CAUSAL BASE FAIRNESS

A methodology to use counterfactual fairness

Orange is engaged in developing fair AI-based services and causal reasoning is of interest.

Purpose is to give every one a key to a responsible digital world.

Responsibility is to give trustworthy, fairness, transparency, accountability and scalable AI.

Absence of any prejudice towards an individual or a group based on their intrinsic traits in the context of decision making.

AI prediction are based on correlation, but correlation is not causality.

Fairness criteria is based on two factors such as statistical criteria and causal based criteria. Which are further sub categorized such as potential outcome and structural causal model.

Casual-effect and counterfactual reasoning has been continuously in increasing trends based upon the report published NourlPS,2021.

Counterfactual fairness has always been the focus point. It is a model which always predict the same output for it's factual and counterfactual instance.

Counterfactual fairness is graph dependent and the value of variable is dependent on the sensitivity parameter.

Discovering the causal graph is difficult and it is done through causal discovery algorithms such as (PC, GES, Lingam).

One dataset and several graph discovered and each graph shows different set of non-descendant variable and thus two different counterfactually fair model.

There are two experimental environment to overcome these measuring issue such as we use counter factual fairness accuracy.

Counterfactual accuracy is the value of total instances which we predicted to its real counterfactual value.

We set up an experimental environment that automatically improves the counterfactual fairness of a ML model.

CHRISTELE TARNEC

MOHAMMED DANISH MUSTAFA