Choosing Aggregation Levels for Forecasting and Fairness

Jade Zhang, Lanqing Du, Jinwook Lee, Matthew J. Schneider

Decision Sciences and MIS, Drexel University

July 12, 2022

Outline

1. Introduction
2. Methodology - Sequential Aggregation
3. Application - Comparison between Standard Aggregation and Sequential Aggregation
4. Conclusions

Outline

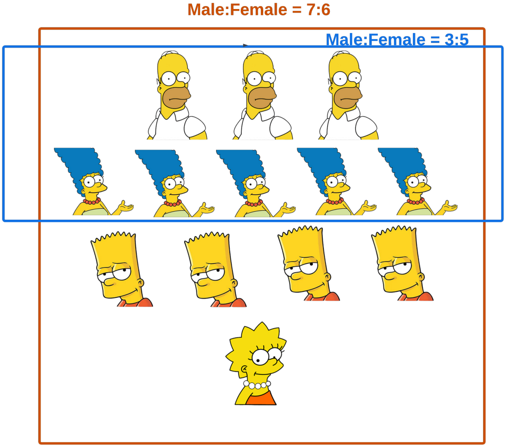
1. Introduction
2. Methodology - Sequential Aggregation
3. Application - Comparison between Standard Aggregation and Sequential Aggregation
4. Conclusions

Optimizing Forecast Accuracy Does Not Optimize Fairness

- Definition of Fairness: the absence of any prejudice or favoritism toward an individual or group based on their inherent or acquired characteristics, an unfair algorithm is one whose decisions are skewed toward a particular group of people (Mehrabi et al., 2021)
- Forecasting with fairness
  - Training the algorithms for the best forecast accuracy may lead to fairness issues (Loukina et al., 2019)
- Enrollment forecasting
  - Forecast at the department, college, and university level for gender/race/...
  - Forecasts are used to plan scholarships, budgets, tuition, etc.

Data Availability Can Change Fairness

- Data aggregation changes fairness metrics
- Different methods of aggregation would lead to different results
  - Simpson's paradox



Literature Review

- Fairness metrics at group level
  - Group fairness notions for binary outcomes in machine learning (Fu et al., 2020)
  - Generalized form for conditional disparity: $\mathbb{P}(x|a = a, z = z) = \mathbb{P}(x|a = a', z = z)$ (Ritov et al., 2015)
- Social good metrics
  - "Presenting both a forecast of a phenomenon and its accuracy alongside the FSG metrics is important" (Rostami-Tabar et al., 2022)
- Choosing aggregation level for hierarchical forecasting
  - Based on properties & nature of data (e.g., aggregate neighborhoods by their locations (Humeau et al., 2013))
  - Based on other metrics or similarity (e.g., hierarchical agglomerative clustering based on Gower's distance (Goehry et al., 2017)
  - Forecast accuracy is improved when defining hierarchical structure based on similarity (Quilumba et al., 2015)

Contribution

- Fairness metrics for time-series data
- Choosing aggregation levels based on fairness (sequential aggregation)

  How does choosing aggregation levels based on fairness affect forecast accuracy?

Data

Annual undergraduate student enrollment data from the University of Cailfornia (UC) Davis from 2010 to 2019

- University of California, Disaggregated Data (2021) (`https://www.universityofcalifornia.edu/about-us/information-center/disaggregated-data`)
- Only contains information at university level (missing college level and department level).
- We choose three attributes: race/ethnicity, gender, first generation status
    - Race/ethnicity: 73 races, categorized into 7 broad categories
    - Gender: male/female
    - First generation status: yes/no
    - 292 time series at the most disaggregated level

Outline

Measuring Fairness without Selection Bias

1. Define pseudo-Boolean conditional parity to measure fairness between X_i and X_j for $1 \le i < j \le N$ based on

$$f_{(x_i,x_j|\epsilon)}(\mathbf{y}) = \sum_{k=1}^{n} c_k y_k + \sum_{1\le k<l\le n} c_{kl} y_k y_l, \tag{1}$$

$$\text{where } y_k = h_{ij}(\epsilon_k) = \begin{cases} 1 & \text{if } \left|P(A_k \mid X_i) - P(A_k \mid X_j)\right| \le \epsilon_k \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{B}^n$ and $c_S \in \mathbb{R}, S \subseteq V, V = [n]$.

2. Define total fairness: average $f(\mathbf{y})$ of all pairs

$$\tau = \frac{\sum_{1\le i<j\le N} f_{(x_i,x_j|\epsilon)}(\mathbf{y})}{\binom{N}{2}} \tag{3}$$

where $f_{(x_i,x_j|\epsilon)}(\mathbf{y})$ as defined in Equation (1), with $c_S = \frac{1}{|S|}$.

Sequential Aggregation

3. Consider w_{ij} the weight of the edge between elements of pairs, which is a linear combination of both fairness f_{ij} and similarity g_{ij} based on Euclidean distance and correlation

$$w_{ij} = \alpha f_{ij} + (1 - \alpha) g_{ij}, \tag{4}$$

4. Solve the perfect matching problem to find optimal pairs of time series to aggregate

$$\begin{aligned} &\max \textstyle\sum_{v_1, v_2} w_{v_1 v_2} x_{v_1 v_2} \\ &\text{subject to} \\ &\sum_{v_1 \in V_1} x_{v_1 v_2} = 1 \\ &\sum_{v_2 \in V_2} x_{v_1 v_2} = 1 \\ &x_{v_1 v_2} \in \{0, 1\}, \end{aligned} \tag{5}$$

5. Repeat the process until we get the desired level of fairness

Outline

Race Aggregation

- **Standard Aggregation**: based on common sense/characteristics
  - 7 broad categories
    - Black: Caribbean, African...
    - Asian: Chinese, Japanese, Korean...
- **Sequential Aggregation**: flexible and numerous ways to aggregate race
  - Keep aggregating until we get 9 broad categories
  - Select multiple sets of parameters to generate different aggregation results

Example of Sequential Aggregation

Label all races as 1, 2, 3,...,73, and set $\epsilon = 0.01, \alpha = 0.5$ for sequential aggregation

	Standard	Seq Agg
Group 1	1 - 5	66, 45, 43, 11, 38, 14, 65, 1
Group 2	6	72, 29, 57, 17, 67, 39, 9, 2
Group 3	7 - 27	26, 13, 47, 12, 51, 46, 62, 3
Group 4	28 - 33	58, 36, 56, 16, 20, 6, 61, 4
Group 5	34 - 40	34, 32, 68, 27, 60, 35, 52, 5
Group 6	41 - 71	37, 24, 42, 10, 18, 53, 31, 44, 7
Group 7	72, 73	48, 30, 41, 23, 59, 21, 28, 8
Group 8		69, 63, 71, 18, 70, 25, 40, 15
Group 9		64, 50, 49, 22, 54, 33, 55, 19

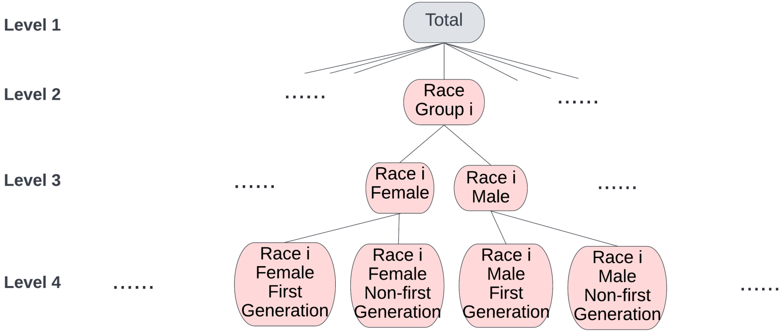
Sequential Aggregation Specifications

Aggregation Outcome		
	ϵ	α
Sequential Aggregation	0.01	0.5
Sequential Aggregation	0.05	0.5
Sequential Aggregation	0.1	0.5

1. α: $\alpha = 0.5$ represents the same weight on fairness as on similarity
2. ϵ: higher ϵ represents greater difference are allowed between groups

Hierarchical Forecasting with Complete Information

- Forecast number of students enrolled in 2019 based on data from 2010 to 2018 using simple exponential smoothing
- Assume data is available at every level, and the forecast is performed for each level



Forecast Accuracy of Fairness at Time $t + 1$

Use simple exponential smoothing to forecast $\widehat{fairness}_{t+1}$ based on historical fairness

$$f_{(x_i,x_j|\epsilon)}(\mathbf{y}) = \sum_{k=1}^{n} 1 y_k + \sum_{1 \le k < l \le n} 0 y_k y_l$$

$$\tau = \frac{\sum_{1 \le i < j \le N} f_{(x_i,x_j|\epsilon)}(\mathbf{y})}{\binom{N}{2}}$$

ϵ	Aggregation	In-sample τ	Forecasted $\hat{\tau}$	$\lvert\hat{\tau} - \tau\rvert$
Level 4 (first-generation status)				
	Standard	0.231	0.137	0.094
0.01	Seq Agg	0.281	0.281	0.000
0.05	Seq Agg	0.281	0.281	0.000
0.1	Seq Agg	0.288	0.287	0.001

Forecast Accuracy of Fairness at Time $t+1$

ϵ	Aggregation	In-sample τ	Forecasted $\hat{\tau}$	$\lvert\hat{\tau}-\tau\rvert$
Level 3 (gender)				
	Standard	0.810	0.477	0.333
0.01	Seq Agg	1.000	1.000	0.000
0.05	Seq Agg	1.000	0.944	0.056
0.1	Seq Agg	1.000	0.944	0.056
Level 2 (race)				
	Standard	0.333	0.194	0.139
0.01	Seq Agg	0.583	0.583	0.000
0.05	Seq Agg	0.588	0.528	0.060
0.1	Seq Agg	0.611	0.611	0.000

Forecast Accuracy of Enrollment Counts at Time $t + 1$

$$AvgMASE = \frac{1}{N} \Big| \frac{\hat{\epsilon}^2_{i,t}}{\frac{1}{T}\sum_t^T = 2|y_t - y_{t-1}|} \Big|,$$

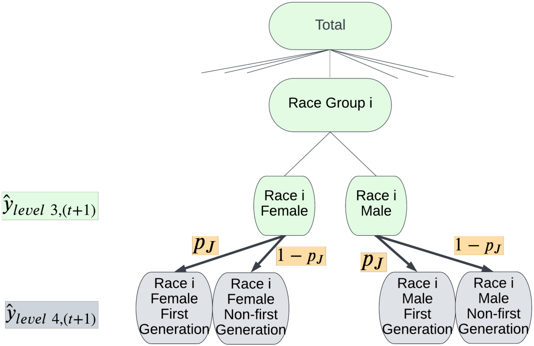
where $\hat{\epsilon}^2_{i,t}$ is Mean Squared Error

Average Mean Absolute Scaled Error				
ϵ	Aggregation	Level 4	Level 3	Level 2
	Standard	0.93	0.56	0.48
0.01	Seq Agg	0.83	0.61	0.67
0.05	Seq Agg	0.75	0.58	0.53
0.10	Seq Agg	0.75	0.50	0.38

Top-down Forecasting without Level 4 Data

Assume the race information and the disaggregated gender information are available, while **only the proportion p_J of first-generation status in the total counts is known**.

- Average Historical Proportions (AHP): $p_J = \frac{1}{T}\sum_{t=1}^{T}\frac{y_{J,t}}{y_t}$
- proportions of the historical averages (PHA): $p_J = \sum_{t=1}^{T}\frac{y_{J,t}}{T} / \sum_{t=1}^{T}\frac{y_t}{T}$



Forecast Accuracy of Fairness at Time $t+1$

$$\widehat{fairness}_{t+1} = f(\widehat{forecast}_{t+1})$$

Only included results from AHP since AHP and PHA provides almost the same results regarding accuracy of forecasted fairness.

ϵ	Aggregation	In-sample τ	Forecasted $\hat{\tau}$	$\lvert\hat{\tau} - \tau\rvert$
Level 4 (first-generation status)				
	Standard	0.231	0.000	0.231
0.01	Seq Agg	0.281	0.000	0.281
0.05	Seq Agg	0.281	0.000	0.281
0.1	Seq Agg	0.288	0.000	0.288

Forecast Accuracy of Enrollment Counts at Time $t + 1$

Average Mean Absolute Scaled Error			
Level 4 (first-generation status)			
ϵ	Aggregation	AHP	PHA
	Standard	5.30	5.30
0.01	Seq Agg	4.05	4.04
0.05	Seq Agg	5.03	5.03
0.1	Seq Agg	4.32	4.32

Outline

Conclusions

- Sequential aggregation can achieve similar or even better forecast accuracy
- Sequential aggregation performs comparatively well at the levels with information loss
- Sequential aggregation achieves high forecast accuracy of fairness with complete information

Future Work

- Explore the theoretical relationships between sequential aggregation and hierarchical forecasting using similarity
- Apply methodology to more sensitive data, such as criminal justice of recidivism studies

Thank you! Questions or Comments?

References I

Runshan Fu, Yan Huang, and Param Vir Singh.
Artificial intelligence and algorithmic bias: Source, detection, mitigation, and implications.
In *Pushing the Boundaries: Frontiers in Impactful OR/OM Research*, pages 39–63. INFORMS, 2020.

Benjamin Goehry, Yannig Goude, Pascal Massart, and Jean-Michel Poggi.
Aggregation of multi-scale experts for bottom-up load forecasting.
IEEE Transactions on Smart Grid, 11(3):1895–1904, 2019.

References II

Samuel Humeau, Tri Kurniawan Wijaya, Matteo Vasirani, and Karl Aberer.
Electricity load forecasting for residential customers: Exploiting aggregation and correlation between households.
In *2013 Sustainable internet and ICT for sustainability (SustainIT)*, pages 1–6. IEEE, 2013.

Anastassia Loukina, Nitin Madnani, and Klaus Zechner.
The many dimensions of algorithmic fairness in educational applications.
In *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 1–10, 2019.

References III

Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan.
A survey on bias and fairness in machine learning.
ACM Computing Surveys (CSUR), 54(6):1–35, 2021.

Franklin L Quilumba, Wei-Jen Lee, Heng Huang, David Y Wang, and Robert L Szabados.
Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities.
IEEE Transactions on Smart Grid, 6(2):911–918, 2014.

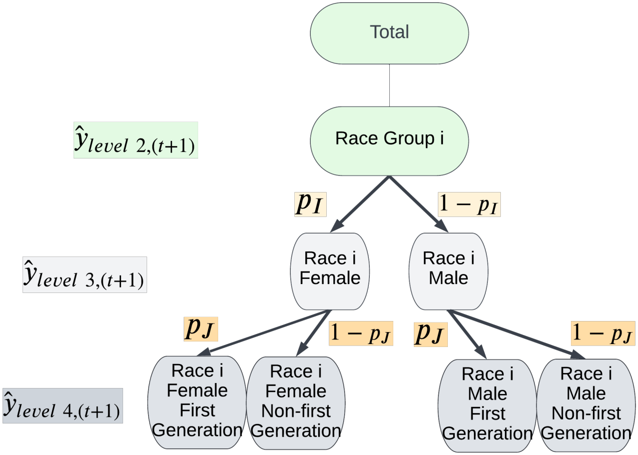
Ya'acov Ritov, Yuekai Sun, and Ruofei Zhao.
On conditional parity as a notion of non-discrimination in machine learning.
arXiv preprint arXiv:1706.08519, 2017.

References IV

Bahman Rostami-Tabar, Mohammad M Ali, Tao Hong, Rob J Hyndman, Michael D Porter, and Aris Syntetos.
Forecasting for social good.
International Journal of Forecasting, 38(3):1245–1257, 2022.

Top-Down Forecasting without Level 3 and Level 4 Data

Assume the race information is available, while **only the proportions of first-generation status p_J and gender p_I in the total counts are known.**



Forecast Accuracy of Fairness at Time $t+1$

ϵ	Aggregation	In-sample τ	Forecasted $\hat{\tau}$	$\lvert\hat{\tau}-\tau\rvert$
Level 4 (first-generation status)				
	Standard	0.231	0.000	0.231
0.01	Seq Agg	0.281	0.000	0.281
0.05	Seq Agg	0.281	0.000	0.281
0.1	Seq Agg	0.288	0.000	0.287
Level 3 (gender)				
	Standard	0.810	0.143	0.667
0.01	Seq Agg	1.000	0.028	0.972
0.05	Seq Agg	1.000	0.222	0.778
0.1	Seq Agg	1.000	0.028	0.972

Forecast Accuracy of Enrollment Counts at Time $t+1$

Average Mean Absolute Scaled Error			
Level 4 (first-generation status)			
ϵ	Aggregation	AHP	PHA
	Standard	5.59	5.57
0.01	Seq Agg	4.28	4.26
0.05	Seq Agg	5.21	5.19
0.1	Seq Agg	4.45	4.44
Level 3 (gender)			
	Standard	1.82	1.78
0.01	Seq Agg	1.51	1.44
0.05	Seq Agg	1.34	1.28
0.1	Seq Agg	1.50	1.42