THE STATE OF MACHINE LEARNING

RSE seminar, University of Sheffield

Tania Allard, PhD

https://trallard.github.io/Talks/RSE-sheffield
TANIA ALLARD

Developer advocate
Research Software Engineer
Data expert

trallard
ixeke
MACHINE LEARNING EVERYWHERE
MACHINE LEARNING EVERYWHERE

So much that it is starting to not make sense anymore... like when you say a word 50 times in a row
For good or for bad it is everywhere:
For good or for bad it is everywhere:

- Deployed in healthcare and warfare
For good or for bad it is everywhere:

- Deployed in healthcare and warfare
- In the creative industry (from music to books)
For good or for bad it is everywhere:

- Deployed in healthcare and warfare
- In the creative industry (from music to books)
- Reading CVs and judging your creditworthiness
For good or for bad it is everywhere:

- Deployed in healthcare and warfare
- In the creative industry (from music to books)
- Reading CVs and judging your creditworthiness
- Making us more Instagram worthy
The big players:

- Apple
- Facebook
- Google
- IBM
- Intel
- Microsoft
- Nvidia
- Open AI
- Twitter
MACHINE LEARNING GENERALISED IN TWO WORKFLOWS

- Model development (R&D)
- Model serving (production for customers consumption)
WHAT ARE THESE GIANTS' ISSUES?
WHAT ARE THESE GIANTS’ ISSUES?

Mainly scale...in multiple areas
If we have a small team we have a smaller number of issues... right?
If we have a small team we have a smaller number of issues... right? 🐒 Small number of models to maintain
If we have a small team we have a smaller number of issues... right?

- Small number of models to maintain
- People have the knowledge in their heads
If we have a small team we have a smaller number of issues... right?

- Small number of models to maintain
- People have the knowledge in their heads
- They have their own methods to track progress
THAT IS THE SMALL TEAM PERFORMANCE FALLACY

We still need processes and best practices in place... so let me get back at this later
AS THE TEAM DEMAND GROWS THE PROBLEMS GROW

- Increased complexity of data flow
- Larger number of workflows
- Managing complexity of flows and scheduling becomes a nightmare
- Resource allocation has to be on point
SERVING MODELS BECOMES HARDER
HOW DO THEY SERVE MILLIONS OF
CUSTOMERS ACROSS THE GLOBE?
Three main players:

- Infrastructure / resources
- Processes
- People
Analytics & Machine Learning

- Snowflake
- Kafka
- Apache Spark
- Amazon EMR
- Amazon Kinesis
- Amazon Redshift
- Elastic
- TensorFlow
- Apache Storm
- Google Cloud Machine Learning
- Amazon Machine Learning

Azure Machine Learning
INFRASTRUCTURE AS A CODE
EVERYTHING AS A CODE

- Version control
- Less ambiguity on the configurations
- Shorter turnarounds
- Deterministic environments
PROCESSES
DATA AND CODE AS FIRST CLASS CITIZENS
PEOPLE

- Data scientist
- Data engineer
- ML Engineer
WHAT DOES ACADEMIA HAVE TO OFFER?

Much more than you think
PEOPLE

- Researchers
- Research software engineers
- Librarians
RESOURCES AND INFRASTRUCTURE

We still need to figure this out... it is pretty much an ad-hoc case
PROCESSES

- Scientific rigour
- Peer review
- Data management
Which areas could benefit from academic collaborations?
META-LEARNING

Humans learn across tasks (learn from experience)
If prior tasks are similar then we can carry prior knowledge
AlphaGo uses some sort of meta-learning
Algorithmic fairness

It has become increasingly important to ensure that models are making justified calls that are free from unintended bias.
ALGORITHMIC FAIRNESS

It has become increasingly important to ensure that models are making justified calls that are free from unintended bias.

The one way to make progress is through interdisciplinary collaboration.
TOWARDS MODEL EXPLAINABILITY

Address the trade-off between performance and interpretability
REINFORCEMENT LEARNING DEADLY TRIAD

Following nature's paradigms RL agents receive awards and then learn to maximise success by performing optimal actions.
How to keep an algorithm learning if there are far too many potential variables or outcomes to be evaluated without being fed ridiculous amounts of data.
IN BRIEF

Focus on the 3 pillars:

- People
- Infrastructure
- Processes
THANK YOU

@ixek
✉️ tania.allard@microsoft.com