CSE474/574 Introduction to Machine Learning
Programming Assignment 3
Report

Group number: 14
Group Members:
Vinita Venktesh Chappar
Prajit Krisshna Kumar
Harsh Hemendra Shah

---

1 Abstract

In this project, we have come up with a fairness system similar to that of COMPAS system. The main goal of the project was to implement and choose such an algorithm which will improve the current system in terms of accuracy or cost and will give unbiased results accross various racial groups. We have implemented the various fairness algorithms which are used as a postprocessing methods in various models and compared the results. After careful consideration, we have come up with solution presented in this report.

2 Introduction

COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) is the system commonly used by many states to assess a criminal defendantâĂŹs likelihood of committing a crime. It was developed in 1998 and since then, it has been used to assess more than 1 million criminals. Many people argue that such a system is highly biased towards a certain race than others. ProPublicaâĂŹs study shows that the COMPAS system was highly biased with the white race than the black race. So much so that black convicts were termed "risky" almost twice as the white convicts.

This study was then countered by Northpointe saying that ProPublicaâĂŹs study lacked a thorough analysis and COMPAS system adopted a racial bias even when there was no data of the person being of a certain race. Even so, it is not a bad idea to use risk assessment systems such as COMPAS. The skills of decision making in humanâĂŹs is sometimes not that accurate at all. You would be surprised to know that a study showed parole boards were more likely to let the convict go free if the judges had a meal break just before the verdict![1]

3 Proposed Solution

A machine learning risk analysis system could discover such inconsistencies stated above. That is the reason we have created one such system to replace the COMPAS. We have tried to address the problems of racial bias using a Group fairness approach [2].

Details of the choosen Fairness Market model:

Market Model of Choice: SVM

Algorithm of choice: Equal Opportunity

Secondary optimization criteria: Cost

Cost of the system based on market model:-

Cost on Training data: $-628,702,592
Cost on Test data: $-142,856,824

Accuracy of the system based on market model:-

Accuracy on Training data: 0.6378390911198701

Accuracy on Test data: 0.6497909893172318

Results of postprocessing method enforce_equal_opportunity:

Parameters	African-American	Caucasian	Hispanic	Other
Thresholds	0.11	0.08	0.05	0.03
Accuracy	0.634774882	0.626884036	0.596802842	0.600591716
FPR	0.512542373	0.503229775	0.513595166	0.497560976
FNR	0.250920568	0.243002153	0.245689655	0.248120301
TPR	0.749079432	0.756997847	0.754310345	0.751879699
TNR	0.487457627	0.496770225	0.486404834	0.502439024
Total Accuracy	0.626959829			
Total Cost	$-757,737,810			

4 Stakeholders

The stakeholders in the situation here the COMPAS is used is not just the owners or stockholders of the system. Any person who has an intereset or is concerned and affected by the system is a stakeholder. Hence, the convict, the judges who will use the system to make a decision or the parole officers will also be the stakeholders here.

5 Problems addressed by the system

Mainly, unfair bias arises in machine learning system due to selection, sampling, reporting bias in the data set or bias in the objective function.[3] Even if the data does not explicitly contain biases, it may have some features known as sensitive features.

These features may affect the way the machine learning algorithm treats certain people having those sensitive features. These features/attributes are generaly the data such as gender, sexuality, age etc. Although the machine learning algorithm may not contain biases, it may happen that due to such biases in the data itself, our algorithm might turn towards a certain criteria and increase the unfair bias.

6 Impact of our solution

Our algorithm has a "Group fairness approach". Which means that every group will get an equal opportunity. To achieve equal opportunity, we pick per-group thresholds such that the fraction of low-risk group members is the same across all groups. i.e Equal TPR. This ensures that all the groups are treated equally since every group will have a similar TPR. By using the secondary optimization criteria as cost, we are not only getting the optimal cost but also reducing the false negative rate.

7 Conclusions

As the model name suggests, equal opportunity will provide similar TPR accross all of the racial lines. Along with TPR, we have other metrics as well which gives our choice of postprocessing algorithm a significant weightage:

i. **equal opportunity vs demographic parity:**
The TPR accross demographic parit had a lot of variance accross races. the TPR for African-American was 0.71 while for others it was 0.75. The demographic parity model would have become more biased if similar test data was tested on it.

ii. **equal opportunity vs maximum profit:**
While accuracies for both are quite similar and ofcourse maximum profit will give us the best cost, but the TPRs were astoundingly variant. Ranging from 0.85 to 0.33, the model was tending towards a biased result.

iii. **equal opportunity vs predictive parity:** Both of them had shown promising results but equal opportunity had the upper hand in terms of having the optimal cost.

iv. **equal opportunity vs single threshold:** Here, again the cost of equal opportunity was far better than the single threshold.

References:

[1] https://advances.sciencemag.org/content/4/1/eaao5580

[2] https://www.biostat.wisc.edu/ craven/cs760/lectures/fairness.pdf

[3] Word embeddings quantify 100 years of gender and ethnic stereotypes

[4] https://blog.acolyer.org/2018/05/07/equality-of-opportunity-in-supervised-learning/