Reading Response
Mar/16/2021

Kritika Chugh
SUID: 882046659

Trying to Define the fairness?

In the first paper, FPR for married/divorced male and female applicants is 0.70 and 0.55, respectively and TPR = 0.86 for both males and females. This classifier would look small at first but this in turn would favor bad credit score males over females. The definition of fairness in loan approval for males and females is as follows: The assignment of males to negative class or positive class should be equal to females being assigned to negative/positive class. But the finding showed that, there was smaller number of male candidates that were incorrectly assigned to the negative class or in other words larger number of male candidates are incorrectly assigned to the positive class. The result is that it is more likely to assign a good credit score to the males who have a running bad credit score, compared to females and thus it violates the definition of fairness. It makes sense for this problem. We have managed to find a definition of fairness for this loan and credit score assignment problem. This means that the classifier is more likely to assign a good credit score to males who have an actual bad credit score, compared to females and it does not fit the definition of fairness.

What should a good classifier be like? Simply it should be having equal prediction accuracy regardless of the gender or race or age. In this example it was the gender that was unequal for both males and females and thus was not accurate.

There are other such examples of such bias. Such as likes of test-fairness in different ethnic groups and many a times the definition of fairness was taken wrong that led to more backlash. “fairness” was reconceptualized again and again to maximize its social utility and to reduce the disadvantage to those who are on the other side of the bargain.

The clear example of biases could be something when we create two entities the same and only differentiated by gender. Now this would be a good litmus to check the fairness of the algorithm. As they in the first review do not seem to differentiate bases on gender. That’s what they did on one of the papers, but the output classification was not the same. But there are so many problems to fight this. Some might suggest separating the data set but then the other might foresee this as a case of reverse bias. Clearly, the views of the public and the measure of fairness by the professionals are almost always different. I personally feel that if the algorithms create an environment which is automated, give equal probability of selection but lack in making a equal selection from the data set then how good are they?

Final note would be if these definition sway away from the public's perception of fairness then there is little chance for these technologies to thrive. Companies invest a huge chunk of money into the advancement of these algorithms but if they lack in the essence of what can be digested as an ethical product then law might have to intervene and start giving verdicts on the ethicality of these very algorithms.