Effect of Ageing of Datasets in Long Term Fairness

Author: Jorden van Schijndel Supervisor: Anna Lukina

[1]TU Delft

Motivation

- **Fairness** in machine learning models is an extensively researched topic, often done in a static context. However, more recent research[1] shows that it should be monitored in a **dynamic** context.
- **Ageing** of datasets is the effect of older data entries becoming outdated over time, causing a loss in accuracy[4] when not taken into account.
- **Fading** is a dynamic forgetting method to prioritise new data when training a machine learning model to anticipate the ageing of a model.
- Monitoring fairness dynamically raises the issue of ageing of datasets, which has not been properly researched thus far.

Research questions

What is the effect of ageing of datasets in long term fairness?

- How do you measure long term fairness?
- When is an instance no longer important?
- How do you modify a dataset to make newer instances more relevant?
- Is it feasible to modify datasets for fairness in machine learning models?

Methodology

In order to find the influence of the age of a data entry on their relevance, we test multiple fading algorithms against each other and a baseline.

- Static Multinomial Naïve Bayes
  - MNB learns once on old data and is then asked to predict the labels of newer data.
  - Baseline to show importance of dynamic models.
- No Fading
  - MNB learns continuously on new data with a consistent weight
  - Does not prioritize new data over old data. Baseline of the dynamic models.
- Abrupt Fading
  - MNB learns continuously on new data with a consistent weight.
  - Makes use of a sliding window, completely forgets all the data outside of this sliding window. For this research, a sliding window of size 2 was chosen.

$$w_i = \frac{y_i - (y_0 - 1)}{n + 1} \tag{1}$$

- Gradual Weight Fading
  - MNB learns continuously on new data with a distinct weight, see Equation 1.
  - Prioritizes new data by giving it a larger weight and giving older data a lower weight.
- Gradual Amount Fading
  - MNB learns continuously on new data with a consistent weight.
  - Prioritizes new data by training on only a portion of the older data, the data is split by a factor of w, see Equation 1.

The models are ran on the Income prediction task from the **Adult** dataset[2]. They are measured by their accuracy and more importantly, the **Equality of Opportunity**[3].

Results

- Figure 1 shows the performance of each model when given test data over years 2009 to 2013. The lower the equality of opportunity, the fairer a model.
- Table 1 shows the mean and the standard deviation of the Equality of Opportunity.
- All dynamic algorithms significantly outperform the static baseline on the Equality of Opportunity.
- All dynamic models have a slightly lower accuracy than the static model, see Table 2, since the difference is so little, it is deemed negligible.
- All models show a similar trend, this trend is not relevant when comparing the different models.

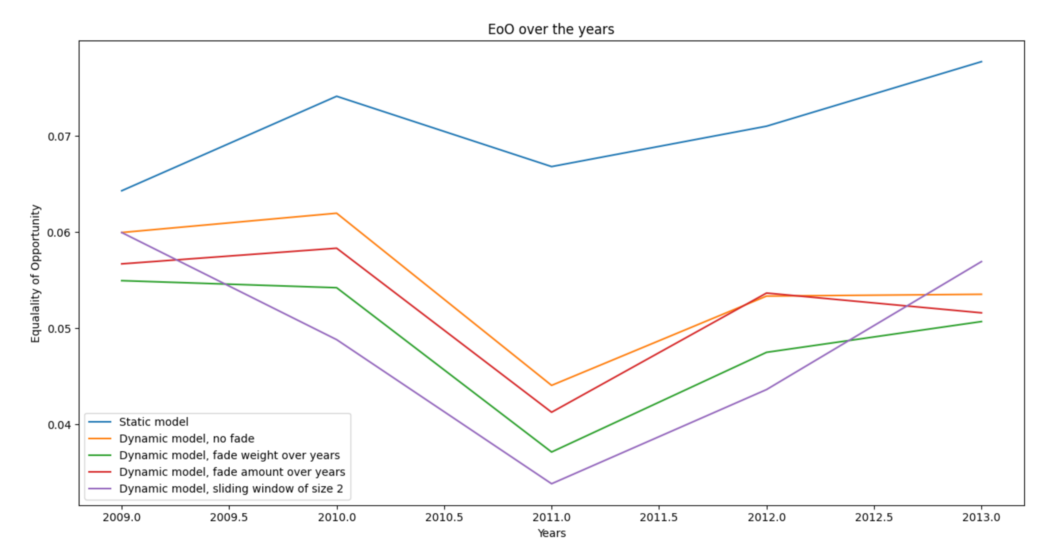

Figure 1. Comparison of EoO the fading algorithms

- **Abrupt fading** performs the best in 2011, but the worst in 2009 and 2013 out of all the dynamic models. This and the high standard deviation in Table 1 show the inconsistency of the model.
- **Gradual Weight Fading** has the lowest mean Equality of Opportunity out of all the models and consistently performs better than most models.
- All dynamic models take longer to train than the static model, especially the gradual weight fading algorithm. This introduces a trade-off between the efficiency of a model and the equality of opportunity.

	Mean EoO	STD EoO
Static	0.069839312	0.008136377
Dynamic no fade	0.049905245	0.00683245
Abrupt	0.046902824	0.011646918
Weight	0.046846232	0.006602121
Amount	0.048661459	0.006823865

Table 1. Mean and standard deviation of the equality of opportunity of the models

	Mean score	STD score
Static	0.67821964	0.006672671
Dynamic no fade	0.673552594	0.006878765
Abrupt	0.6728605	0.005736844
Weight	0.672950157	0.006772536
Amount	0.67333697	0.006847099

Table 2. Mean and standard deviation of the accuracy of the models

Discussion

- To calculate the weight, the total amount of data has to be known. This limits the Gradual algorithms to data in a set range. For actual continuous data, the formula has to be reworked.
- The Adult dataset sorts data in years, and the entries do not contain timestamps. This means more fine-grained weight formulas are possible for other datasets.

Conclusion and Future work

- The results show that when measuring the fairness in the Adult dataset, the ageing of the dataset should be taken into consideration. The dynamic models that prioritise new data obtain a significantly lower equality of opportunity; **A fading algorithm should be used when performing a prediction task on the Adult dataset for fairness.**
- The accuracy of a model is not heavily influenced over time.
- The Gradual Weight Fading algorithm performed the best, however this model takes the longest to train, introducing a **trade-off between efficiency and fairness.**
- One can merge different fading algorithms to find an optimal point in the efficiency-fairness trade-off, e.g. the Abrupt Fading and Gradual Weight Fading algorithms.
- The fading algorithms can be tested on different datasets and on longer time periods, leaving plenty of room for future work.

References

[1] Alexander D'Amour, Hansa Srinivasan, James Atwood, Pallavi Baljekar, D. Sculley, and Yoni Halpern.
Fairness is not static: deeper understanding of long term fairness via simulation studies.
In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, page 525–534, New York, NY, USA, 2020. Association for Computing Machinery.

[2] Frances Ding, Moritz Hardt, John Miller, and Ludwig Schmidt.
Retiring adult: New datasets for fair machine learning.
Advances in Neural Information Processing Systems, 34, 2021.

[3] Moritz Hardt, Eric Price, and Nathan Srebro.
Equality of opportunity in supervised learning, 2016.

[4] Sebastian Wagner, Max Zimmermann, Eirini Ntoutsi, and Myra Spiliopoulou.
Ageing-based multinomial naive bayes classifiers over opinionated data streams.
volume 9284, 11 2015.

j.j.vanschijndel@student.tudelft.nl