Development of an Equity Strategy for Recommendation Systems

Rafael V. M. Santos[1], **Giovanni V. Comarela**[2]

[1]Departamento de Informática – Instituto Federal do Espírito Santo (IFES)
Caixa Postal 15.064 – 91.501-970 – Cachoeiro de Itapemirim – ES – Brazil

[2]Departamento de Informática – Universidade Federal do Espírito Santo (UFES)
Vitória, ES.

rafaelv@ifes.edu.br, gc@inf.ufes.br

Abstract. *As a highly data-driven application, recommender systems can be affected by data distortions, culminating in unfair results for different groups of data, which can be a reason to affect system performance. Therefore, it is important to identify and resolve issues of unfairness in referral scenarios. We therefore developed an equity algorithm aimed at reducing group injustice in recommender systems. The algorithm was tested on two existing datasets (MovieLens and Songs) with two user clustering strategies. We were able to reduce group unfairness in both data sets by considering the two clustering strategies.*

1. Introduction

Recommendation systems are crucial to various online platforms that exert significant influence over the choices we make in our everyday lives. From social media platforms such as Facebook and Twitter to streaming services such as Netflix and transportation apps such as Uber, these systems shape our preferences and decisions. However, as our reliance on these systems increases, it is important to consider potential inadvertent social harms that may arise.

Computational models are not free from bias, as they are built and elaborated based on human reasoning and decisions [Taso et al. 2023]. These models may present biases and privilege certain groups over other groups [Ruback et al. 2021]. Therefore, due to non-neutrality they can make discriminatory decisions [Niemiec et al. 2022].

In recent studies, it has been highlighted how recommendation systems, by predicting user preferences, can unintentionally perpetuate inequalities and injustices. For instance, [Wang et al. 2023] and [Tang et al. 2023] indicate the possibility of such systems offering unfair or unequal quality of service to certain individuals or user groups. Furthermore, it is important to emphasize that these systems can also contribute to social polarization, widening the divergence between individual or user group preferences, as demonstrated by [Cinus et al. 2022].

In job platforms, algorithms can develop gender biases, favoring male candidates over female candidates [Kumar et al. 2023]. In online education, courses from more developed regions may be more prestigious in recommendations, perpetuating regional imbalances in access to quality education [Gómez et al. 2021]. On music streaming platforms, promoting popular artists can limit the visibility of lesser-known talent, perpetuating the success cycle for established artists [Mehrotra et al. 2018]. Recommendation systems in e-commerce can reinforce consumption inequalities, with higher-spending customers receiving more exclusive options [Li et al. 2021].

These examples emphasize the importance of reflecting on the social impacts of recommendation algorithms, highlighting the need for equity and fairness in these systems. Relevant works include [Barocas and Selbst 2016a] on algorithmic bias and the study by [Sweeney 2013] on racial discrimination in online ads.

Recommendation accuracy is often used as a metric to evaluate the performance of a recommendation algorithm-how well it can predict whether a user may like an item or not, i.e., its utility. However, the issue of user fairness arises when it is necessary to consider the unequal effects of recommendations on certain groups.

In this article, we introduce an algorithm that incorporates metrics to capture injustice in recommendation systems, as well as a strategy to reduce group unfairness. Furthermore, we examine the relationship between improvements in socially relevant measures and changes in the overall system accuracy.

This article is divided into four additional sections: In Section 2, we present the results of a literature review; in Section 3, we describe the materials used, datasets, the proposed approach, and the experimental methodology; in Section 4, we present and discuss the obtained results; and finally, in Section 5, we conclude the work and outline possible directions for future research.

2. Related Work

We begin this section by providing definitions of the justice concepts to be applied in this project.

Justice is a topic of growing interest in the field of machine learning. After the discussion in this section, we consider a recommendation system fair if it provides equal quality of service (i.e., prediction accuracy) to all users or user groups [Zafar et al. 2017].

Next, we address how our fairness measures for recommendation systems relate to those presented in previous research.

Justice in Machine Learning and Recommendation Systems: In recent years, there has been increasing awareness of the potential social harms caused by the use of machine learning algorithms in decision-making scenarios [Barocas and Selbst 2016b, Boyd and Crawford 2012]. In response, researchers have proposed various notions and metrics of fairness for machine learning tasks, including classification [Hardt et al. 2016, Bilal Zafar and Gummadi 2017, Zafar et al. 2017, Zemel et al. 2013], regression [Berk et al. 2017], ranking [Biega et al. 2018, Wang and Gong 2018, Zehlike et al. 2022], and set selection [Celis et al. 2016]. These proposals can be grouped into two main categories: those that measure fairness at the level of individual users and those that measure fairness at the level of user groups [Dwork et al. 2011].

Relative to the research on learning tasks such as classification and regression, few researchers have explored notions of fairness in the context of recommendation systems. Recently, Burke et al. [Burke et al. 2018] noted that recommendation systems that predict user preferences for items should consider fairness from both sides: the perspective of users receiving recommendations and the perspective of items being recommended. Some of the early works by Kamishima et al. [Kamishima and Akaho 2017, Kamishima et al. 2012, Kamishima et al. 2018] focused on notions of group-level fairness, modifying the learning model to ensure that item recommendations were independent of user characteristics such as race and gender. More recently, Beutel et al. [Beutel et al. 2017] and Yao et al. [Yao and Huang 2017] defined notions of group-level fairness in recommendation systems based on the prediction accuracy across different user or item groupings.

Innovative Contributions: Unlike conventional strategies needing continuous adjustments [Kamishima et al. 2012, Burke et al. 2018], our approach avoids direct recommendation algorithm modifications for each fairness principle. Additionally, unlike [Rastegarpanah et al. 2018], which suggests preprocessing, we offer post-processing.

3. Materials and Methods

In this section, the methodology adopted in the computational experiments reported in this study is detailed.

3.1. Database

Case Study 1 used the MovieLens 1M dataset[1], which contains approximately 1 million ratings of approximately 4000 movies made by approximately 6000 users, with ratings on a 5-point scale [Harper and Konstan 2015]. We filtered the top 300 users with the most ratings, along with the top 1000 most rated movies.

In Case Study 2, we used the Songs dataset[2], which contains approximately 16000 ratings of approximately 19993 songs made by 16000 users, with ratings on a 5-point scale. We also filtered the top 300 users with the most ratings, along with the top 1000 most rated songs.

In both studies, after the predictions were calculated by the recommendation algorithm, we performed two types of user clustering: hierarchical clustering analysis and 95-5 clustering analysis. In the latter, we considered the number of ratings made by users.

We also considered a recommendation system that estimates unknown ratings by solving the matrix factorization problem. The alternating least squares algorithm [Hardt 2013, Hastie et al. 2014] was used to find the factors.

The module 1 of the algorithm, detailed in section 3.2, calculates the social justice measures in the proposed case studies. Module 2 of the algorithm, as described in section 3.3, was employed to calculate a recommendation matrix that minimizes group unfairness, i.e., that maximizes group fairness for social measures in recommendation systems.

Finally, the results of group justice measures and recommendation accuracy were reported. We compared these measures by considering the estimated matrix calculated by a traditional recommendation system $\hat{X}$ with the estimated matrix calculated by the equity algorithm $\hat{X}_{\pi}$.

[1] https://github.com/ravarmes/recsys-algorithm-impartiality/tree/main/data/MovieLens-1M
[2] https://github.com/ravarmes/recsys-algorithm-impartiality/tree/main/data/Songs

3.2. Module 1 of the Algorithm: Calculation of Social Measures

In light of all the specifications and discussions from the previous section, we formally define the metrics that specify the objective functions associated with individual justice and group justice. It is pertinent to mention that all implementations of the fairness measures used in the proposed equity algorithm were based on the work of [Rastegarpanah et al. 2018], providing a solid foundation for our approach to dealing with social justice in recommendation systems.

We start by presenting the system configuration, the notation, and the problem definition. Let us suppose that $X \in \mathbb{R}^{n \times m}$ is a partially observed rating matrix of n users and m items, where the element x_{ij} denotes the rating given by user i to item j. Let Ω be the set of indices of known ratings in X. Furthermore, let Ω_i denote the indices of known item ratings for user i, and let Ω_j denote the indices of known user ratings for item j.

For a matrix A, $P_\Omega(A)$ is a matrix whose elements at $(i, j) \in \Omega$ are a_{ij}, and zeros elsewhere. Similarly, for a vector a, $P_{\Omega_j}(a)$ is a vector whose elements at $i \in \Omega_j$ are the corresponding elements of a, and zeros elsewhere. Throughout the paper, we denote the j-th column of A by the vector a_j and the i-th row of A by the vector a^i. All vectors are column vectors.

Given a traditional recommendation system, an estimated matrix of recommendations $\hat{X} = [\hat{X}_{ij}]_{n \times m}$ is generated. In this recommendation problem, we assume users in a set $\{u_1, u_2 ..., u_n\}$ and items in a set $\{v_1, v_2 ..., v_m\}$.

Individual Justice. For each user i, we define ℓ_i, the user loss for i, as the estimate of the mean squared error over the known ratings of user i. Individual unfairness R_{indv} as the variation of user losses. To enhance individual justice, we aim to minimize R_{indv}.

$$\ell_i = \frac{||P_{\Omega^i}(\hat{x}^i - x^i)||_2^2}{|\Omega^i|} \qquad R_{indv}(X, \hat{X}) = \frac{1}{n^2} \sum_{k=1}^{n} \sum_{l>k} (\ell_k - \ell_l)^2 \tag{1}$$

Group Justice. Let I be the set of all users/items and $G = \{G_1, G_2, ..., G_g\}$ be a partition of users/items into g groups, i.e., $I = U_{i \in \{1,2,...,g\}} G_i$. We define the group loss as the estimate of the mean squared error over all known ratings in group i. For a given partition G, the unfairness of the group R_{grp} is the variation of all group losses. Again, to improve group justice, we minimize R_{grp}.

$$L_i = \frac{||P_{\Omega_{G_i}}(\hat{X} - X)||_2^2}{|\Omega_{G_i}|} \qquad R_{grp}(X, \hat{X}, G) = \frac{1}{g^2} \sum_{k=1}^{g} \sum_{l>k} (L_k - L_l)^2 \tag{2}$$

3.3. Module 2 of the Algorithm: Social Equity Algorithm

In this module, we provide a framework that can generate fairness-aware recommendations based on a reranking method with fairness constraints.

Therefore, considering a traditional recommendation system that generates an estimated recommendation matrix $\hat{X}$, each user u_i receives a set of recommendations

$\{v_1, v_2..., v_m|u_i\}$, which, from the perspective of equity calculation, represents an individual loss ℓ_i.

We use the estimated matrix $\hat{X}$ to generate h other estimated matrices $\hat{X}_1$, $\hat{X}_2$..., $\hat{X}_h$. These h estimated matrices are generated with random variations, bounded within $-\ell_i/4$ and $+\ell_i/4$, for each value of $\hat{x}^i$ corresponding to user i. This perturbation applies not only to known ratings but also to all estimated recommendations.

The new values of cells $\hat{x}^i$ in each of the estimated matrices $\hat{X}_p$ can consider a perturbation strategy based on the variance of rating differences versus recommendations. In this context, we set a maximum variance of 16 $(5-1)^2$, as the largest difference between an actual and recommended value can be 4. For instance, we can consider an actual rating of 1 for a specific item compared to a recommendation for the same item calculated at a value of 5. Thus, we normalize the random recommendation value by dividing it by four times the individual unfairness ℓ_i.

We apply the reranking algorithm to choose n rows $\{v_1, v_2..., v_m|u_i\}$, generating a single estimated matrix $\hat{X}_\pi$.

For each estimated matrix, we calculate n individual losses (ℓ_i), corresponding to the n users. Therefore, for each estimated matrix $\hat{X}_p$, where $\{1 \leq p \leq h\}$, we have a list of n individual losses $\{\ell_1, \ell_2..., \ell_n|\hat{X}_p\}$.

We define the matrix of individual losses $Z = [Z_{ij}]_{n\times h}$ to represent the n individual losses calculated for each of the h estimated matrices $\hat{X}_p$, where $Z_{ij} \in \{\mathbb{R}_+\}$, and $\{1 \leq i \leq n\}$, and $\{1 \leq j \leq h\}$, index users and estimated matrices, respectively.

We define the binary matrix $W = [W_{ij}]_{n\times h}$ to indicate whether individual loss j is considered for a user i in forming the final estimated matrix $\hat{X}_\pi$, where $W_{ij} \in \{0, 1\}$, $\{1 \leq i \leq n\}$, and $\{1 \leq j \leq h\}$ index users and individual losses, respectively. Specifically, if individual loss j is considered for user i, then $W_{ij} = 1$; otherwise, $W_{ij} = 0$.

We apply the re-ranking algorithm to select n rows $\{v_1, v_2..., v_m|u_i\}$, generating a single estimated matrix $\hat{X}_\pi$. In this algorithm, we aim to minimize the sum of individual loss scores under equity constraints. Thus, we formulate the optimization procedure for the fairness-aware recommendation problem:

$$R_{grp}(X, \hat{X}, G) = \frac{1}{g}\sum_{k=1}^{g}(L_k - \mu)^2 \qquad R_{grp}^{min} = \frac{1}{g}\sum_{k=1}^{g}(L_k - \mu)^2 \tag{3}$$

where:

$$\ell_i = \sum_{i=1}^{n}\sum_{j=1}^{h} W_{ij}Z_{ij} \qquad L_k = \frac{1}{|\Omega_{G_i}|}\sum_{i=1}^{|\Omega_{G_i}|}\ell_{G_i} \tag{4}$$

A general idea of the social equity algorithm, Algorithm 1, can be visualized in Figure 1. The source code of this project is also available [3].

[3]https://github.com/ravarmes/recsys-algorithm-impartiality

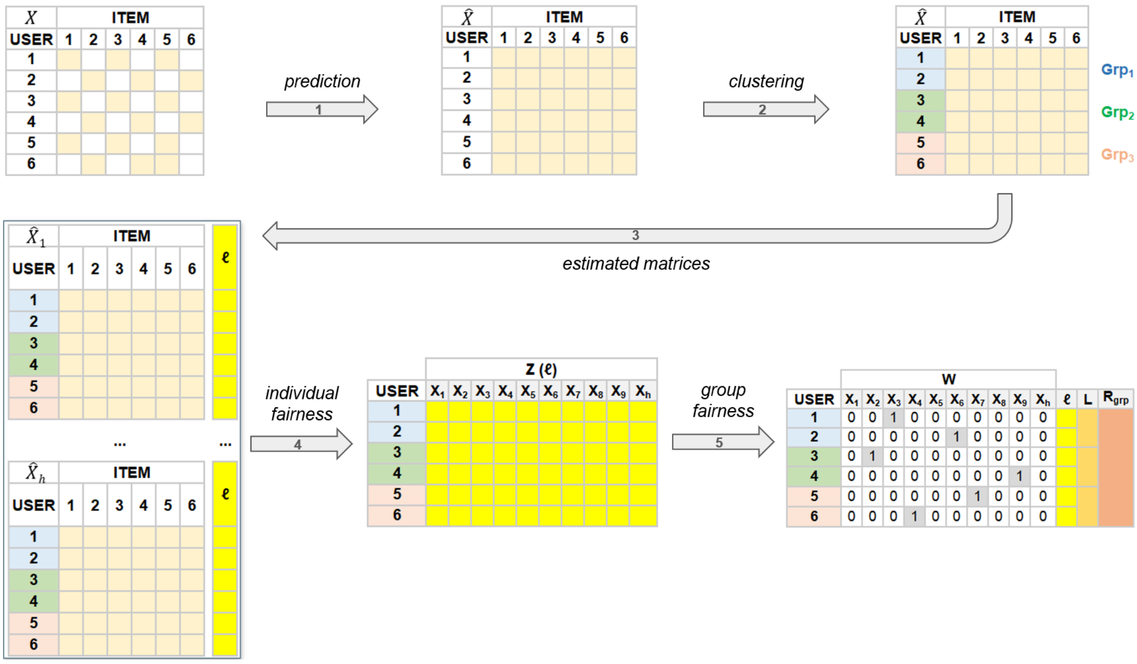

Figura 1. Scheme of the equity algorithm

The five steps of Figure 1 can be detailed as follows:

1. Prediction: the matrix X filled with some ratings is considered by a traditional recommendation system to make predictions of recommendations in $\hat{X}$;
2. Clustering: users are grouped considering some common characteristic;
3. Estimated matrices: h estimated matrices are generated from perturbations of the matrix $\hat{X}$;
4. Individual fairness: the individual losses of each user in each of the h matrices are calculated to assemble the matrix Z;
5. Group fairness: with the help of an optimization algorithm, the matrix W is generated, which represents the best combination of recommendations that minimizes R_{grp} (group unfairness).

Algorithm 1 Equity Algorithm

Require: Partially observed rating matrix $X \in \mathbb{R}^{n \times d}$ of n users and d items, user groups $G = \{G_1, G_2..., G_g\}$, number of estimated matrices to be generated h

Ensure: Estimated matrix considering lower individual losses $\hat{X}_\pi$

Calculate $\hat{X}$

Calculate R_{indv}, R_{grp}, $RMSE$ of $\hat{X}$

Calculate h estimated matrices $\{\hat{X}_1, \hat{X}_2..., \hat{X}_h\}$

$p \leftarrow 0$

while $p \leq h$ **do**

  Calculate R_{indv}, R_{grp}, $RMSE$ of $\hat{X}_p$

end while

Calculate $Z \in \mathbb{R}^{n \times h}$ Matrix of individual losses for n users and h estimated matrices

$\hat{X}_\pi$ = Optimization Algorithm applied to matrix Z

4. Results and Discussions

In this section, we showcase the performance of our reclassification method by emphasizing the recommendation quality and the effectiveness of impartiality compared to those of the traditional recommendation algorithms without awareness of impartiality. Experimental settings regarding the types of user clustering and the number of estimated matrices will be detailed below. In the results presented in tables, cells with the greatest reductions in group unfairness are highlighted in red.

4.1. Experimental Settings

User Clustering. User clustering was conducted under two possible configurations:

- Hierarchical Clustering: Users were grouped into three clusters by using the hierarchical method (Agglomerative Clustering[4]). The main goal in this case was to identify nonobvious clusters.
- 95-5 Clustering: Users were divided into two groups. One group contained the top 5% of users with the highest number of ratings, while the remaining 95% of users were placed in another group. The 5% group was considered privileged users, while the 95% group represented nonprivileged users.

In both configurations, the variable used for identifying clusters was the number of ratings given by users.

Number of Calculated Estimated Matrices (h). To determine the optimal number of estimated matrices to be calculated by the equity algorithm, we tested four possible values of h: 3, 5, 10, 15, and 20. We conducted 10 repetitions for each h value.

The results of the experiments are presented in tables (1, 2, 3 and 4), providing the following information:

- Dataset: Name of the dataset used;
- h: Quantity of calculated estimated matrices;
- Mean: Mean resulting from 10 repetitions of the equity algorithm execution;
- Standard Deviation: Standard deviation resulting from 10 repetitions of the equity algorithm execution;
- (%): Percentage reduction or increase comparing the original mean and the mean resulting from the equity algorithm execution.

Considering the MovieLens data from Table 1, we can observe that as we increase the value of h, the group injustice R_{grp} decreases. For $h = 20$, we managed to reduce the value of R_{grp} to 0.000143071, representing a reduction of 23.07% in R_{grp} calculated prior to executing the equity algorithm.

Table 1 also displays the efficiency of recommendations after applying the equity algorithm. For the MovieLens dataset, as h increases, the $RMSE$ also increases. However, this is a relatively small increase. Considering $h = 20$, the root mean squared error was 0.888179990, representing only a 0.19% increase in $RMSE$. This implies that the recommendations did not significantly lose efficiency after applying the equity algorithm.

[4]https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html

Tabela 1. R_{grp} **and** $RMSE$ **for Hierarchical Strategy on MovieLens Dataset**

Hierarchical Strategy for User Grouping $\{G_1, G_2, G_3\}$ Fairness Measure: Group Injustice (R_{grp}) — Original Mean 0.000185967 Efficiency Measure: Root Mean Squared Error ($RMSE$) — Original Mean 0.886507607							
Dataset	h	$R_{grp}(\mu)$	$R_{grp}(\sigma)$	(%)	$RMSE(\mu)$	$RMSE(\sigma)$	(%)
Movies	3	0.000153749	0.000003509	-17.32	0.890267822	0.000088645	0.42
	5	0.000150642	0.000004826	-19.00	0.889607785	0.000077187	0.35
	10	0.000148635	0.000005320	-20.07	0.889146792	0.000937324	0.30
	15	0.000143388	0.000004164	-22.90	0.888427079	0.000046692	0.22
	20	0.000143071	0.000002504	-23.07	0.888179990	0.000045354	0.19

Tabela 2. R_{grp} **and** $RMSE$ **for Hierarchical Strategy on Songs Dataset**

Hierarchical Strategy for User Grouping $\{G_1, G_2, G_3\}$ Fairness Measure: Group Injustice (R_{grp}) — Original Mean 0.534167908 Efficiency Measure: Root Mean Squared Error ($RMSE$) — Original Mean 0.886811055							
Dataset	h	$R_{grp}(\mu)$	$R_{grp}(\sigma)$	(%)	$RMSE(\mu)$	$RMSE(\sigma)$	(%)
Songs	3	0.416342698	0.012090458	-22.06	0.851455821	0.001627002	-3.99
	5	0.372863539	0.009088738	-30.20	0.848962814	0.001855544	-4.27
	10	0.331597033	0.012828769	-37.92	0.844854076	0.001735787	-4.73
	15	0.311728809	0.007120742	-41.64	0.843491505	0.001043880	-4.88
	20	0.300869422	0.007406372	-43.68	0.842314881	0.001559770	-5.02

Tabela 3. R_{grp} **and** $RMSE$ **for 95-5 Strategy on MovieLens Dataset**

95-5 Strategy for User Grouping $\{G_1, G_2\}$ Fairness Measure: Group Injustice (R_{grp}) — Original Mean 0.000030415 Efficiency Measure: Root Mean Squared Error ($RMSE$) — Original Mean 0.886507607							
Dataset	h	$R_{grp}(\mu)$	$R_{grp}(\sigma)$	(%)	$RMSE(\mu)$	$RMSE(\sigma)$	(%)
Movies	3	0.000008842	0.000000244	-70.93	0.890821795	0.000106706	0.49
	5	0.000008578	0.000000674	-71.80	0.890640562	0.000175177	0.47
	10	0.000008483	0.000000265	-72.11	0.890861685	0.000101311	0.49
	15	0.000008338	0.000000164	-72.59	0.891311053	0.000295554	0.54
	20	0.000008502	0.000000227	-72.05	0.891029760	0.000056276	0.51

Tabela 4. R_{grp} **and** $RMSE$ **for 95-5 Strategy on Songs Dataset**

95-5 Strategy for User Grouping $\{G_1, G_2\}$ Fairness Measure: Group Injustice (R_{grp}) — Original Mean 0.005148096 Efficiency Measure: Root Mean Squared Error ($RMSE$) — Original Mean 0.886811055							
Dataset	h	$R_{grp}(\mu)$	$R_{grp}(\sigma)$	(%)	$RMSE(\mu)$	$RMSE(\sigma)$	(%)
Songs	3	0.000782695	0.000457652	-84.80	0.887953802	0.002763840	0.13
	5	0.000315979	0.000067582	-93.86	0.893862257	0.003343720	0.08
	10	0.000315450	0.000027016	-93.87	0.897279633	0.001907185	1.18
	15	0.000334982	0.000029861	-93.49	0.893956985	0.005758564	0.81
	20	0.000354839	0.000068558	-93.11	0.897136207	0.001446050	1.16

The behavior of the data for the Songs dataset followed the same trend regarding group injustice. The larger the h value is, the lower the group injustice R_{grp}. In this case, the reduction in injustice was greater compared to the reduction in the MovieLens dataset. For $h = 20$, the value of R_{grp} was 0.300869422, representing a reduction of 43.68%.

However, the efficiency measurement of recommendations in the Songs dataset exhibited a different behavior compared to efficiency in the MovieLens dataset. For the Songs dataset, with $h = 20$, the root mean squared error was 0.842314881, resulting in a 5.02% reduction in $RMSE$. This indicates that the recommendations improved in efficiency after applying the equity algorithm.

Tables 3 and 4 show the results of the equity algorithm applied to a 95-5 grouping strategy. The reductions in injustice were even greater.

In Figure 2, we highlight the percentage reduction of group injustice R_{grp}. We compare the group injustices of the estimated matrix resulting from the application of the equity algorithm $\hat{X}_\pi$ with the group injustice calculated for the estimated matrix $\hat{X}$, resulting from the application of a traditional recommendation algorithm. Both datasets (MovieLens and Songs) were considered, as well as the two user grouping strategies (hierarchical and 95-5).

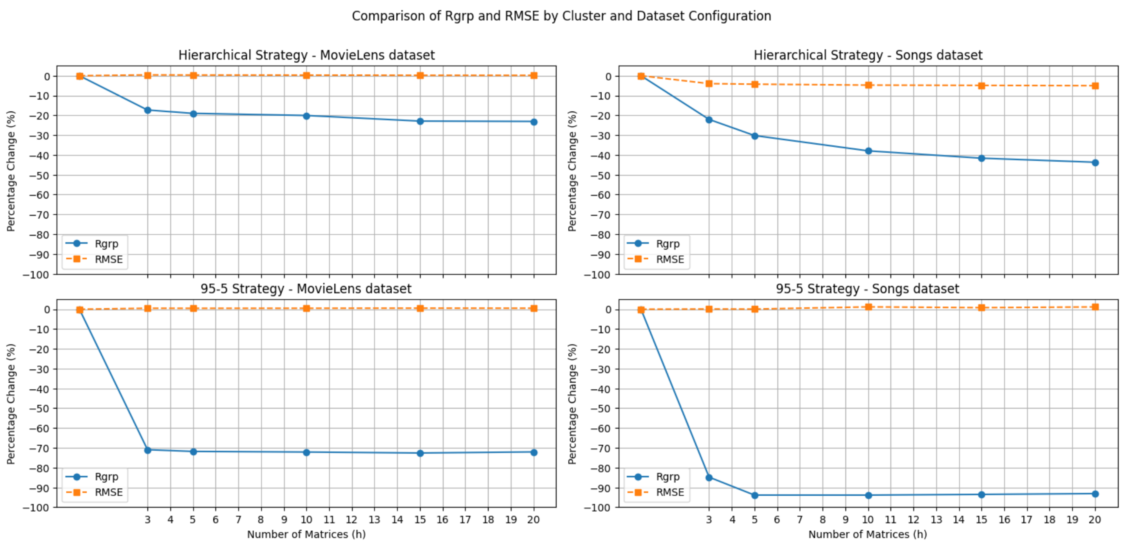

Figura 2. Percentage Reductions of Group Injustice R_{grp}

A crucial aspect to highlight is the significant decrease in group inequality observed across all databases when the 95-5 clustering strategy is adopted. This result aligns with the data-centric nature of recommendation systems, where users with a larger volume of reviews tend to be disproportionately favored in both databases analyzed.

The algorithm's rapid convergence after the inclusion of 5 matrices indicates that the 95-5 clustering strategy efficiently mitigates the initial data inequalities. After reaching an equity threshold, the additional benefits of extra matrices are marginal, suggesting the algorithm's efficiency in correcting inequalities with few modifications. This underscores the importance of identifying the optimal point of equity versus computational cost to enhance the effectiveness and efficiency of recommendation systems.

5. Conclusion

In the present study, we develop and evaluate the equity algorithm in recommendation systems considering different user grouping strategies. The equity algorithm managed to decrease group injustice in both grouping strategies for both datasets. The most significant reductions were observed in the Songs dataset. For instance, for a $h = 10$ in the 95-5 grouping strategy, a reduction of 93.87% in Group Injustice R_{grp} was observed. Regarding the efficiency of recommendations, we note that even considering substantial reductions in injustice, there were no significant losses in efficiency.

For future work, we plan to test the algorithm on additional datasets from diverse contexts, aiming to evaluate its effectiveness across various domains. Furthermore, we intend to analyze the results obtained from different recommendation strategies to gain a deeper understanding of the algorithm's robustness and adaptability. By conducting these experiments, we aim to further validate the applicability of the proposed algorithm and refine its performance under different circumstances.

Referências

Barocas, S. and Selbst, A. D. (2016a). Big data's disparate impact. *Calif. L. Rev.. California Law Review*, 104(IR):671.

Barocas, S. and Selbst, A. D. (2016b). Big data's disparate impact. *California Law Review*, 104(3):671–732.

Berk, R., Heidari, H., Jabbari, S., Joseph, M., Kearns, M. J., Morgenstern, J., Neel, S., and Roth, A. (2017). A convex framework for fair regression. *CoRR*, abs/1706.02409.

Beutel, A., Chi, E. H., Cheng, Z., Pham, H., and Anderson, J. (2017). Beyond globally optimal: Focused learning for improved recommendations. In *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017*.

Biega, A. J., Gummadi, K. P., and Weikum, G. (2018). Equity of attention: Amortizing individual fairness in rankings. *CoRR*, abs/1805.01788.

Bilal Zafar, Isabel Valera, M. G.-R. and Gummadi, K. P. (2017). Training fair classifiers. In *AISTATS'17: 20th International Conference on Artificial Intelligence and Statistics, 2018*.

Boyd, D. and Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15(5):662–679.

Burke, R., Sonboli, N., and Ordonez-Gauger, A. (2018). Balanced neighborhoods for multi-sided fairness in recommendation. In *FAT*.

Celis, L. E., Deshpande, A., Kathuria, T., and Vishnoi, N. K. (2016). How to be fair and diverse? *CoRR*, abs/1610.07183.

Cinus, F., Minici, M., Monti, C., and Bonchi, F. (2022). The effect of people recommenders on echo chambers and polarization. *Proceedings of the International AAAI Conference on Web and Social Media*, 16(1):90–101.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., and Zemel, R. S. (2011). Fairness through awareness. *CoRR*, abs/1104.3913.

Gómez, E., Shui Zhang, C., Boratto, L., Salamó, M., and Marras, M. (2021). The winner takes it all: Geographic imbalance and provider (un)fairness in educational recommender systems. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '21, page 1808–1812, New York, NY, USA. Association for Computing Machinery.

Hardt, M. (2013). On the provable convergence of alternating minimization for matrix completion. *CoRR*, abs/1312.0925.

Hardt, M., Price, E., and Srebro, N. (2016). Equality of opportunity in supervised learning. *CoRR*, abs/1610.02413.

Harper, F. M. and Konstan, J. A. (2015). The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, 5(4):1–19.

Hastie, T., Mazumder, R., Lee, J., and Zadeh, R. (2014). Matrix completion and low-rank svd via fast alternating least squares.

Kamishima, T. and Akaho, S. (2017). Considerations on recommendation independence for a find-good-items task. In *In 11th ACM Conference on Recommender Systems*.

Kamishima, T., Akaho, S., and Asoh, H. (2012). Enhancement of the neutrality in recommendation. In *In Proc. of the 2nd Workshop on Human Decision Making in Recommender Systems*, pages 8–14.

Kamishima, T., Akaho, S., Asoh, H., and Sakuma, J. (2018). Recommendation independence. In Friedler, S. A. and Wilson, C., editors, *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, volume 81 of *Proceedings of Machine Learning Research*, pages 187–201. PMLR.

Kumar, D., Grosz, T., Rekabsaz, N., Greif, E., and Schedl, M. (2023). Fairness of recommender systems in the recruitment domain: an analysis from technical and legal perspectives. *Frontiers in Big Data*, 6.

Li, Y., Chen, H., Fu, Z., Ge, Y., and Zhang, Y. (2021). User-oriented fairness in recommendation. In *Proceedings of the Web Conference 2021*, WWW '21. ACM.

Mehrotra, R., McInerney, J., Bouchard, H., Lalmas, M., and Diaz, F. (2018). Towards a fair marketplace: Counterfactual evaluation of the trade-off between relevance, fairness & satisfaction in recommendation systems. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, New York, NY, USA. Association for Computing Machinery.

Niemiec, W., Borges, R., and Barone, D. (2022). Artificial intelligence discrimination: how to deal with it? In *Anais do III Workshop sobre as Implicações da Computação na Sociedade*, pages 93–100, Porto Alegre, RS, Brasil. SBC.

Rastegarpanah, B., Gummadi, K. P., and Crovella, M. (2018). Fighting fire with fire: Using antidote data to improve polarization and fairness of recommender systems. *CoRR*, abs/1812.01504.

Ruback, L., Avila, S., and Cantero, L. (2021). Vieses no aprendizado de máquina e suas implicações sociais: Um estudo de caso no reconhecimento facial. In *Anais do II Workshop sobre as Implicações da Computação na Sociedade*, pages 90–101, Porto Alegre, RS, Brasil. SBC.

Sweeney, L. (2013). Discrimination in online ad delivery.

Tang, J., Shen, S., Wang, Z., Gong, Z., Zhang, J., and Chen, X. (2023). When fairness meets bias: a debiased framework for fairness aware top-n recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems*, RecSys '23, page 200–210, New York, NY, USA. Association for Computing Machinery.

Taso, F., Reis, V., and Martinez, F. (2023). Discriminação algorítmica de gênero: Estudo de caso e análise no contexto brasileiro. In *Anais do IV Workshop sobre as Implicações da Computação na Sociedade*, pages 13–25, Porto Alegre, RS, Brasil. SBC.

Wang, B. and Gong, N. Z. (2018). Stealing hyperparameters in machine learning. *CoRR*, abs/1802.05351.

Wang, Y., Ma, W., Zhang, M., Liu, Y., and Ma, S. (2023). A survey on the fairness of recommender systems. *ACM Trans. Inf. Syst.*, 41(3).

Yao, S. and Huang, B. (2017). Beyond parity: Fairness objectives for collaborative filtering. *CoRR*, abs/1705.08804.

Zafar, M., Valera, I., Rodriguez, M., and Gummadi, K. P. (2017). Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1171–1180.

Zehlike, M., Sühr, T., Baeza-Yates, R., Bonchi, F., Castillo, C., and Hajian, S. (2022). Fair top-k ranking with multiple protected groups. *Information Processing & Management*, 59(1):102707.

Zemel, R., Wu, Y., Swersky, K., Pitassi, T., and Dwork, C. (2013). Learning fair representations. In Dasgupta, S. and McAllester, D., editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 325–333, Atlanta, Georgia, USA. PMLR.