Interpretable to whom? A role-based model for analyzing interpretable Machine Learning systems

Abraham Jose
ID :5068109, CAP6614
abraham@knights.ucf.edu

1 Summary

The author explores to whom a machine learning system should be interpreted and defines the relationship with an agent associated to a deep learning system for better interpretability. This will also help us to understand how important explainability and transparency are to the agents or to the roles like creators, examiners, operators, executers, data and decision-subjects in a machine learning eco-system.

2 Strengths of the proposal

1. Helps us to better understand and to define interpret and to explain a model based on the role/relationship with the neural network model, it is possible to reduce the complexity and helps better understanding.
2. Through the role of examiner, it is possible to audit or to perform a forensic investigation on the neural network model/ecosystem, ensuring there is no non-compliance with the rights of the data and the decision subjects.

3 Weaknesses of the proposal

1. It is not very clear how they will define the the decision-subject in ecosystem for a GAN system(which can be defined as a separate encoder and decoder neural network with adversarial effects on each other) or a complex multi-neural model.
2. Since the roles are not mutually exclusive, agent lack the flexibility and conflicts with his own role which is not accounted in the ecosystem model for the roles in a deep learning model.

4 Results

The author creates a ecosystem model to fit in all the roles, which can be mutually independent by its nature(but one actual agent can have multiple roles, since they are not mutually exclusive) and it is clearly possible to interpret(define) and to explain the model for each roles. The roles in the ecosystem includes creators, examiners, data-subjects, operators,executors and decision subject.

5 Discussion

The paper discusses how we can treat a model for each roles like creators, examiners, operators, executors, data and decision-subjects within an ecosystem. For each of these roles, the intrepretability, amount of transparency required and explainability varies a lot since it is based on each roles. There can be a lot of issues in defining these roles when multiple roles can be carried out by single entity/agent.