 DOI: 10.11576/dataninja-1157

Is it Possible to Characterize Group Fairness in Rankings in Terms of Individual Fairness and Diversity?

Chiara Balestra CHIARA.BALESTRA@CS.TU-DORTMUND.DE
TU Dortmund, Germany

Abstract

Rankings are ever-present in everyday life. Examples are the results of personalized recommendations and web search queries. Rankings can result from an algorithm, importance scores and human-based rankings of items. Till we are not concerned with societal applications, the "fairness" of the ranking is often irrelevant; however, problems appear when switching from depersonalized items to individuals. Then, suddenly, fairness becomes an issue.

We investigate the relationships among group fairness, individual fairness, diversity, and Shapley values. Far from being a comprehensive survey of fairness-related papers or proposing a new method, we want to raise awareness of the chaos we are trying to navigate and propose some new research direction we are trying to follow.

Keywords: Shapley values, individual fairness, group fairness, diversity.

1. Fairness and Shapley values

We start from where it "all" started. Fairness in machine learning is a relatively new branch; the necessity to study algorithms from a fairness perspective derives from the unpleasant discovery that some algorithms, implemented in critical contexts, were being racist. Implementing machine learning algorithms for risk assessment [1] by financial institutions, training image classification models on biased data, and selecting candidates for job positions represent a few of the critical societal applications where fairness is essential; different groups of individuals need to be fairly and "equally" treated.

But what is "fairness"? How can we say that a prediction, a ranking, or an algorithm is fair? We start by introducing "individual fairness", and particularly by introducing Shapley values, said "fair" scores in Cooperative Game Theory [2]. A *cooperative game* is a pair $(\mathcal{N}, f)$ where $\mathcal{N}$ is a finite set of players $\mathcal{N} = \{1, \ldots, N\}$ and f is a function over the power set of players $\mathcal{P}(\mathcal{N})$, i.e., $f : \mathcal{P}(\mathcal{N}) \mapsto \mathbb{R}$. f is the *value function* of the game; the role of the value function is to assign to sets of players a real number, and it is usually assumed to satisfy some mathematical properties, i.e., $f(\emptyset) = 0$, it is "non-negative" and it is "monotone". Under the monotonicity assumption, the *grand coalition* $\mathcal{N}$ is the set assuming the maximum of the value function f.

Shapley values assign to each player his worth in the game $(\mathcal{N}, f)$, their values sum up to $f(\mathcal{N})$ and they are "fair" concerning the value brought by each player to the coalitions. The Shapley value of player i is formally defined as

$$\phi_f(i) = \sum_{\mathcal{A} \subseteq \mathcal{N} \setminus i} \frac{1}{N\binom{N-1}{|\mathcal{A}|}} \left[f(\mathcal{A} \cup i) - f(\mathcal{A}) \right]. \tag{1}$$

Shapley values derive their popularity from their "nice to have" properties, i.e., the *Pareto optimality*, the *dummy*, the *linearity*, and the *symmetry property* [2]. Particularly interesting for us are the *dummy property*, stating that given $i \in \mathcal{N}$ such that $f(\mathcal{A} \cup \{i\}) = f(\mathcal{A})$ for each $\mathcal{A} \subseteq \mathcal{N}$ it holds $\phi_f(i) = 0$, and the *symmetry property*, claiming that given two players $i, j \in \mathcal{N}$ such that $f(\mathcal{A} \cup \{i\}) = f(\mathcal{A} \cup \{j\})$ for each $\mathcal{A} \subseteq \mathcal{N}$ it holds $\phi_f(i) = \phi_f(j)$.

The definition of fairness in the Oxford Dictionary reads, "Fairness is the quality of treating people equally or in a way that is reasonable". The definition is quite fuzzy; nevertheless, it was needed to formalize it mathematically. This resulted in several proposals of mathematical definitions; the first distinction is between "individual" and "group fairness". Individual fairness refers to the similar treatment of "similar" individuals. Group fairness refers to the treatment of "different" groups and usually includes the ethical concerns of gender parity, race, and sexual orientation; the so-called "protected attributes" are usually defined by law and morality concerns. Individual fairness, instead, does "not necessarily" care of morality issues: the similarity among individuals can be defined with respect to "any" attribute, either protected or not by law.

2. Contradictions within fairness

The intrinsic fairness of Shapley values derives from two of their properties, i.e., the dummy and symmetry properties [2], which guarantee that two players with similar characteristics obtain similar Shapley values. On the other side, two recent works [3, 4] show how the result of these two properties is essentially a "redundancy unawareness" of the importance scores obtained through Shapley values. The concepts of "redundancy unawareness" and "individual fairness" are eventually the same. So why do we claim in some contexts that they represent an advantage and in others that they represent a disadvantage? To understand this, we need to relate it with the *(group) fairness* in rankings.

2.1. Fairness in rankings

Rankings are spread in any field, from everyday life to complex machine learning algorithms. Rankings are nothing more than ordered lists of elements, items, or individuals. Rankings often go hand in hand with importance scores; however, if rankings are trivial to obtain from the corresponding importance scores, the opposite does not hold.

Given the several applications in society [5–7], issues relative to the fairness of the rankings and their evaluation play an essential role. The need to explore the rankings fully, which is not fulfilled in most real-world contexts, implies that elements not ranked in the top positions suffer from low visibility; this fact is usually referred to as *position bias*, and it is particularly relevant when it affects the items belonging to different various groups in a dissimilar manner. Position biases can affect differently protected and unprotected communities and potentially propagate gender, sex, and sexual orientation biases against marginalized groups. To address the issue, one could define *group fairness* constraints, with the aim of guaranteeing the same treatments in the various groups; Singh and Joachims [8] define constraints for "fairness of exposure" in ranking outputs, and the work by Zehlike et al. [9] deals with the fair top-k ranking problem. We stated that Shapley values satisfy the "individual fairness", but this might still be in contrast with the definition of "group fairness".

2.2. Diversity and individual fairness achieve group fairness?

So far, we have introduced group fairness, individual fairness, and Shapley values. The individual fair ranking derived from Shapley values does not respect any group fairness condition; this can be easily concluded by observing that elements in groups, e.g., highly correlated groups, are similarly ranked. Furthermore, the pruning criteria proposed in [3, 4] avoid that highly correlated elements are similarly ranked; in other words, the pruning criteria include "diversity" in the rankings. Diversity [10] was introduced in Recommender Systems and refers to the property of the recommendations to propose items that are new to the user and not "too similar" to the already seen elements in various parts of the ranking.

We claim that a combination of "diversity" and "individual fairness" in importance scores can induce a ranking that respects the "group fairness" property. The claim can be potentially generalized to any importance scores, independently of their derivation and it is not limited to Shapley values importance scores. Our claim is supported by the works by Balestra et al. [3, 4]; furthermore, the concern is becoming relevant to the community [11], where the authors study the connection between fairness and novelty in RS. Although some preliminary experiments showed the connection between group fairness and the simultaneous satisfaction of diversity and individual fairness in Shapley values, additional analysis must be performed to prove it more generally. Therefore, our claim is still far from being generally proven.

3. Conclusions

We introduced the relationship between Shapley values, well known for providing individual fair rankings, and the lack of diversity in the provided rankings. We propose a new theory, claiming that the pruning criteria proposed in [3, 4] can be interpreted as adding "diversity" in the rankings; more generally, we claim that under some specific (and still under study) conditions, the equation "diversity" plus "individual fairness" equals "group fairness" holds.

Acknowledgments

This research was supported by the research training group Dataninja funded by the German federal state of North Rhine-Westphalia.

References

[1] Jonathan N Crook, David B Edelman, and Lyn C Thomas. Recent developments in consumer credit risk assessment. *European Journal of Operational Research*, 183, 2007.

[2] L. S. Shapley. A value for n-person games. In *Contributions to the Theory of Games*, volume II. 1953.

[3] Chiara Balestra, Florian Huber, Andreas Mayr, and Emmanuel Müller. Unsupervised features ranking via coalitional game theory for categorical data. In *DaWaK*, 2022.

[4] Chiara Balestra, Carlo Maj, Emmanuel Müller, and Andreas Mayr. Redundancy-aware unsupervised ranking based on game theory: Ranking pathways in collections of gene sets. *Plos one*, 18, 2023.

[5] Peter Emerson. The original borda count and partial voting. *Social Choice and Welfare*, 40, 2013.

[6] Christian List. Social choice theory. 2013.

[7] Cynthia Dwork, Ravi Kumar, Moni Naor, and Dandapani Sivakumar. Rank aggregation methods for the web. In *WWW*, 2001.

[8] Ashudeep Singh and Thorsten Joachims. Fairness of exposure in rankings. In *KDD*, 2018.

[9] Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. Fa*ir: A fair top-k ranking algorithm. In *CIKM*, 2017.

[10] Pablo Castells, Neil Hurley, and Saul Vargas. Novelty and diversity in recommender systems. In *Recommender systems handbook*. 2021.

[11] Yuying Zhao, Yu Wang, Yunchao Liu, Xueqi Cheng, Charu Aggarwal, and Tyler Derr. Fairness and diversity in recommender systems: a survey. *arXiv preprint arXiv:2307.04644*, 2023.