Thales XAI Platform: Adaptable Explanation of Machine Learning Systems - A Knowledge Graphs Perspective*

Freddy Lécué[1,2], Baptiste Abeloos[1], Jonathan Anctil[1], Manuel Bergeron[1], Damien Dalla-Rosa[1], Simon Corbeil-Letourneau[1], Florian Martet[1], Tanguy Pommellet[1], Laura Salvan[1], Simon Veilleux[1], and Maryam Ziaeefard[1]

[1] CortAIx, Thales, Canada
[2] Inria, France

Abstract. Explanation in Machine Learning systems has been identified to be the main asset to have for large scale deployment of Artificial Intelligence (AI) in critical systems. Explanations could be example-, features-, semantics-based or even counterfactual to potentially action on an AI system; they could be represented in many different ways e.g., textual, graphical, or visual. All representations serve different means, purpose and operators. We built the first-of-its-kind XAI (eXplainable AI) platform for critical systems i.e., Thales XAI Platform which aims at serving explanations through various forms. This paper emphasizes on the semantics-based explanations for Machine Learning systems.

1 Explainable AI in Critical Systems

Motivation: The current hype of Artificial Intelligence (AI) mostly refers to the success of Machine Learning (ML) and its sub-domain of deep learning. However industries operating with critical systems are either highly regulated, or require high level of certification and robustness. Therefore, such industry constraints do limit the adoption of non deterministic and ML systems. Answers to the question of explainability will be intrinsically connected to the adoption of AI in industry at scale. Indeed explanation, which could be used for debugging intelligent systems or deciding to follow a recommendation in real-time, will increase acceptance and (business) user trust. Explainable AI (XAI) is now referring to the core backup for industry to apply AI in products at scale, particularly for industries operating with critical systems.

Focus: Thales XAI Platform is designed to provide explanation for a ML task (classification, regression, object detection, segmentation). Although Thales XAI Platform does provide different levels of explanation e.g., example-based, features-based, counterfactual using textual and visual representations, we emphasis only on the semantics-based explanation through knowledge graphs.

Critical Applications: From adapting a plane trajectory, stopping a train, refitting a boat to reconfiguring a satellite, all are examples of critical situations where explanation is a must-to-have to follow an AI system decision.

2 Why Knowledge Graphs for Explainable AI?

State-of-the-Art Limitations: Most approaches limits explanation of ML systems to features involved in the data and model, or at best to examples, prototypes or counterfactuals. Explanation should go beyond correlation (features importance) and numerical similarity (local explanation).

Opportunity: By expanding and linking initial (training, validation and test) data with entities in knowledge graphs, (i) context is encoded, (ii) connections and relations are exposed, and (iii) inference and causation are natively supported. Knowledge graphs are used for encoding better representation of data, structuring a ML model in a more interpretable way, and adopt a semantic similarity for local (instance-based) and global (model-based) explanation.

3 Thales XAI Platform: A Knowledge Graph Perspective

(Semantic) Perspective: The platform is combining ML and reasoning functionalities to expose a human-like rational as explanation when (i) recognizing an object (in a raw image) of any class in a knowledge graph, (ii) predicting a link in a knowledge graph. Thales XAI Platform is using state-of-the-art Semantic Web tools for enriching input, output (class) data with DBpedia ($4,233,000$ resources) and domain-specific knowledge graphs, usually enterprise knowledge graphs. This is a crucial step for contextualizing training, validation, test data.

Explainable ML Classifications: Starting from raw images, as unstructured data, but with class labels augmented with a domain knowledge graph, Thales XAI Platform relies on existing neural network architectures to build the most appropriate models. All confidence scores of output classes on any input image are updated based on the semantic description of the output classes. For instance, an input classified as a *car* will have a higher overall confidence score in case some properties of *car* in the knowledge graph are retrieved e.g., having *wheels*, being on a *road*. In addition the platform is embedding naturally explanation i.e., properties of the objects retrieved in both the raw data and knowledge graph.

Explainable Relational Learning: Starting from relational data, structured as graph, and augmented with a domain knowledge graph, Thales XAI Platform relies on existing knowledge graph embeddings frameworks to build the most appropriate models. Explanation of any link prediction is retrieved by identifying representative hotspots in the knowledge graph i.e., connected parts of the graphs that negatively impact prediction accuracy when removed.