On the Role of Domain Knowledge in Explainable Machine Learning

Journées de Statistique de la SFdS 2021
(Virtual Event)
June 11, 2021

Freddy Lecue, Chief AI Scientist (Thales Canada, Inria France)
http://www-sop.inria.fr/members/Freddy.Lecue/
@freddylecue
Critical Applications
AI Adoption: Requirements

Valid AI

Privacy-preserving AI

Responsible AI

What is the rational?

Human Interpretable AI

What is the rational?

Machine Interpretable AI

Explainable AI
Knowledge Graph in Machine Learning (1)

Augmenting (input) features with more semantics such as knowledge graph embeddings / entities

https://stats.stackexchange.com/questions/230581/decision-tree-too-large-to-interpret
Knowledge Graph in Machine Learning (2)

Augmenting machine learning models with more semantics such as knowledge graphs entities

https://stats.stackexchange.com/questions/230581/decision-tree-too-large-to-interpret
Knowledge Graph in Machine Learning (3)

- **Input Layer**: Training Data
  - Neurons respond to simple shapes
- **Hidden Layer**:
  - 1st Layer: Neurons respond to more complex structures
  - 2nd Layer: Neurons respond to highly complex, abstract concepts
- **Output Layer**:
  - Low-level features to high-level features
  - Augmenting (intermediate) features with more semantics such as knowledge graph embeddings / entities
Knowledge Graph in Machine Learning (4)

Jesse Mu, Jacob Andreas: Compositional Explanations of Neurons. NeurIPS 2020

Open question: What is the impact of semantic representation on units in Neural Networks?
Knowledge Graph in Machine Learning (5)

- Input Layer
- Hidden Layer
- Output Layer

Training Data

Neurons respond to simple shapes

Neurons respond to more complex structures

Neurons respond to highly complex, abstract concepts

1st Layer

2nd Layer

nth Layer

Input (unlabeled image)

Low-level features to high-level features

Augmenting (input, intermediate) features – output relationship with more semantics to capture causal relationship
Knowledge Graph in Machine Learning (6)

Description 1: This is an orange train accident

Description 2: This is a train accident between two speed merchant trains of characteristics X43-B and Y33-C in a dry environment

Description 3: This is a public transportation accident

Augmenting models with semantics to support personalized explanation
Knowledge Graph in Machine Learning (7)

“How to explain transfer learning with appropriate knowledge representation?

Augmenting input features and domains with semantics to support interpretable transfer learning.
Knowledge Graph in Machine Learning (8)

"How to explain concept drift in Machine Learning?"

Augmenting input features and domains with semantics to interpret concept drift in Machine Learning.

How Does it Work in Practice?
State of the Art
Machine Learning
Applied to Critical Systems
Object (Obstacle) Detection Task
Object (Obstacle) Detection Task State-of-the-art ML Result

Lumbermill - .59
Object (Obstacle) Detection Task State-of-the-art ML Result

- Boulder - 0.09
- Railway - 0.11
- Lumbermill - 0.59
State of the Art
XAI
Applied to Critical Systems
Object (Obstacle) Detection Task
State-of-the-art XAI Result

Lumbermill - .59
Object (Obstacle) Detection Task
State-of-the-art XAI Result

Lumbermill - .59
Object (Obstacle) Detection Task
State-of-the-art XAI Result

Lumbermill - .59
Unfortunately, this is of NO use for a human behind the system
Let’s stay back

Why this Explanation?
(meta explanation)
After Human Reasoning...

Lumbermill - .59

A sawmill or lumber mill is a facility where logs are cut into lumber. Prior to the invention of the sawmill, boards were rived (split) and planed, or more often sawn by two men with a whipsaw, one above and another in a saw pit below. The earliest known mechanical mill is the Hierapolis sawmill, a Roman water-powered stone mill at Hierapolis, Asia Minor dating back to the 3rd century AD. Other water-powered mills followed and by the 11th century they were widespread in Spain and North Africa, the Middle East and Central Asia, and in the next few centuries, spread across Europe. The circular motion of the wheel was converted to a reciprocating motion at the saw blade. Generally, only the saw was powered, and the logs had to be loaded and moved by hand. An early improvement was the developm (en)
What is missing?

Lumbermill - .59
Context matters

About: Boulder

An Entity of Type: place, from Named Graph: http://dbpedia.org, within Data Space: dbpedia.org

In geology, a boulder is a rock fragment with size greater than 25.6 centimetres (10.1 in) in diameter. Smaller pieces are called cobbles and pebbles, depending on their “grain size”. While a boulder may be small enough to move or roll manually, others are extremely massive. In common usage, a boulder is too large for a person to move. Smaller boulders are usually just called rocks or stones. The word boulder is short for boulder stone, from Middle English boulderston or Swedish boulderen. Boulder sized clasts are found in some sedimentary rocks, such as coarse conglomerate and boulder clay.

Property Value

dw:abstract

- In geology, a boulder is a rock fragment with size greater than 25.6 centimetres (10.1 in) in diameter. Smaller pieces are called cobbles and pebbles, depending on their “grain size”. While a boulder may be small enough to move or roll manually, others are extremely massive. In common usage, a boulder is too large for a person to move. Smaller boulders are usually just called rocks or stones. The word boulder is short for boulder stone, from Middle English boulderston or Swedish boulderen. Boulder sized clasts are found in some sedimentary rocks, such as coarse conglomerate and boulder clay.

About: Rail transport

An Entity of Type: software, from Named Graph: http://dbpedia.org, within Data Space: dbpedia.org

Rail transport is a means of conveyance of passengers and goods on wheeled vehicles running on rails, also known as tracks. It is also commonly referred to as train transport. In contrast to road transport, where vehicles run on a prepared flat surface, rail vehicles (rolling stock) are directionally guided by the tracks on which they run. Tracks usually consist of steel rails, installed on ties (sleepers) and ballast, on which the rolling stock, usually fitted with metal wheels, moves. Other variations are also possible, such as slab track, where the rails are fastened to a concrete foundation resting on a prepared sub-surface. Rolling stock in a rail transport system generally encounters lower frictional resistance than road vehicles, so passenger and freight cars (cargoes and wagons) can be coupled into larger trains. The operation is carried out by a railway company, providing transport between train stations or freight customer facilities. Power is provided by locomotives which either draw electric power from a railway electrification system or produce their own power, usually by diesel engines. Most tracks are accompanied by a signaling system. Railways are a safe and land transport system when compared to other forms of transport. Railway transport is capable of high levels of passenger and cargo utilization and energy efficiency, but is often less flexible and more capital-intensive than road transport, when lower traffic levels are considered. The oldest, man-powered railways date back to the 7th century BC, with Persia, one of the Seven Ages of Greece.
• Hardware: High performance, scalable, generic (to different FGPA family) & portable CNN dedicated programmable processor implemented on an FPGA for real-time embedded inference

☑️ Software: Knowledge graph extension of object detection

This is an **Obstacle: Boulder** obstructing the train: XG142-R on **Rail.Track** from City: Cannes to City: Marseille at **Location: Tunnel VIX** due to **Landslide**
XAI Thales Platform

Higher accuracy with no intensive fine-tuning
Human interpretable explanation
Running on the edge at inference time
Train operating on Rail Track Boulder obstrucing Railway - .90 Boulder - .81 Tunnel - .74 Tunnel obstructing Landslide

Train operating on Rail Track Boulder obstrucing Railway - .90 Boulder - .81 Tunnel - .74 Tunnel obstructing Landslide

Train operating on Rail Track Boulder obstrucing Railway - .90 Boulder - .81 Tunnel - .74 Tunnel obstructing Landslide

Train operating on Rail Track Boulder obstrucing Railway - .90 Boulder - .81 Tunnel - .74 Tunnel obstructing Landslide

Train operating on Rail Track Boulder obstrucing Railway - .90 Boulder - .81 Tunnel - .74 Tunnel obstructing Landslide
Knowledge Graph in Machine Learning - An Implementation


Ecosystem
XAI in Canada: An instance of a Scientific eXplainable AI project

DEEL (Dependable Explainable Learning) Project
2019-2024

With support of

Industrial partners

Academic partners

And 3 aerospace partners

Certificability
➤ Structural warranties
➤ Risk auto evaluation
➤ External audit
Human-in-the-Loop

Journées de Statistique de la SFdS, 2021, June 11th

Freddy Lecue, Chief AI Scientist (Thales Canada, Inria France)

@freddylecue