DECISION MAKING WITH FAIR RANKING

Zorica A. Dodevska
The Institute for Artificial Intelligence Research and Development of Serbia

Boris Delibašić
University of Belgrade

Sandro Radovanovic
University of Belgrade

DECISION MAKING WITH FAIR RANKING

Zorica A. Dodevska*[12], Boris Delibašić[1], Sandro Radovanović[1]
[1]Faculty of Organizational Sciences, University of Belgrade
[2]Research and Development Institute Lola Ltd, Belgrade
*Corresponding author, e-mail: zd20185004@student.fon.bg.ac.rs

Abstract: *Ranking is a sensitive process because it involves working with sensitive attributes that can discriminate alternatives. Due to the availability of a large amount of data for automated processing, ranking is increasingly in use. Therefore, concepts of algorithmic fairness in the field of classification in machine learning find their place in fair ranking methods. This paper provides an overview of fair ranking terms, fair ranking challenges, and fair ranking algorithms from the state-of-the-art literature.*

Keywords*: decision making, fair ranking, algorithmic fairness, discrimination*

1. INTRODUCTION

Ranking is at the core of decision making. Decision-makers first rank alternatives according to a set of criteria and then choose the best or top k best alternatives. Ranking is part of the standard procedure of many decision-making established methods, for example, ranking options according to their obtained utility in the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) or in the Analytic Hierarchy Process (AHP). It can also provide input data in decision-making processes. In collective decision making, decision-makers can easily collect data by using an ordinal scale. For example, voters can rank alternatives according to some single-winner voting rules (e.g., Borda, Condorcet) or multi-winner voting rules (e.g., Chamberlin-Courant), or participants can rank options in questionnaires (e.g., by using the Lickert scale).

Because ranking simplifies decision-making information (Kulhman et al., 2019), it has frequent applications. Ranking objects can be universities (Johnes, 2018), researchers and research teams (Vavryčuk, 2018), job applicants (Encheva, 2019), etc. Also, an increasing number of algorithms process online data based on which personalized recommended systems provide users with ranked options according to their preferences. However, since simplicity produces inequity (Kleinberg and Mullainathan, 2019), ranking from the viewpoint of ethical consideration of algorithmic discriminatory bias is the focus of interest in the field of fair machine learning.

Machine learning algorithms can deepen the already existing bias among historical data. Considering *sensitive* attributes (such as race, gender, age, religion, ethnicity, health status, income, etc.), they learn from already presented discriminatory practices and embed them into future decisions. Even acting against intentional discrimination does not guarantee to eliminate bias. In the case of excluding *sensitive* attributes from consideration (for which *protected* attributes is a synonym in the law), there is still a correlation in the data (Hajian et al., 2016). For example, regardless of gender hiding, occupation-gender associations from historic human-like biases still exist (Caliskan et al., 2017); therefore, a proxy for the *protected* attribute. Eliminating *sensitive* attributes can, on the contrary, additionally harm already disadvantaged groups (Corbett-Davies and Goel, 2018). Hence, ranking models used in automated decision making demand development of fairness ranking measures to protect underprivileged groups appropriately.

The paper aims to contribute to the existing corpus of knowledge in the field of fair ranking, by summarizing fair ranking terms, identifying challenges of fair ranking, and resuming proposed algorithms from the state-of-the-art literature. Conclusion remarks contain guidelines for further work to support the efforts in this field.

2. RELATED WORK

Fairness doctrine acts against discrimination of minority groups (group-level fairness) or individuals (individual fairness). In machine learning, fairness principally focuses on the classification outcome of observed subjects (e.g., an applicant is awarded funds or not). Group-level fairness techniques include a set of criteria that support the equalization of positive outcomes across groups (e.g., *demographic parity* or *statistical parity* criterion requires equality in the proportion of positive outcomes across sub-populations). Here, a feature (i.e., *sensitive* attribute) divides the population into two or more disjoint groups, from which at least one is disadvantaged because of past discriminatory stereotypes that become a part of wrongly grounded predictions. Individual fairness advocates consistency, in the sense that similar individuals should have similar outcomes. One example of individual fairness measure is *counterfactual fairness.* According to Kusner et al. (2017), it requires the same predictions for an individual in the actual world and a counterfactual world (where the individual is a member of a different demographic group).

Gajane and Pechenizkiy (2018) point out that algorithmic fairness in machine learning except considering parity in impact (i.e., result), can also require equality in treatment (i.e., the process of classification). For example, *unawareness* (or *„anti-classification"*) requires not explicitly usage of *sensitive* attributes in the decision-making process. Also, algorithmic fairness can focus on accuracy and errors. For example, *equalized odds* require equal accuracy – *true-positive rates* and *false-positive rates* across all groups (Hardt et al., 2018). The same authors introduce *equal opportunity* that imposes an equal *true-positive rate* independently from the group membership. C*alibration* observes applicants with a specified risk score (Corbett-Davies and Goel, 2018); it requires equal *false-negative rates* among them.

2.1. Fair ranking terms

A fair ranking is a ranking without discrimination of ranked items, with particular emphasis on items that belong to disadvantaged/protected groups. Having that in mind, according to Castillo (2018), fair ranking should satisfy at least the following conditions:

- It should avoid statistical bias (which means sufficient presence of items across groups);
- It should fulfill individual fairness (which demands a consistent treatment of similar items); and
- It should provide a proper representation of items in the ranked population.

Literature about fair ranking includes various terms – criteria, metrics, and measures (listed and defined in Table 1). Kuhlman et al. (2019) present the following criteria: *rank equality*, *rank calibration*, and *rank parity*, and supporting pairwise error metrics for two observed groups that compare the actual ranking and the learned ranking. Zehlike et al. (2017) suggest a fair top-k ranking set of criteria: *selection utility*, *ordering utility*, and *ranked group fairness*. Different metrics are in circulation, for example, for search (Wang et al., 2013). Yang and Stoyanovich (2017) provide three measures of *statistical parity* (for more information, please see Table 1.). Concerning the actors, Burke (2017) identifies several criteria in recommendation systems. Castillo (2018) distinguishes attention-based and probability-based measures.

Table 1: Fair ranking terms

Name of the term	What does it present?	Definition	Author(s)
Rank equality	Criterion	"No group should be unfairly privileged or penalized compared to another group" when ranking.	Kuhlman et al. (2019)
Rank equality error	Measure	It represents the ratio of the number of discordant mixed pairs and the total number of mixed pairs (they include one object from each group).	Kuhlman et al. (2019)
Rank calibration	Criterion	A probabilistic classifier should predict the ranking of objects from each group appropriately.	Kuhlman et al. (2019)
Rank calibration error	Measure	It represents the ratio of the number of discordant pairs that contain objects from the target group and the total number of pairs that contain at least one object from the target group.	Kuhlman et al. (2019)
Rank parity	Criterion	*Statistical parity* in the top-k prefix of a ranking.	Yang and Stoyanovich (2017), Kuhlman et al. (2019)
Rank parity error	Measure	It represents the ratio of the number of pairs from the learned ranking that favors one group over another, and the total number of mixed pairs.	Kuhlman et al. (2019)
Selection utility	Criterion	Each top-k applicant is more qualified compared to each applicant outside of the top-k.	Zehlike et al. (2017)
Ordering utility	Criterion	Top-k applicants "should be ordered by decreasing qualifications."	Zehlike et al. (2017)
Ranked group fairness	Criterion	Top-k applicants "should fairly represent the protected group."	Zehlike et al. (2017)
Normalized discounted	Metrics	It represents a normalization of DCG measure, which is a weighted sum of relevance degree of	Wang et al. (2013)

cumulative gain (NDCG)		ranked items. Commonly used in the search, the weight serves a decreasing function of object rank (i.e., position).	
Rank drop	Metrics	It shows the maximum number of positions that one object has lost on the ranked list.	Kuhlman et al. (2019)
Normalized discounted difference (rND)	Measures	It calculates the difference in the proportion of applicants from the protected group between top-k and the overall population.	Yang and Stoyanovich (2017)
Normalized discounted KL-divergence (rKL)	Measures	It calculates the expectation of the difference of protected group membership between top-k and the overall population.	Yang and Stoyanovich (2017)
Normalized discounted ratio (rRD)	Measures	It is similar to *rND*, but it can be applied only if the protected group numerically encompasses at most 50% of the population (i.e., a smaller part), and when fairness probability is less than 0.5.	Yang and Stoyanovich (2017)
C-fairness	Criterion	It requires fairness for consumers (or subjects), i.e., the disparate impact of the recommendation on their protected classes.	Burke (2017)
P-fairness	Criterion	It requires fairness for providers (or objects) only.	Burke (2017)
CP-fairness	Criterion	It requires fairness for both consumers and providers, which is a case in the reciprocal recommendation or when both belong to protected groups.	Burke (2017)
Attention-based measures	Measures	They measure whether actual or potential attention (e.g., the *fairness of exposure* (Singh and Joachims, 2018) vs. disparate treatment).	Castillo (2018)
Probability-based measures	Measures	They measure deviations between the expected and observed characteristics of the ranking.	Castillo (2018)

3. CHALLENGES OF FAIR RANKING

Several problems of fair ranking impose essential limitations in the literature:

- The design of ranking systems should suit the ranking issue (Asudeh et al., 2019), so there are no guarantees for universal solutions.
- Ranking accuracy is often important for all positions in the rankings, not just among the top k ranks (Kuhlman et al., 2019). One such situation is applying for funds when applicants on positions below the line remain without funds. Still, if they received funds, then the significance of their positions/ranks becomes less important.
- Measuring errors in ranking models requires metrics according to the task (Kuhlman et al., 2019), for example, binary notation of classes for *true-positive rate* and *true-negative rate*, and development of fair metrics for ranking objects - e.g., researchers, etc.
- The ranking process is susceptible. Singh and Joachims (2018) point out that even “small differences in item relevance can cause a large difference in exposure and therefore economic opportunity across groups.”
- Transparency of ranking models and explainable rankings become increasingly important requirements. Many recommendation systems have their ranking algorithms. However, the question arises as to the transparency of ranking procedures. How does the rank join the ranking objects? What is behind the rank? Does it justify the quality in terms of meeting the prescribed ranking criteria? Although transparency would help build trust, it is still rare in search engines and web platforms (Castillo, 2018).
- Gajane and Pechenizkiy (2018) conclude that it is difficult to quantify and mathematically formalize social issues (e.g., such as unequal access to resources). However, they stress the importance of finding a way to incorporate those issues in fairness formalizations.
- It is important to know what you try to rank because the accurate and understandable ranking models are not possible without extensive and proper knowledge about details of ranking objects (Schoenhagen, 2019). So, ranking systems should first effectively elicit data from users in order to satisfy their needs (Schoenhagen, 2019).
- Harmonizing individual and group fair metrics at the same time is a challenging task, and often unattainable. Therefore, in many cases, it is justified to define an acceptable threshold. But again, the question remains on what basis experts define those thresholds.

4. FAIR RANKING ALGORITHMS

Various fair ranking algorithms from the literature propose to resolve fair ranking problems. Table 2 summarizes some of them.

Table 2: Fair ranking algorithms

Name of algorithm	What does it do?	Author(s)
FARE (Fair Auditing based on Rank Error)	It provides fairness diagnostics for error-based fairness criteria customized for ranking.	Kuhlman et al. (2019)
FA*IR	It is an algorithm for resolving the fair top-k ranking problem.	Zehlike et al. (2017)
DELTR (Disparate Exposure in Learning to Rank)	It addresses the potential issue of disparate exposure in ranking in-processing, i.e., at training time.	Zehlike et al. (2020a)
FairSearch	It is an open-source library that provides fairness in the ranked search results.	Zehlike et al. (2020b)
Ranking with Fairness Constraints	It is a linear time approximation algorithm of constrained ranking maximization problem used for processing ethical data.	Celis et al. (2018)
Designing Fair Ranking Schemes	It provides scoring ranking functions that use a weighted sum of numeric attribute values.	Asudeh et al. (2019)
Fairness of Exposure in Rankings	It allows expression of fairness constraints on rankings concerning exposure allocation.	Singh and Joachims (2018)

5. CONCLUSION REMARKS

The use of ranks simplifies information for decision-makers, but the process of assigning ranks to objects is undoubtedly not a simple one. Fair ranking is a complex task; as there are different forms and sources of discrimination, and therefore different criteria and measures. This paper gives an overview of them. To overcome the identified challenges in this area, Figure 1 shows guidelines for in-processing steps of the decision-making process with fair ranking.

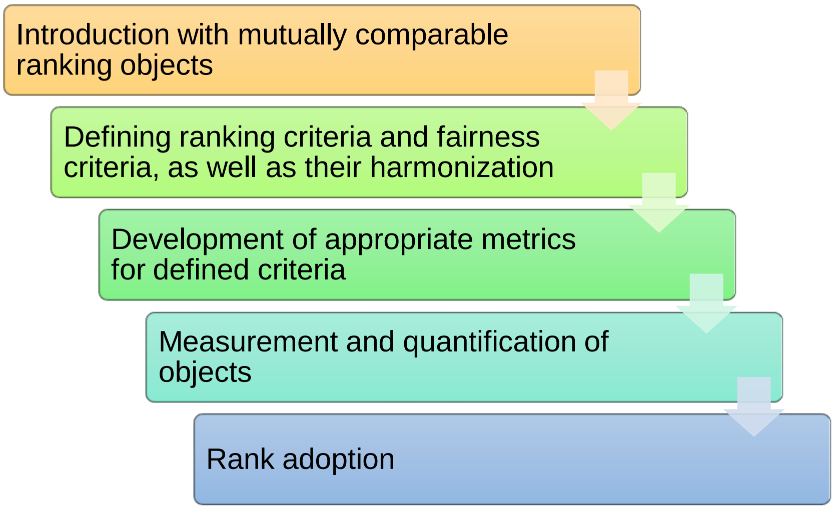

Figure 1: In-processing steps of the decision-making process with fair ranking

The ability to find and rank data in the age of the Internet and Big data is a great advantage (Schoenhagen, 2019) but, at the same time, a responsibility. The paper points out some of the prominent algorithms in this area. Therefore, fair machine learning has a task to prevent further development of algorithmic discriminatory practices.

Acknowledgments. This paper is the result of the project ONR - N62909-19-1-2008 supported by the Office of Naval Research, the United States: *Aggregating computational algorithms and human decision-making preferences in multi-agent settings*. The Ministry of Education, Science and Technological Development of the Republic of Serbia also supports the paper.

REFERENCES

[1] Asudeh, A., Jagadish, H. V., Stoyanovich, J., & Das, G. (2019, July). Designing Fair Ranking Schemes. In *2019 International Conference on Management of Data (SIGMOD '19)* (18 pages), Amsterdam, Netherlands. New York, NY, USA: ACM. doi: 10.1145/3299869.3300079

[2] Burke, R. (2017, August). Multisided Fairness for Recommendation. Presented at the *2017 Workshop on Fairness, Accountability, and Transparency in Machine Learning (FATML'17*), Halifax, NS, Canada.

[3] Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. Science, *356*(6334), 183–186. doi:10.1126/science.aal4230

[4] Castillo, C. (2018, October). Fairness and Transparency in Ranking. Paper presented at the *ACM International Conference on Information and Knowledge Management (CIKM'18)*. Turin, Italy.

[5] Celis, L., Elisa, Straszak, D., & Vishnoi, Nisheeth K. (2018). Ranking with fairness constraints. arXiv:1704.06840v4 [cs.DS] 30 Jul 2018

[6] Corbett-Davies, S., & Goel, S. (2018). The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning. arXiv:1808.00023v2 [cs.CY] 14 Aug 2018

[7] Encheva, S. (2019). Similarity Measures for Ranking Job Applicants. In X-S. Yang., S. Sherratt, N. Dey, A. Joshi (Eds.) *Third International Congress on Information and Communication Technology.* Advances in Intelligent Systems and Computing, vol 797 (pp. 943–949). Singapore: Springer. doi: 10.1007/978-981-13-1165-9_86

[8] Gajane, P., & Pechenizkiy, M. (2018). On Formalizing Fairness in Prediction with Machine Learning. arXiv:1710.03184v3 [cs.CY] 28 May 2018

[9] Hajian, S., Bonchi, F., & Castillo, C. (2016, August). Algorithmic Bias: From Discrimination Discovery to Fairness-aware Data Mining. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)* (pp. 2125–2126). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/2939672.2945386

[10] Hardt, M., Price, E., & Srebro, M. (2016, Dec). Equality of Opportunity in Supervised Learning. Paper presented at the *30th Conference on Neural Information Processing Systems (NIPS 2016)*, Barcelona, Spain.

[11] Johnes, J. (2018). University rankings: What do they really show? *Scientometrics, 115*, 585–606. doi: 10.1007/s11192-018-2666-1

[12] Kleinberg, J., & Mullainathan, S. (2019). Simplicity Creates Inequity: Implications for Fairness, Stereotypes, and Interpretability. arXiv:1809.04578v2 [cs.LG] 2 Jun 2019

[13] Kuhlman, C., VanValkenburg, MA., & Rundensteiner, E. (2019, May). FARE: Diagnostics for Fair Ranking using Pairwise Error Metrics. In Proceedings of the *2019 World Wide Web Conference (WWW '19)* (7 pages), San Francisco, CA, USA. New York, NY, USA: ACM. doi: 10.1145/3308558.3313443

[14] Kusner, M., Loftus, J., Russell, C., & Silva, R. (2017, Dec). Counterfactual Fairness. Paper presented at the *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA.

[15] Schoenhagen, P. M. (2019). Fair Ranking System. Retrieved from https://digitalcommons.wpi.edu/mqp-all/6902

[16] Singh, A., & Joachims, T. (2018, August). Fairness of Exposure in Rankings. In *Proceedings of the KDD 2018 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2219–2228), London, UK.

[17] Vavryčuk, V. (2018). Fair ranking of researchers and research teams. PLoS ONE, *13*(4): e0195509. doi: 10.1371/journal.pone.0195509

[18] Wang, Y., Wang, L., Li, Y., He, D., Chen, W. & Liu, T.-Y. (2013). A theoretical analysis of NDCG type ranking measures. *COLT*.

[19] Yang, K., & Stoyanovich, J. (2017, June). Measuring Fairness in Ranked Outputs. In *Proceedings of SSDBM '17* (6 pages), Chicago, IL, USA. doi: 10.1145/3085504.3085526

[20] Zehlike, M., & Castillo, C. (2020a, April). Reducing Disparate Exposure in Ranking: A Learning To Rank Approach. In *Companion Proceedings of the Web Conference 2020 (WWW '20 Companion)* (7 pages) Taipei, Taiwan. New York, NY, USA: ACM. doi:10.1145/3366424.3383534

[21] Zehlike, M., Bonchi, F., Castillo, C., Hajian, S., Megahed, M., & Baeza-Yates, R.A. (2017, November). FA*IR: A Fair Top-k Ranking Algorithm. In *Proceedings of the CIKM 2017 ACM on Conference on Information and Knowledge Management* (pp. 1569–1578), Singapore. New York, NY, USA: Association for Computing Machinery. doi: 10.1145/3132847.3132938

[22] Zehlike, M., Sühr, T., Castillo, C., & Kitanovski, I. (2020b, April). FairSearch: A Tool For Fairness in Ranked Search Results. In *Companion Proceedings of the Web Conference 2020 (WWW '20 Companion)* (4 pages), Taipei, Taiwan. New York, NY, USA: ACM. doi: 10.1145/3366424.3383534