Dynamic Algorithmic Fairness in Machine Learning

ADAPTIVE RUNTIME FAIRNESS MONITORING FOR CREDIT SCORING DURING ECONOMIC FLUCTUATIONS

AUTHOR

Eigard Alstad
Alstad@student.tudelft.nl

SUPERVISOR

Anna Lukina

1. INTRODUCTION

Background

- Fairness in credit scoring models is crucial as these systems influence individuals access to financial services.
- Periods of economic fluctuations - economic growth or decline - can impact the distribution of input data for credit scoring models.

Research Gap:

- Fairness monitors track the statistics of model decisions to ensure fairness, but assume a static distribution which isn't true during economic fluctuations.

RQ: Can we track the underlying economic state through a credit scoring dataset and use this information to detect fairness violations more quickly?

2. DATA GENERATION

We have generated a synthetic credit scoring dataset with the following attributes, using real-world dataset statistics to ensure realism.

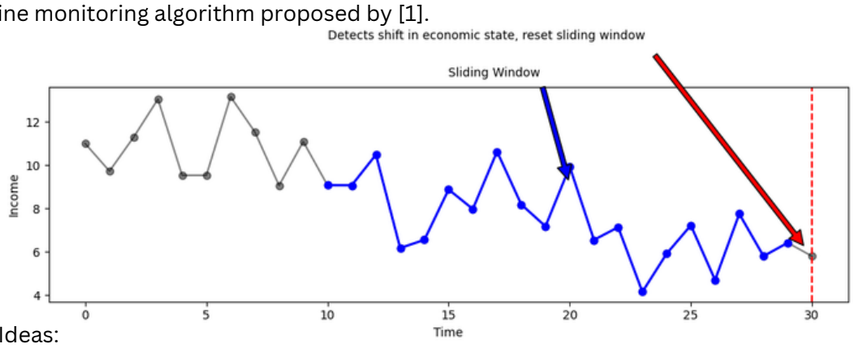
Income (Numerical)	Employment Status (Boolean)	Minority Status (Boolean)	Loan Approval (Boolean)

We simulate economic recessions and booms with adjustments to income and employment status. The final dataset consists of 1900 data instances representing a stable period followed by a recession and subsequent boom.

3. ADAPTIVE FAIRNESS MONITORING ALGORITHM

We focus on one fairness metric, namely **Demographic Parity**, meaning that the probability of loan approval should be independent of minority status.

Our main contribution is an adaptive fairness monitoring algorithm building upon the baseline monitoring algorithm proposed by [1].

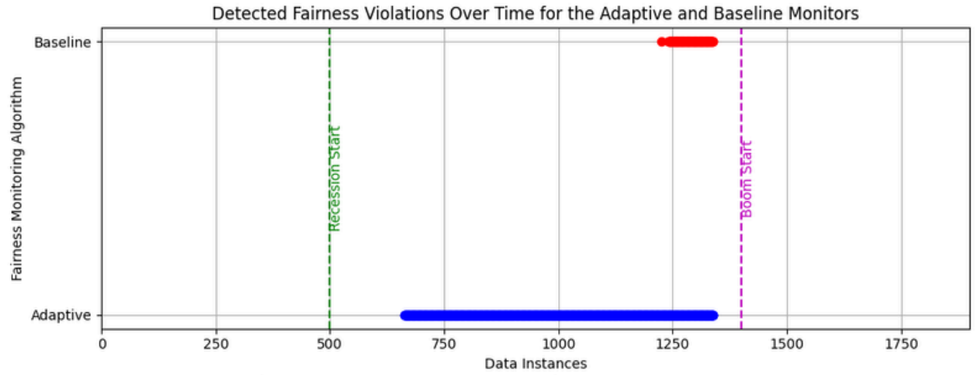


Main Ideas:

- Compute fairness metric over a sliding window instead of all data instances
- Reset the sliding window when economic fluctuations are detected through income distributions, in order to disregard less representative samples.

RESULTS

We compared the fairness violations reported by both our adaptive algorithm and the baseline algorithm that computes fairness metrics over all data instances.



- The plot displays the fairness violations reported from a credit scoring model trained on a dataset representing a stable economy.
- The fairness specification for this experiment was that the probability of loan approval for the minority group be at least 80% of that for the majority group..
- The adaptive algorithm begins to detect unfair behavior approximately 500 data instances earlier than the baseline.
- This preliminary result indicates that the monitor was able to detect economic fluctuations through income distributions and use this to detect fairness violations more quickly.

DISCUSSION AND CONCLUSION

Limitations

- Focused on one fairness metric.
- Did not address computational costs.
- Results based on single synthetic dataset.
- Simplified transitions between distinct economic states.

Conclusion: Incorporating economic state detection into fairness monitoring for credit scoring models can enhance the speed of detecting fairness violations, though further research is needed to address limitations.

Future Work:

- Explore additional fairness metrics.
- Address computational complexity.
- Broader application to other domains.

REFERENCES

[1] - Aws Albarghouthi and Samuel Vinitsky. Fairness-aware programming. In Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* ’19, page 211â219, New York, NY, USA, 2019. Association for Computing Machinery.

TU Delft Delft University of Technology