Project Proposal

on

Degree of Fairness in Machine Learning Algorithms

MASTERS OF COMPUTER SCIENCE
MACHINE LEARNING

BY

Komal Bodhankar (A20492705) , Mousam Sarkar (A20493354)

Under the guidance of

Dr. Yan Yan

DEPARTMENT OF COMPUTER SCIENCE

Illinois Institute of Technology
Chicago, Illinois, United States of America

Fall 2022 (Aug 2022 - Dec 2022)

INTRODUCTION TO PREVIOUS WORK(S)

Humans tend to be biased in nature. These biases are reflecting in our decision making skills and are impacting the data that is readily available online. Machine learning field, even now, is in an infant stage while deriving predictions from this unreliable data, resulting in unaccountable predictions proposed by different algorithms. In recent years, attention has been focused on how predictive models may be biased - a now overloaded word that, in popular media, has come to mean that the model's performance (however defined) unjustifiably differs along social axes such as race, gender, and class. Therefore impacting various human lives. To address this issue, the measure of fairness and various fairness notions of an algorithm are taken into consideration. The existing research considers Derivable Conditional Fairness Regularizer (DCFR) as a metric in In-processing framework to track the trade off between precision and fairness of major decision making algorithms. Additionally, it addresses the paramount importance of fairness in algorithms for both academic and practical research applications.

PROPOSED PROJECT

We propose to implement the usage of DCFR as one of a metric in pre-processing and post-processing of different Machine Learning algorithms and derive the relation between its precision and fairness with respect to predictions. We demonstrate our goal by utilizing real world datasets and portraying the advantages of fairness notion and DCFR in aforementioned algorithmic domains.

MILESTONES

1. Finalizing appropriate real-world datasets for the proposed problem. (3 days)
2. Study compatibility and finalization of different Machine Learning algorithms with DCFR(2 weeks)
3. Clean and prepare the data as per requirement for the pre-processing stage. (3 days)
4. Different model developments by utilizing DCFR as one of the fairness metrics. (2 weeks)
5. Analyze and portray various observations of developed models in the post-processing stage. (1 week)

REFERENCES

[1] Renzhe Xu, Peng Cui, Kun Kuang. 2020. Algorithmic Decision Making With Conditional Fairness. In proceedings of KDD 2020, Virtual Event, USA.

[2] Junyi Chai, Xiaoqian Wang. 2022. Fairness with Adaptive Weights. In Proceedings of the 39th International Conference on Machine Learning, Baltimore, Maryland, USA, PMLR 162, 2022.

[3] Mariya I. Vasileva, 2020. The Dark Side of Machine Learning Algorithm: How and why they can leverage bias, What can be done to pursue algorithm fairness. In Proceedings of 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 2020.

[4] Shira Mitchell, Eric Potash, Solon Barocas. 2021. Algorithm Fairness: Choices, Assumptions, and Definitions. In proceedings of the Annual Review of Statistics and its Applications, Virtual Event, USA, November 2020.