Hadamard, 1945, chap. 7, p. 84) and Kuhn (Kuhn, 2012, 1969 postscript, p. 182)
  - Less support for the idea that notation etc. must be standardized
  - The recent focus on Explainable ML makes standard models relevant
    - *"What if ... we would only have to learn and explain one model"*
    - *"What if ... all new models could be explained with reference to the standard model*
- Hu and Xing proposed a Standard Model for Machine Learning (Hu and Xing, 2022, Harvard Data Science Review, 2022)

Standard models = general statements = Explainable ML

- ML as a science is often heuristic and probabilistic
- Standard models can facilitate general statements
  - According to the deductive-nomological model of scientific explanation (Woodward et al., 2021), good explanations refer to back to generally accepted results
- General statements can therefore facilitate Explainable Machine Learning

Explainable ML is a property of ML's scientific identity

- A task force under the Association for Computing Machinery (ACM) worked from 2017 to 2021 on defining data science and ML (ACM, 2021, p.1)
- Emphasized multidisciplinarity (computer science, statistics, mathematics, application domains)
- The word 'broad' appears 32 times over the 132 pages of the final report
- The official curricula for Danish ML programmes have the same 'broad' approach
- Justified from a ressource perspective, but commits the ML sin of "too rich a hypothesis space = overfitting"

As ML's scientific identity matures, will Explainable ML improve?

- Proposal:
  - 'Computer science first': Define a clear hierachy of sciences for ML
  - ML is a proper subset of computer science.
  - The subset that analyzes uncertainty and ambiguity better than traditional computer science

My project

- Does Machine Learning need a Theory of Everything? supervised by Professor, Ph.d. Henrik Kragh Sørensen, 2023

`https://github.com/TimMondorf/Does-Machine-Learning-need-a-Theory-of-Everything`

Tim Mondorf

`cjk681@alumni.ku.dk, +45 42765599`

MSc Economics, student MSc Computer Science, University of Copenhagen (DIKU)

UNIVERSITY OF COPENHAGEN