Generating Fair Rankings with Multinomial Protected Attributes

Situation: Rankings are very widely used in our life to set priorities, which are closely related to making decisions. Accordingly, our concerns about algorithmic bias in rankings, in which the rankings are generated unfairly, have grown bigger. Therefore, generating fair rankings is an indispensable work as it builds the basis for our decisions and indeed, ranking algorithms have risen to a significant topic in machine learning. According to prior studies, the algorithm for generating a fair top-k ranking: FA*IR, which assures the appearance of specific percentage of protected group from a binomial protected attribute, has already been invented. Now our task remains to further develop this algorithm for multinomial protected attributes.

Proposed Solution: For the test of the algorithm, the fairness constraints should be specified clearly at the beginning. A fairness test should be performed with the output of the algorithm to see if all the constraints are still satisfied.

1. Receive as parameters:
   - The number of total items to be ranked (m)
   - The number of items to be selected for the ranking (k)
   - The number of groups in the protected attributes (g)
   - The percentage of each protected group (p_1, p_2 … p_n) that must be assured in the ranking as parameters.
2. Since each group should be treated individually, they should all be separated from an initial list of items and each of them forms its own group ranking in decreasing order of its quality. Every item that is placed on the first position of these rankings should now have the highest quality in the group, which indicates that when an item should be selected into the final ranking, only the heads of the rankings should be considered.
3. Calculate the least number of each group of items that should be placed in the ranking on each position according to their p and save the values into an array size of the number of protected groups (L).
   - Based on FA*IR: Statistical method to calculate the least number of each group of items that should be guaranteed. The two methods below are proposed:
     a. Multinomial Cumulative Distribution Function (CDF): As the protected attributes have multinomial distribution, their probability of cumulative distribution function should be used as the p-value.
     b. Binomial Distribution Function (CDF): As each component of Multinomial Distribution follows Binomial Distribution with each parameter p_g, use Binomial CDF to calculate the least number of group populations for protected attribute.
   - Significance parameter α corresponding to the Type I error should be specified.

4. Creates another array size of the number of protected groups (R) and initialize it with 0 in every position. This array will be used to record the number of each group of items that are selected into the rank.
5. Generate the final ranking iteratively (n-times):
    - Compare L with R first. If any item record (R) is smaller than the least item number required (L), take the head item of this group ranking.
    - Otherwise, compare the quality of all the head items from each group ranking and take the one with the best quality.

Benefits:

- Fair ranking system could be achieved with multinomial protected attributes, which will assure a certain percentage of protected attributes to appear in the ranking.
- This algorithm could form the base for the further development of ranking algorithms.

Possible Obstacles:

- The "quality" of items should be defined precisely. The standard for measuring the quality is yet too vague.
- The algorithm could be very slow if there are too many protected attributes, which are to be used to generate the rankings.
- An item might belong to more than one protected attribute.
- At each position of the ranking, only one item could be picked, while the percentage of the population changes for all the groups. It might violate the fairness constraints when more than one protected group items must be placed at once as it moves to the next position.